

KHAZAR UNIVERSITY

Faculty: Graduate School of Science, Arts and Technology

Department: English Language and Literature

Speciality: Translation

MASTER'S THESIS

Theme: A comparative quality assessment of ChatGPT-4 and human translation of scientific texts

Student: Sabina Ahmadova

Supervisor: Prof. Dr. Huseynagha Rzayev

Baku – 2024

XƏZƏR UNIVERSİTETİ

Fakültə: Təbiət elmləri, Sənət və Texnologiya Yüksək Təhsil

Departament: İngilis dili və Ədəbiyyatı

İxtisas:– Tərcümə (İngilis dili)

MAGİSTR TEZİSİ

Mövzu: ChatGPT-4 və İnsan tərəfindən elmi mətnlərin tərcüməsinin müqayisəli keyfiyyət qiymətləndirilməsi

Magistrant: Səbinə Əhmədova

Elmi rəhbər: Prof. Hüseynağa Rzayev

Bakı-2024

TABLE OF CONTENTS

INTRODUCTION.....	4
CHAPTER I. LITERATURE REVIEW.....	7
1.1 The history of Machine Translation.....	7
1.2 Neural Networks.....	10
1.3 Deep Learning from 2006 to the present.....	13
1.4 ChatGPT-4 and ChatGPT-3.5 models.....	14
1.5 Metrics.....	20
CHAPTER II. METHODOLOGY AND DATA ANALYSIS.....	23
2.1 Data Gathering Instruments of the Study.....	23
2.2 Corpus and Preparing the Data for Analysis.....	24
2.3 The Procedure of Data Analysis.....	24
CHAPTER III. DISCUSSION AND RESULTS.....	25
CONCLUSION.....	82
DEFINITION OF KEY TERMS.....	83
REFERENCES.....	88
REFERENCES OF SAMPLES.....	93
APPENDIX.....	98

INTRODUCTION

We are on the edge of 4th Industry Revolution. This term “Industry-4.0”, initially was offered at the Hannover Exhibition in 2011. Klaus Schwab (Schwab, 2016) in his book “The 4th Industrial Revolution” highlighted current attributes of new coming era, such as overlapping of the real and virtual worlds, and interconnecting different technologies are, produced new decision and breakthroughs, machine learning and Artificial Intelligence (AI).

According to “The AI Index 2024 Annual Report” (Maslej et.al, 2024, p3) - “Progress accelerated in 2023. New state-of-the-art systems like GPT-4 is impressively multimodal: They can generate fluent text in dozens of languages, process audio, and even explain memes. As AI has improved, it has increasingly forced its way into our lives. Additionally, GPT-4 achieved an impressive 0.96 mean win rate score on the comprehensive Holistic Evaluation of Language Models (HELM) benchmark, which includes MMLU (Measuring Massive Multitask Language Understanding) among other evaluations”.

We will go into that later, but initially we cast our minds back to the history of Translation science which plays a crucial role in our life; it is especially important to underline its impact on interconnection between nations. The history of Translation traces to ancient times of Cicero (46 B.C) who believed that translation should be done by meaning “sense-for sense”, not by letters or words. Western world’s close connection to religion gave to start translation of canonical texts like Bible. We can take examples from the English language in 1550. The important event (the defeating Spanish Armada in 1558) gave the opportunity to many British merchants and pirates to explore new areas; of course, the translation was that instrumental tool for trade and relationship between countries. It should be also mentioned consecutive interpreting method, which was widely used in ancient times among emissaries and merchants from various cultures and across several ancient civilizations. More detailed discussion provided by Dr. Abbasbeyli Elvin, the first Azerbaijani to join the AIIC (International Association of Conference Interpreters), in his article (Abbasbeyli, 2021) "The role of the translation bureau (Tercümə Odası) in preparing translators, interpreters, and diplomats of the Ottoman Empire," highlights the historical significance of consecutive interpreting. The 19th and 20th centuries translation field was reanalyzed and formalized as an academic discipline which emerged as various schools of thoughts and divisions.

Delving into the period after World War 2, necessity of using simultaneous interpreting was put forward by the Nuremberg Trials. The trials' multilingual nature and the need for productive processes made traditional consecutive interpreting unpractical. The success of the process was achieved by simultaneous translating into 4 foreign languages realized by 50 interpreters, who firstly used IBM's International Translation System. That new method was the beginning of the era of simultaneous interpreting.

Turning into the structuring of the thesis, the followings should be mentioned:

Motivation of the theme. The current research sheds light on capabilities of an advanced language processing models such as ChatGPT-4 and ChatGPT-3.5 as compared to human translation. As we consider them assistant tools for translators, the comprehensive view of translations made by both newly appeared large language models (LLM) in 2022 and 2023, is very crucial. The analysis process, discussed in Chapter 3, investigates pros and cons of the models and their contribution to translation of scientific texts from English to Russian language.

Objects of investigation: The objects of the research are the language models (LLM) ChatGPT-4 and ChatGPT-3.5, based on deep learning system.

The subject of investigation: The subject of investigation is translation of scientific texts from English to Russian languages.

The aim and objectives: The main purpose of this thesis is to analyze the translation capabilities of large language models based on deep learning systems, to identify the percentage deviations between both models in translation procedures as well as with respect to human translations, and to examine strengths and weaknesses of both models and human translation in scientific texts.

Theoretical and pragmatic significance: The results/gains of the study will be useful for scientists and developers of the large language models ChatGPT-4 and ChatGPT-3.5, as well as not only translators and language professionals, but also for translation companies in terms of the strengths and limitations of AI models in translation of scientific texts. The findings will be helpful in enhancing research development in the field of AI models in furtherance.

Research questions:

1. Can ChatGPT-4 be used as a viable alternative to human translation in scientific texts?
2. To what extent can ChatGPT-4 accurately capture the nuances of bilingual translating scientific texts from English to Russian, including terms and conventionalized identities in comparison to human translation?
3. To what extent is the large language model ChatGPT-4 better in comparison to its predecessor, the ChatGPT-3.5?

Method of research: This research uses a mixed approach. Both qualitative and quantitative methods, complementing each other, were useful in providing the comprehensive result.

Structure of the thesis: The structure of the current thesis consists of the introduction, three chapters, conclusion and references. The first chapter describes the theoretical background of the study, the second chapter consists of the methodology, and the third chapter includes the analysis process and deals with metalinguistic analysis of the results, shedding the light on pedagogical implications of the gains as well as identifies the perspective aspects of the problem.

CHAPTER I. LITERATURE REVIEW

1.1 The history of Machine Translation

The history takes us to the year of 1646 in Leipzig (Germany), where Gottfried Wilhelm Leibniz was born. The scientist offered binary calculus system in 1685, which was described in his letters to De Remond (Лейбниц, 2005). He explained and found similarity of his system with the Chinese traditional math method of calculation. It was presented in the sequence from 0-9; when they reach 10, they continue to count as 10-in ten times as a hundred. Instead of this, he used only 0 and 1 symbols and nothing else, to calculate everything reducing the numbers by simple principle. This binary calculus system was a foundation of binary code. The scientist, also known as a philosopher, using the term of power related to substance, meant that substances can be as many as power, and the feature of such substance is the ability of action. He wrote that power equals to substance and called substance as “monada” (Greek word monos is one).

Leibniz wrote- “When there are disputes among persons, we can simply say, ‘Let us calculate’, and without further ado, see who is right.”

The binary calculus system theory, was applied in the development of computers. This binary code consists of two digits known as bits, and used in computers, allows to process, store, and transmit data.

EXPLICATION DE L'ARITHMETIQUE BINAIRE. QUI SE SERT DES SEULS CARACTÈRES 0 ET 1, AVEC DES REMARQUES SUR SON UTILITÉ, ET SUR CE QU'ELLE DONNE LE SENS DES ANCIENNES FIGURES CHINOISES DE FOHY¹

Le calcul ordinaire d'Arithmétique se fait suivant la progression de dix en dix. On se sert de dix caractères, qui sont 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, qui signifient zéro, un et les nombres suivans jusqu'à neuf inclusivement. Et puis allant à dix, on recommence, et on écrit dix par 10, et dix fois dix ou cent par 100, et dix fois cent ou mille par 1000, et dix fois mille par 10000, et ainsi de suite.

Mais au lieu de la progression de dix en dix, j'ai employé depuis plusieurs années la progression la plus simple de toutes, qui va de deux en deux, ayant trouvé qu'elle sert à la perfection de la science des Nombres. Ainsi je n'y employe point d'autres caractères que 0 et 1, et puis allant à deux, je recommence. C'est pourquoi deux s'écrit ici par 10, et deux fois deux ou quatre par 100, et deux fois quatre ou huit par 1000, et deux fois huit ou seize par 10000, et ainsi de suite. Voici la Table des Nombres de cette façon, qu'on peut continuer tant que l'on voudra.

On voit ici d'un coup d'oeil la raison d'une propriété célèbre de la progression Géométrique double en Nombres entiers, qui porte que si on n'a qu'un de ces nom-

0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	1
0	0	0	0	0	0	0	0	1	0	2
0	0	0	0	0	0	0	1	0	0	3
0	0	0	0	0	1	0	0	0	0	4
0	0	0	0	1	0	0	0	0	0	5
0	0	0	1	0	0	0	0	0	0	6
0	0	0	1	0	0	0	0	1	0	7
0	0	1	0	0	0	0	0	0	0	8
0	0	1	0	0	0	0	0	0	1	9
0	0	1	0	0	0	1	0	0	0	10
0	0	1	0	0	1	0	0	0	0	11
0	0	1	0	0	1	0	0	0	1	12
0	0	1	0	1	0	0	0	0	0	13
0	0	1	0	1	0	0	0	0	1	14
0	0	1	1	0	0	0	0	0	0	15
0	1	0	0	0	0	0	0	0	0	16
0	1	0	0	0	0	0	1	0	0	17
0	1	0	0	0	1	0	0	0	0	18
0	1	0	0	1	0	0	0	0	0	19
0	1	0	1	0	0	0	0	0	0	20
0	1	0	1	0	0	0	0	0	1	21
0	1	0	1	0	1	0	0	0	0	22
0	1	0	1	1	0	0	0	0	0	23
0	1	1	0	0	0	0	0	0	0	24
0	1	1	0	0	0	1	0	0	0	25
0	1	1	0	1	0	0	0	0	0	26
0	1	1	0	1	0	0	0	1	0	27
0	1	1	1	0	0	0	0	0	0	28
0	1	1	1	0	0	1	0	0	0	29
0	1	1	1	1	0	0	0	0	0	30
0	1	1	1	1	1	0	0	0	0	31
1	0	0	0	0	0	0	0	0	0	32

¹ Leibnizens mathematische Schriften. Herausgegeben von C. J. Gerhardt. Bd. VII. Abt. 2. Halle, 1863. S. 223 – 227.

Fig. 1

John Hutchins (Hutchins, 2003) describes the invention made by P. Trojanskij patented in 1933 as the ‘translating machine’, or (as the title of his 1933 patent describes it) - “a machine for selecting and typing words when translating from one language into another or several others simultaneously.”

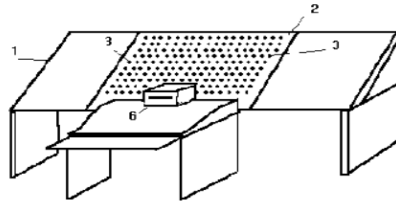


Fig. 2

Patent's description about the machine consisted of "a smooth sloping desk, over which moving easily and freely in different directions is belt provided with perforations which position the belt in front of an aperture" (see fig. 2)

This wide belt was a large dictionary, with entries in six languages in parallel columns. The operator located a word of the source language and moved the belt to display in the aperture the corresponding word of the target language. Then operator would type in a code indicating the grammatical category or role of the word in question – codes that Trojanskij referred to as 'signs for logical parsing' – and the combination of target word and code were then photographed onto a tape. Then the next source word would be located and 'translated' in the same way. From a tape of the target language words in sequence, a typist would then produce a 'coherent text' for a reviser to substitute the correct morphological forms for each word based on the assigned codes. As a final stage a 'literary editor' would produce the final target text. Unlike the operator and the reviser, who needed to know only their own languages (those of the source and the target respectively), the editor would need to know both languages in order, in Trojanskij's words, "to extract the meaning of the translation, to choose synonyms, to polish the unevenness, i.e. to do general literary finishing." (Hutchins, 2003).

Next development was described by Prof. Dr. C. Dianne Martin in her study (Martin, 1995), focusing on J. Mauchly's construction of an "electronic calculator" in 1942. It emerged from the use of high speed vacuum tubes and was funded by the government next year. The Project PX, the secret classification for ENIAC (the Electronic Numerical Integrator and Computer) began in June, 1943 until November, 1945. ENIAC weighed 30 tons, covered 1,500 square feet of floor space, used over 17,000 vacuum tubes and cost about \$500,000. Later, The ENIAC was used to provide calculations in weather forecasting, wind tunnel design for Army until 1955 in Aberdeen Proving Ground in Maryland.

In 1949, Warren Weaver in his memorandum of "Translation" (Weaver, 1949) underlined, "There is no need to do more than mention the obvious fact that a multiplicity of language impedes cultural

interchange between the peoples of the earth, and is a serious deterrent to international understanding. The present memorandum, assuming the validity and importance of this fact, contains some comments and suggestions bearing on the possibility of contributing at least something to the solution of the world-wide translation problem through the use of electronic computers of great capacity, flexibility, and speed”.

The Georgetown-IBM experiment was presented on January 7, 1954 in collaboration with Georgetown University and IBM corporation. The experiment included automated translation of 60 sentences from English to Russian. The system only based on 6 grammar rules and vocabulary, which consisted of 250 words in Russian language, “several short messages, within the 250-word range of the device, were tried.” (Hutchins, 2004).

In November 1966, Automatic Language Processing Advisory Committee (ALPAC,1966) published the estimation report for the research progress of machine translation (MT). It was regrettable. In fact, all financial assistance was stopped due to the factors of slow speed, quality and charge in comparison to human translators.

Despite all challenges, the machine translation experienced a rebirth in 1980, stimulated by advancement in computer technology. The growing internationalization of science in the 19th century and emergence of the Terminology discipline, which undergone several stages, led to the development of the second stage featuring crucial innovations such as the development of mainframe computers and documentation techniques (Teresa, 1998, p.6).

1975-1980 is marked by the dissemination of personal computers and in 1976 the European Communities accepted the operational system Systran (translated from Russian to English), developed by Peter Toma. This system replaced the IBM Mark II system and become one of the first MT marketable systems at that period. I. M. Pigott in his paper (Pigott,1986 p.144) wrote, “I am pleased to report that significant progress has been made with the result that in 1985 machine translation is no longer looked upon as an isolated application but has, for the majority of users...”. He underlined that Systran system has been developed with addition translations from English and French language into German and now supports 5 “language pairs”. At that time many various systems were introduced and occurred all over the world, for instance, Mitsubishi (MELTRAN), Hitachi(HICATS), the METAL German-English system (1988) used to translate documents of data processing and telecommunications ‘fields.

The statistics-based approach emerged in 1990 within the project “Candide” by an IBM team (Berger et al., 1994, pp.157-162). The Candide was a system for automatic translation (MT) from French to English and used methods of statistics and information theory, developing a probability model of translation process.

The method utilized mathematical models to predict the probability that of one sequence of words in a language appropriate to a sequence of another language. The accuracy of translated texts was only 48% but reduced the work of human translators by 60%.

The significant development was made by Google Translate(GT) in April 2006, introducing its own machine translation model. Google Translate started utilizing statistical machine translation but in November 2016 transitioned to the method named “Neural Machine Translation”.

1.2. Neural Networks

Parallel to development of machine translation the remarkable work was done in the period of 1940s-1960s called (Cybernetics), which was mentioned by emergence of neural networks conception. In 1943, Warren S. McCulloch and Walter Pitts (McCulloch & Pitts, 1943) published a description of the first neural network model but without using this term they noted in the abstract of the paper, - “Because of the “all-or-none” character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expressions satisfying certain conditions one can find a net behaving in the fashion it describes.”

The idea to describe how the brain’s basic processing element might function, they “showed how simple electrical circuits connecting groups of ‘linear threshold functions’ could compute various logical functions”. There were only 3 references at the end of that paper, without the main reference to Nicholas Rashevsky’s great influence in developing of neural networks. However, the Joint Research Centre (JRC) in 2018, mentioned his impact in the Science for Policy Report- “Mathematical models of neural networks were first developed by Nicolas Rashevsky in the early 1930s, and they became famous when his student Walter Pitts interpreted biological neural networks in 1942 as networks of logical switches.” (Tuomi I., Joint Research Centre (JRC) 2.2.1 Data-based neural AI, p10).

The American scientist Frank Rosenblatt mainly known as a father of deep learning, was inspired by the work of scientists Warren S. McCulloch and Walter Pitts, and created a neural network called –Perceptron (by meaning of the word perception). He was encouraged by possibility of creating similar model that can imitate a function of human brain’s neurons (human brain consists of 100 billion neurons) and how they can process the information. One of the key’s achievement was the learning algorithm, permitted to do tasks by samples. i.e. it could be trained on positive and negative answers. (Rosenblatt, 1958).

So how do perceptrons work? A perceptron takes several binary inputs, x_1, x_2, \dots , and produces a single binary output:

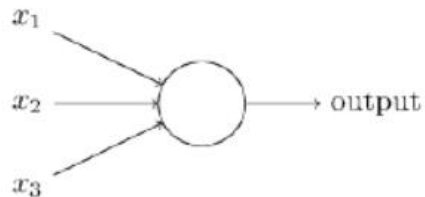


Fig.3

The perceptron has three inputs, x_1, x_2, x_3 . In general, it could have more or fewer inputs. Rosenblatt proposed a simple rule to compute the output. He introduced weights, w_1, w_2, \dots , real numbers expressing the importance of the respective inputs to the output. The neuron’s output, 0 or 1, is determined by whether the weighted sum $\sum w_j x_j$ is less than or greater than some threshold value. Just like the weights, the threshold is a real number which is a parameter of the neuron. (Nielsen, 2015).

$$\text{output} = \begin{cases} 0 & \text{if } \sum_j w_j x_j \leq \text{threshold} \\ 1 & \text{if } \sum_j w_j x_j > \text{threshold} \end{cases}$$

Fig.4

Weights played a crucial role in perceptron model and impact on result. The Perceptron consisted of a single layer in comparison with artificial neural networks, that consist of multiple connected layers.

The period of “AI winter” began with the tract, published by M. Minsky and S. Papert (Minsky & Papert, 1969, p.18) with negative judgements of Perceptrons, as they noted - “Perceptrons have been widely publicized as “pattern recognition” or “learning” machines and as such have been discussed in a large number of books, journal articles, and voluminous “reports.” Most of this

writing ... is without scientific value. We do not see that any good can come of experiments which pay no attention to limiting factors that will assert themselves as soon as the small model is scaled up to a usable size.”. Frank Rosenblatt couldn’t answer their critiques due to his sudden death in Chesapeake Bay on his boat trip.

In 1982 J. J. Hopfield introduced and developed recurrent neural network named the Hopfield network. (Hopfield,1982) The network consisted of neurons and the key ability of which was to remember patterns. So, Hopfield proposed a method for using a neural network as a Content-Addressable Memory (CAM). According to this type of network, where collected data could be saved in neurons’ memory researchers used it for automated cancer diagnosis and in medical image processing operations.

In 1997, Hochreiter and Schmidhuber (Hochreiter & Schmidhuber, 1997) developed the Long-Term Short Memory (LTSM). The Recurrent neural network could remember long term dependencies. All recurrent neural networks have the chain of repeating modules with a simple structure as single layer, but in LTSM the repeating module has an unequal structure with four layers. The key to LTSM is the cell state, where LTSM use capability to remove or add information.

Neural networks modify linear models in two important ways. The first is the use of multiple layers. Instead of computing the output value directly from the input values, a hidden layer is introduced. It is called hidden, due to inputs and outputs in training instances, but not the mechanism that connects them — this use of the concept hidden is similar to its meaning in hidden Markov models (Pickens & Batjargal, 2020).

Moving forward two decades to 2015, Nal Kalchbrenner et.al. (Kalchbrenner, et al.,2015) introduced the Grid Long Short-Term Memory network (Grid LSTM), which uses LSTM cells along all of the dimensions and modulates in a novel fashion the multi-way interaction. Researchers mentioned the advantages of the cells compared to regular connections in solving tasks such as parity, addition and memorization and ways of applying the model to character prediction, machine translation and image classification, showing strong performance across the board.

1.3. Deep Learning from 2006 to the present

Deep learning is a subset of machine learning, which in turn also is a subset of AI (Artificial Intelligence). AI is a technique that enables a machine to mimic human behavior. Deep learning systems in other words “learn” when its networks trained on a large data sets, for instance, after

learning the system can identify the needed object from images through thousand ones. It is like a many-layer pie referred to the term “deep”, to emphasize that the neural network able to train deeper, it has a structure of many layers each of which can handle the data and send its output to the next layer. Such a very useful process, due to the final result, taken from system, is précised.

An American scientist Marvin Minsky, known as a pioneer of development and father of AI, worked on vision system in 1966. The project consisted of connecting computer to a camera and asking to describe “what it saw”, but the task was too much complex at that time.

The modern era of deep learning began in 2006 when Geoffrey Hinton (Hinton, et al., 2006) developed a hybrid model, where the top two hidden layers form an undirected associative memory and the other hidden layers form a directed acyclic graph that converts the representations in the associative memory into observable variables such as the pixels of an image. This hybrid model is notable for its greedy learning algorithm that can find a fairly good set of parameters quickly, even in complex deep networks with millions of parameters and many hidden layers, the learning algorithm is unsupervised but adapts well to labeled data by learning a model, generating both the label and the data. The generative model makes it easy to interpret the distributed representations in the deep hidden layers. The inference required for forming a percept is both fast and accurate.

Researchers demonstrated the possibility of learning a deep, densely-connected, belief network one layer at a time. The obvious way to do this is to assume that the higher layers do not exist when learning the lower layers, but this is not compatible with the use of simple factorial approximations to replace the intractable posterior distribution. This discovery, along with the increasing availability of large data sets and powerful GPU (Graphics Processing Units), led to major breakthroughs.

By 2012, Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton (Krizhevsky, et al., 2012) made a phenomenal result with their AlexNet model, which drastically outperformed other competitors in the ImageNet competition, reducing the error rate by a substantial margin.

Nowadays, the deep learning systems are improved significantly, we can see their influence in various fields, such as in autonomous driving, natural language processing, and medical diagnosis.

1.4 ChatGPT-4 and ChatGPT-3.5 models

The book “The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma” written by Mustafa Suleyman and Michael Bhaskar (CO Founder of Deepmind and Inflection AI) in September 2023 was bestseller. In his book he confirms the fact of “the coming wave, defined by two core technologies: artificial intelligence (AI) and synthetic biology. Together these fields usher in a new dawn for humanity, creating wealth and surplus unlike anything ever seen.” (Mustafa Suleyman & Bhaskar, 2023, p.22).

During the “coming wave” that increased scientists’ interest in the AI field in the 21st century, OpenAI Inc. was founded in December 2015 by Samuel Harris Altman, an American investor, entrepreneur, and prominent figure (referred to as the “golden brain”), and Elon Musk, a visionary known for finding solutions in force majeure situations. Elon Musk is the Founder, CEO, and Chief Engineer of SpaceX, CEO and Product Architect of Tesla, Inc., Owner, CTO, and Executive Chairman of X (formerly Twitter), President of the Musk Foundation, and Founder of the Boring Company, X Corp., and xAI. He is also the co-founder of Neuralink, OpenAI, Zip2, and X.com (part of PayPal). The founding team of OpenAI included notable figures such as Sam Altman, Greg Brockman, Ilya Sutskever, Wojciech Zaremba, and John Schulman. ChatGPT-3.5 is a language model based on deep learning system and trained on a huge text data was launched in 2022, by OpenAI Inc. In just 5 days of its launch, the model reached approximately 1 million users.

Initially, work was conducted on that model, which was included in the thesis to compare its capabilities with the model ChatGPT-4. The excitement led to using it in May 2023 with the idea of making a comparison in translation between a pair of languages: English and Russian. ChatGPT-3.5 provided significant assistance in translation, especially in overcoming challenges due to the lack of cultural knowledge in English. For instance, the sentence containing the phrase 'the kontras were saints' was translated (see Fig. 5), and the language model provided information about the meaning of the phrase.

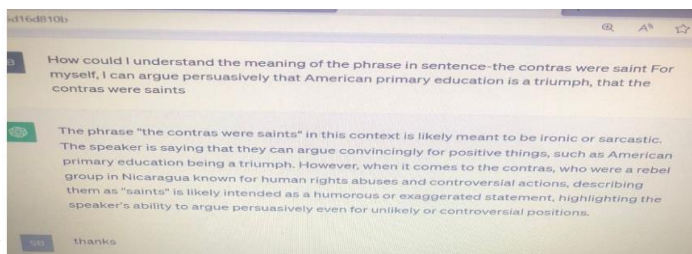


Fig.5

Additionally, the model was asked to help in creating a new translation app.

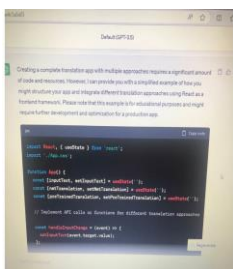


Fig.6

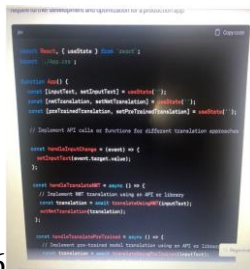


Fig.7



Fig.8

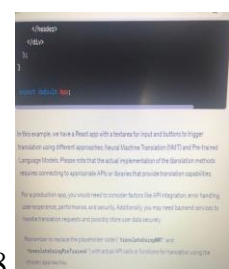


Fig.9

As an assistant tool, it is very user-friendly, can generate texts on various topics and in different fields and it is, in fact, best interlocutor. Limitations: incorrect answers, called “hallucination, the model’s information of events occurred only until January, 2022; couldn’t translate low-resource languages well. For instance, the idiom- “It is raining cats and dogs” translated to Azerbaijani language as “Göydən pişiklər və itlər tökülür”.

So, what is the actual meaning of abbreviations “GPT”? In a simple way here an explanation, for G: generative-predicting next word; Pre-trained: previously trained on a large amount of data, T-Transformer: Encoder-Decoder based neural network. It should be mentioned that we should use the specific prompts for both models, if you give an appropriate prompts and correctly explain what you need, model will give you more precise answers.

In the Preliminary Study “Is ChatGPT a good translator?”, Wenxiang Jiao (Jiao et al., 2023) focused on three aspects: if the style of prompts may affect the quality of translation; how does it provide the multilingual translation on various language pairs (e.g., high vs. low) and language family (e.g., European vs. Asian); can it perform translation accuracy well in domain-specific or even noisy sentences (including grammar mistakes, jargon, complex constructions that make text difficult to read or understand). By selecting four languages: German, English, Romanian and Chinese, the researchers additionally used BLEU score metric. In evaluation process, Wenxiang Jiao et al., found that ChatGPT worked competitively with commercial translation products (Google Translate and DeepL translate) on high-resource European Languages-German-English and English-German, but lags behind significantly on low-resource languages. Researchers explored a strategy named pivot prompting asking ChatGPT to translate the source sentence into a high-resource pivot language before into the target language, which improved significantly the translation performance. As for the translation accuracy, ChatGPT did not perform as well as the commercial systems but is potentially a good translator for spoken language.

Tom Kocmi and Christian Federmann in the study “Large Language Models Are State-of-the-Art Evaluators of Translation Quality” (Kocmi & Federmann, 2023), experimented with four prompt templates and evaluated seven various models of GPT, showing that only GPT-3.5 and larger models are capable of translation quality assessment. By applying the GEMBA metric which was considered not reliable enough, but only for system-level evaluation.

Sai Cheong Siu in his study “ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation” (Siu, 2023) noted that ChatGPT can clearly explain the meanings of various textual elements, including words, phrases, and clauses, by taking into account their context within a sentence or paragraph. This aspect is particularly useful when dealing with words that have multiple meanings, which can pose challenges for translators. Additionally, Sai Cheong Siu mentioned the ability to offer explanations that consider cultural and historical contexts, which is especially beneficial when dealing with culture-specific expressions, idioms, or phrases that may not be easily understood by individuals from different cultural backgrounds. For consideration, the scholar provided some examples to assess the ability to clarify technical terms, paraphrase the academic language into simple one, that helps the translator gain a better understanding of the source text, ensuring accurate conveyance of meaning in the target language. Similarly, he underlined, when given a medical document, the chatbot can interpret medical terminology and offer layperson-friendly explanations. Another key feature of ChatGPT is an ability to generate automat translation. For instance, Sai Cheong Siu showed an example by using prompt for GPT “Translate the following into [target language]”, which immediately was done by GPT providing several draft suggestions for choice. The model also helps to detect mistakes, provides quality of translation, maintaining grammatical structure of sentences, gives numeral synonyms of the given words.

For considerable advantages, the scientist claimed the speed and cost, giving as an example; it took 8 seconds for Chinese-English translation by GPT-3.5.

However, as mentioned by Bang (Bang et al., 2023) in “A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity”

ChatGPT suffers from hallucination problems like other LLMs. It generates more extrinsic hallucinations – factual statements that cannot be verified from the source, in other words, creating materials that is factually wrong or untrue).

Nick McKenna, Tianyi Li et al., (McKenna et al., 2023) in the study “Sources of Hallucination by Large Language Models on Inference Tasks” established two biases originating from pretraining which predict much of their behavior, and showed that these are major sources of hallucination in generative LLMs. First, memorization at the level of sentences: regardless of the premise, models falsely label NLI test samples as entailing when the hypothesis is attested in training data, and that entities are used as “indices” to access the memorized data. Second, statistical patterns of usage learned at the level of corpora: showed a similar effect when the premise predicate is less frequent than that of the hypothesis in the training data. Researchers underlined that the further exploration should be done in future work.

Sukhpal Singh Gill and Rupinder Kaur examined foundations, vision, research challenges and advantages of ChatGPT (Gill & Kaur,2023). Researchers noted the significant opportunities coming with ChatGPT models, which technologies can alter and improve our lifestyles. However, there are many problems that have to be solved such as improvement of ability to remain consistent during the conversations, to enhance the capability of model to understand and retain long extracts of content, to correct the producing unreliable contents and answers as well as lack of knowledge’s comprehensiveness of specific fields.

Faten Khoshafah in the article “ChatGPT for Arabic-English Translation: Evaluating the Accuracy” (Khoshafah, 2023) provided the translation analysis of the texts from English to Arabic languages and explored the capabilities of ChatGPT without specifying which of the models were explored. The researcher recommended consulting with professional translators and not relying solely on translations made by GPT models, also suggested improving the AI models, by using more precise legal terminology, improving phrasing, and eliminating redundancy.

Çağrı Işım, Dr. Yavuz Selim Balcıoğlu, in their study “ChatGPT: Performance of translate” (Işım & Balcıoğlu,2023) measured the translation performance of AI GPTmodel from English to Turkish. They selected a total of 50 Turkish paragraphs from the field of education for analysis. The researchers determined the successful performance as 68% on the translation point with 34 out of 50 paragraphs were translated without error. They underlined the success in simple translations, but noted that the output produced by the model needs to be seriously controlled.

Long Ouyang, Jeff Wu et.al, developed and improved a language model by training it with examples of advisable responses and then refining it further based on human evaluations of its

outputs. They developed InstructGPT, that provides more truthful responses, fine-tuned with human feedback in order to improve its performance (Ouyang, et.al, 2022).

A comprehensive survey was provided by Yiheng Liu, Tianle Han (Liu, et.al, 2023) and described the capabilities of ChatGPT in various fields. The researchers highlighted competitiveness with other commercial translation products on high-resources European languages, but noted a decline in performance with distant languages. The integration of Reinforcement Learning from Human Feedback (RLHF) has further enhanced the models' performance, making them highly efficient in processing natural language.

ChatGPT-4 released in March, 2023 and became available for premium users. Additionally, it generates images by integration of model Dalle-E3 into ChatGPT-4, noticeable improvement in translation as well. In January 2024, the OpenAI company kindly offered the builder system, which helps to create your own chatbot with further subsequent placement at the GPT store. The GPT store is a marketplace for customers and an excellent opportunity to promote own advanced GPT agent, offering specific functions.

We have created GPT Agent called "AIVerse". This user-friendly translation assistant chatbot includes 4 languages: English, Russian, Azerbaijani and Hebrew, a huge dataset of formal academic language dictionaries, cultural nuances of languages, multiple idiomatic expressions. Using prompts, the chatbot can translate from screenshots, images and pictures (consisted of words), does multiple translation in one click and working on translation with subtitles for short cartoons or videos. There are several reasons by creating the specific assistant tool. Firstly, to help Azerbaijani students quickly get and translate from English to Azerbaijani and vice versa, to learn other languages as well as idioms. Secondly, we live in a multicultural country and we have Jewish and Russian communities, offering to learn Azerbaijani language by using this chatbot we may achieve mutual cooperation between all nations of our population and even among tourists, annually visiting our homeland. It also creates an impact on cooperation and friendship between Azerbaijan Republic and the other countries. Finally, this is potential development for our future generation.

1.5 Metrics

Human evaluation of translation have been always considered as the gold standard, but besides being an expensive one, it requires significant human labor. Automated evaluations are less expensive and more easily used for quantitative analysis.

Automatic evaluation metrics are evolving better mimic human assessment by incorporating advanced machine learning techniques, including models based on deep learning models like ChatGPT. Automatic metrics for Machine Translation (MT) are considered as a core component in MT research and advancement. These metrics enable the fast assessment and comparison of MT systems on a huge compilations of texts.

These metrics mainly fall into several categories: lexical-based, embedding-based, and supervised (neural fine-tuned) metrics.

The first well-known evaluation metric for machine translation, BLEU (Bilingual Evaluation Understudy) was introduced in 2002 (Papineni, et.al.,2002), BLEU was developed by researchers at IBM and quickly became a standard metric for evaluating the quality of text translated by machine translation systems. It compares the source translation text to one or more reference translations, measuring the precision of matched sequences of words or phrases. BLEU was pioneering, provided a quantitative way to assess the quality of machine-translated text, allowing for more systematic and reproducible evaluation across different translation systems.

However, two decades later, according to Markus Freitag in the study “The Results of WMT22 Metrics Shared Task: Stop Using BLEU – Neural Metrics Are Better and More Robust” (Freitag et al., 2022, pp. 46 – 68) the results demonstrated the domination of neural-based learned metrics and proof that traditional metrics like BLEU, SPBLEU, or CHRF poorly correlate with human ratings. (see Fig.10)

Metric	avg rank
METRIX XXL	1.20
COMET-22	1.32
UNITE	1.86
BLEURT-20	1.91
COMET-20	2.36
MATESE	2.57
COMETKIWI*	2.70
MS-COMET-22	2.84
UNITE-SRC*	3.03
YISI-1	3.27
COMET-QE*	3.33
MATESE-QE*	3.85
MEE4	3.87
BERTSCORE	3.88
MS-COMET-QE-22*	4.06
CHRF	4.70
F101SPBLEU	4.97
HWTSC-TEACHER-SIM*	5.17
BLEU	5.31
REUSE*	6.69

Fig.10

According to the final opinion of researchers, neural-based machine translation evaluation metrics are significantly better than non-neural metrics. We have opted the neural-based machine translation evaluation metrics COMET-22 and BLEURT-20 for comparative analysis of 30 scientific articles, assessing the efficiency and capability of GPT models, and gaining more precise quantitative result as well.

Ricardo Rei proposed (Ricardo Rei, et.al, 2022) COMET-22 – is an ensemble between a COMET estimator model trained with Direct Assessments and a newly offered multitask model trained to predict sentence-level scores along with word-level tags. These models are ensemble together using a hyper-parameter search that weights different features extracted from both evaluation models and combines them into a single score. The researcher underlined that this model architecture can be used with and without access to a reference translation.

The neural fine-tuned metrics have capability to use large multilingual data during training to better compare and evaluate the quality of state-of-the-art MT models by better outperforming lexical-based metrics.

BLEURT-20- (Sellam et al., 2020), was presented as a learnt reference-based text generation metric for Natural Language Generation. It is built using multiple phases of transfer learning starting from a Pre-trained model BERT (Delvin et al., 2018). Due to the fact that metric is trained end-to-end, BLEURT can model human assessment with superior accuracy.

The process of estimation on both metrics consists of uploading needed texts into Python (programming language), and dividing them on paragraphs; then we group them according to translations i.e. ST (original text in English), TT in Russian language (human translation, model of ChatGPT-3.5 and finally ChatGPT-4). As a result, we have 4 texts. The process (tokenization) continues by splitting texts into sentences and transferring them into downloaded from Hugging Face (Hugging Face Platform) COMET-22 and BLEURT-20 metrics, where process of evaluation starts. (see Fig.11).

```
4984 50 238 00 13 (8%)
[[{'src': 'The concept of 'Three Poles of the Cold' (or simply the Three Poles) has not yet received an unambiguous interpretation among the students of political science.',
  'mt': 'Концепция "Трёх полюсов холода" (или просто Трёх полюсов) до сих пор не получила однозначной интерпретации среди студентов политологии.',
  'ref': 'Понятие «Три полюса холода» (или «Три полюса») пока не получило однозначной трактовки в мировом политическом и научном сообществах.'},
 {'src': 'For some, mainly climatologists and glaciologists from the Greater Himalayan region, this is all about only one Pole of the Cold, i.e.',
  'mt': 'Для некоторых, в основном климатологов и гляциологов из региона Великого Гималаевского хребта, это связано только с одним Полюсом Холода, то есть Гималаями.',
  'ref': 'Для одних, например, для климатологов и гляциологов из стран «Большого гималайского региона», речь идёт только об одном полюсе холода – Гималаях.'}]
Котировки:
https://www.bleurt.com/ru/ru-ru-bleurt.html
```

Fig.11

It should be underlined, that COMET-22 neural fine-tuned metric is a more powerful metric, trained on a wider and more diverse dataset and focusing on the comprehensive semantic similarity between source text and translation. As a result, it is less sensitive to minor errors or deviations in translation, than the indicators of BLEURT-20.

CHAPTER II. METHODOLOGY AND DATA ANALYSIS

This chapter explores the research methodology employed in the current thesis. We investigate the translation of scientific texts, assessing the efficiency and capability of GPT models and comparing them with human translation.

According to Eugene Nida (Nida, 1964), “There are no identical languages, either in the meanings attributed to corresponding symbols or in the manner in which these symbols are arranged in phrases and sentences, which means that absolute correspondence between languages is impossible”. However, the precise and correct translation into the target text(TT) originates in deep knowledge of both the source language (SL) and target language (TL). Accurate translation is achieved through analysis, interpretation, consideration of cultural and thematic nuances, and adaptation, ensuring that the tone, meaning and intent of the source text (ST) are faithfully conveyed into target language.

As we applied the translation criticism in this research as a fundamental and core method in the comparison of the texts, it is important to mention the influential work of Holmes (1979), who identified translation criticism as a branch of applied translation studies that deals with assessing translated texts.

2.1 Data Gathering Instruments of the Study

Some data collection instruments were applied to obtain appropriate data and address the research questions raised in the current thesis. In order to conduct comprehensive research by applying qualitative and quantitative methods, data from thirty examples in both source texts and target texts thoroughly searched on reputable websites and in scientific journals. For quantitative data we used Python (programming language) and opted newly neural fine-tuned metrics: COMET-22 and BLEURT-20, in order to estimate ratios. Our choice fell on scientific journal Acta Naturae (Acta Naturae, 2009), as it was the most appropriate source, consisting of an archive of scientific papers in both English and Russian languages.

2.2 Corpus and Preparing the Data for Analysis

A collection of scientific texts, consisting of different themes such as historical, medical, and biological was thoroughly selected. The texts were download from a free resource archive of the

online scientific journal *Acta Naturae* and divided into extracts, preparing them for the translation process of GPT models. Finally, thirty examples of short extracts, each contained at least 650 words, were carefully analyzed to ensure a comprehensive representation and prepared for the detailed and rigorous examination of the translation capabilities of the language models in this study.

2.3 The Procedure of Data Analysis

By applying mixed-method, the procedure of data analysis was complex and took a considerable amount of time. The large dataset of 30 scientific texts was initially divided in small extracts and numbered for quantitative method research. Source texts were copied and paste into GPT models in turns by using specific prompts like- “Translate into Russian Language correctly or do accurate translation”. The extracts were meticulously divided into paragraphs, and then were grouped according to translations: ST (original text in English), TT in Russian language (human translation, translation by the model of ChatGPT-3.5 and finally ChatGPT-4). As a result, we had 4 types of texts for each sample. For the process of estimation, which included graphics and tables, we uploaded the necessary texts into Python (programing language), split the texts into sentences and transferred them into metrics downloaded from Hugging Face (Hugging Face Platform).

The process of qualitative research involved deeply analyzing all extracts of each sample through translation criticism, including the use of English-Russian Explanatory Scientific and Technical Dictionaries as (В.И. Кочергин, 2008), (Э. Пройдаков, Л. Теплицкий, 2021) and (Михаил Циммерман, Клавдия Веденева, 2003). The analysis focused on readability, accuracy, and fidelity. We analyzed all sentences in the translated extracts, identifying numerous mismatches in terms, repetition by the AI models, non-equivalent and misinterpreted idiomatic expressions, and the avoidance of grammatical rules of gender agreement in the sentences in both human translations and GPT models equally. This analysis allowed us to determine the gaps in translation between source texts and target texts, ensuring high quality, precision and adherence to translation standards.

CHAPTER III. DISCUSSION AND RESULTS

This chapter provides a deep analysis and discussions of all considered examples of translations made by human translators and the language models ChatGPT-3.5 and ChatGPT-4, using a mixed-method empirical approach.

We have analyzed 30 scientific articles in English language (SL) and equivalent translations (TL) in Russian language, translated by individuals, then compared with translations made by ChatGPT-3.5 and ChatGPT-4. The translation criticism was applied in current qualitative method by analyzing short extracts from scientific articles.

1st Sample

Resource Text

1. This section “Forum” is about the pharmaceutical industry in Russia. We were encouraged to debate this topic after the unveiling of the Strategy of Development of the Pharmaceutical Industry in the Russian Federation developed by the Ministry of Industry and Trade of the Russian Federation.

2. Without the development and use of technologies, without possession of a certain amount of them, without the infrastructure for permanent innovation in this area, it is impossible to achieve a serious improvement in the quality of our health care sector.

Target Text made by human translators

1. Нынешний раздел «Форум» посвящен фармацевтической промышленности России – к этой теме нас подтолкнуло появление и последующее активное обсуждение Стратегии развития фармацевтической промышленности Российской Федерации до 2020 г., разработанной по заказу Министерства промышленности и торговли Российской Федерации.

2. Не создавая и не используя высокие технологии, не владея достаточным их количеством, не владея инфраструктурой для постоянного воспроизводства инноваций в этой области нельзя добиться серьезного улучшения качества нашего здравоохранения.

ChatGPT-3.5

1. Этот раздел "Форум" посвящен фармацевтической индустрии в России. **Мы были призваны** обсудить эту тему после представления Стратегии развития фармацевтической промышленности Российской Федерации, разработанной Министерством промышленности и торговли Российской Федерации.

2. Без развития и использования технологий, без наличия определенного их количества, без инфраструктуры для постоянных инноваций в этой области невозможно серьезное улучшение качества нашей здравоохранительной системы.

ChatGPT-4

1. Этот раздел "Форум" посвящен фармацевтической промышленности в России. **Нас побудили** обсудить эту тему после того, как была представлена Стратегия развития фармацевтической промышленности Российской Федерации, разработанная Министерством промышленности и торговли Российской Федерации.

2. Без развития и использования технологий, **без владения определенным количеством их**, без инфраструктуры для постоянных инноваций в этой области невозможно добиться серьезного улучшения качества нашего здравоохранения.

Discussion-Considering the given extract of the scientific article, the 1st sentence by human translation was done using explicitation technique, in ST we didn't mention the collocation (the strategy up to 2020). However, it may be used to emphasize and reflect the concise knowledge of the strategy and appropriate for specific auditory. The phrase "**We were encouraged to debate this topic**" was misinterpreted by the both models; by ChatGPT-3.5 (**Мы были призваны**) and ChatGPT-4 (**нас побудили обсудить**) as well.

The 2nd sentence translated by human was incorrect, in translating the meaning word- development. Instead of не развивая or без развития, the translator made interpretation- "without creating or using high technologies". Adding the word **high** and gerundive construction, we consider the human translation inappropriate as for the translation by ChatGPT-4, it is incorrect for the phrase **без владения определенным количеством их** and ignoring definite final elements of the sentence.

2nd Sample

Resource Text

The concept of ‘Three Poles of the Cold’ (or simply the Three Poles) has not yet received an unambiguous interpretation among the students of political science. For some, mainly climatologists and glaciologists from the Greater Himalayan region, this is all about only one Pole of the Cold, i.e. the Himalayas. These scientists are mainly interested in the factors that cause the melting of glaciers in this mountain range, as well as in the climatic and socio-economic consequences that the process has for Asian nations. For others, it is important to establish the relationship between the natural and climatic processes occurring in the Arctic, the Antarctic and the Himalayas to arrive at a global climate model. For international diplomats and political scientists, it is more interesting to see how the positive experience of scientific and environmental cooperation gained in the Arctic and Antarctic can be applied to the Himalayas and Tibet in solving the existing bi- and multilateral problems in this region.

Target Text made by human translators

Понятие «Три полюса холода» (или «Три полюса») пока не получило однозначной трактовки **в мировом политическом и научном сообществах**. Для одних, например, для климатологов и гляциологов из стран «Большого гималайского региона», речь идёт только об одном полюсе холода — Гималаях. Этим учёным интересуют в основном факторы, влияющие на таяние ледников в этом горном массиве, а также климатические и социально-экономические последствия этого процесса для азиатских стран. Для других представителей естественных наук важно установить взаимосвязь между природно-климатическими процессами, происходящими в Арктике, Антарктике и Гималаях, для того чтобы создать глобальную климатическую модель. Для дипломатов и политологов-международников интересно посмотреть, как позитивный опыт сотрудничества в научной области и в сфере охраны окружающей среды, накопленный в Арктике и Антарктике, может быть применён в Гималаях и Тибете и полезен для решения имеющихся дву- и многосторонних проблем в этом регионе.

ChatGPT-3.5

Концепция "Трёх полюсов холода" (или просто Трёх полюсов) до сих пор не получила однозначной интерпретации среди студентов политологии. Для некоторых, в основном климатологов и гляциологов из региона **Великого Гималаевского** хребта, это связано только с одним Полюсом Холода, то есть Гималаями. Эти ученые в основном интересуются

факторами, вызывающими таяние ледников в этом горном массиве, а также климатическими и социальноэкономическими последствиями, **которые этот процесс имеет** для азиатских наций. Для других важно установить связь между природными и климатическими процессами, происходящими в Арктике, Антарктике и Гималаях, чтобы прийти к глобальной климатической модели. Для международных дипломатов и политологов более интересно увидеть, как положительный опыт научного и экологического сотрудничества, накопленный в Арктике и Антарктике, может быть применен к Гималаям и Тибету для решения существующих **би- и мультилатеральных проблем** в этом регионе.

ChatGPT-4

Концепция "Трех полюсов холода" (или просто Три Полюса) пока не получила однозначного толкования среди политологов. Для некоторых, в основном климатологов и гляциологов из региона Большого Гималая, речь идет только об одном Полюсе Холода, то есть о Гималаях. Эти ученые в основном заинтересованы в факторах, вызывающих таяние ледников в этом горном массиве, а также в климатических и социо-экономических последствиях этого процесса для азиатских народов. Для других важно установить связь между природными и климатическими процессами, происходящими в Арктике, Антарктике и Гималаях, чтобы прийти к глобальной модели климата. Для международных дипломатов и политологов более интересно увидеть, как можно применить положительный опыт научного и экологического сотрудничества, полученный в Арктике и Антарктике, к Гималаям и Тибету при решении существующих двусторонних и многосторонних проблем в этом регионе.

Discussion- at the given extract the phrase **among the students of political science** translated by human as **в мировом политическом и научном сообществах**, applying generalization technique in order to cover an expanded auditory and to suggest rising a wide interest to the article. However, the specific meaning of the sentence might be lost. The ChatGPT-3.5 misinterpreted the term of Greater Himalayan region **for Великого Гималаевского хребта**, as well as making incorrect translation as word-by-word for phrase- that the process has for Asian nations- **которые этот процесс имеет** для азиатских наций and bi- and multilateral problems –as **би- и мультилатеральных проблем**. The translation by ChatGPT-4 was precise.

3rd Sample

Source text

The treatment of MDs should be initiated as soon as possible, but the diagnosis of mitopathologies is extremely difficult because of their heterogeneity and overlapping clinical features. The molecular pathogenesis of mitochondrial diseases is investigated using animal models: i.e. animals carrying mutations causing MD symptoms in humans. The use of mutant animal models opens new opportunities in the study of genes encoding mitochondrial proteins, as well as the molecular mechanisms of mitopathology development, which is necessary for improving diagnosis and developing approaches to drug therapy.

Target Text made by human translators

Терапию митохондриальных болезней необходимо начинать как можно раньше, однако их диагностика затрудняется гетерогенностью заболеваний и перекрыванием спектров их клинических проявлений. Для понимания молекулярного патогенеза митохондриальных **болезней создают животные модели**, т.е. животных с мутациями, проявления которых напоминают симптомы митопатологий человека. Использование животных моделей открывает новые возможности в исследовании функций генов, кодирующих митохондриальные белки, молекулярных механизмов возникновения и развития митопатологий, что необходимо для усовершенствования диагностики и разработки подходов к лекарственной терапии.

ChatGPT-3.5

Лечение МЗ должно быть начато как можно скорее, однако диагностика митопатологий крайне затруднительна из-за их гетерогенности и перекрывающихся клинических проявлений. **Молекулярная патогенез** митохондриальных заболеваний изучается с использованием **животных моделей**: то есть животных с мутациями, вызывающими симптомы МЗ у людей. Использование мутированных животных моделей открывает новые возможности для изучения генов, кодирующих митохондриальные белки, а также молекулярных механизмов развития митопатологий, что необходимо для улучшения диагностики и разработки подходов к лекарственной терапии.

ChatGPT-4

Лечение МЗ следует начинать как можно скорее, однако диагностика митопатологий чрезвычайно сложна из-за их гетерогенности и совпадающих клинических признаков.

Молекулярная патогенез митохондриальных заболеваний исследуется на моделях животных, то есть на животных, несущих мутации, вызывающие симптомы МЗ у людей. Использование мутантных **животных моделей** открывает новые возможности в изучении генов, кодирующих митохондриальные белки, а также молекулярных механизмов развития митопатологии, что необходимо для улучшения диагностики и разработки подходов к лекарственной терапии.

Discussion- The 2nd sentence consisted of the phrase-is **investigated using animal models** was interpreted as **создают животные модели** by translator, however more concise translation could be interpreted as made by language models-изучается/исследуется. Human translator uses literal translation and equivalence techniques to ensure accuracy while restructuring sentences slightly for better readability in Russian. The both of GPT models failed in translation of the term- **Молекулярная патогенез** not following the appropriate, established grammatical rule of gender agreement in Russian language. Similarly, the words' combination - **животные модели** as mentioned in 3 target texts, might be translated more precise as-**модели животных.**

4th Sample

Source text

This review aims to assess the current state of soft X-ray microscopy and its impact on our understanding of the subcellular organization. It also attempts to look into the future of X-ray microscopy, particularly as relates to its seamless integration into the cell biology toolkit.

Target Text made by human translators

Настоящий обзор ставит своей целью **рассмотреть вклад и вычлениТЬ принципиально** новые возможности, которые уже дала биологическая микроскопия в мягком рентгеновском диапазоне, проанализировать особенности этого метода и оценить потенциальные направления его дальнейшего развития применительно к клеточно-биологическим исследованиям.

ChatGPT-3.5

Этот обзор направлен на оценку текущего состояния мягкой рентгеновской микроскопии и ее влияния на наше понимание субклеточной организации. **Он также пытается взглянуть**

в будущее рентгеновской микроскопии, особенно в отношении ее **безпроблемной интеграции** в инструментарий клеточной биологии.

ChatGPT-4

Этот обзор направлен на оценку текущего состояния мягкой рентгеновской микроскопии и ее влияния на наше понимание субклеточной организации. Также предпринимается попытка **взглянуть** в будущее рентгеновской микроскопии, особенно в отношении ее беспроблемной интеграции в инструментарий биологии клетки.

Discussion-there is an inadequately translation made by human translator and a significant deviation from the source text, losing the meaning of entire text. In fact, the word - seamless (беспрепятственная, плавная) and whole term of seamless integration was omitted. The translation changes the point of view slightly to fit Russian idiomatic expressions better, using modulation technique. For example, "its seamless integration into the cell biology toolkit" is translated as "применительно к клеточно-биологическим исследованиям" (applicable to cell biological research), which shifts the focus slightly but maintains the overall meaning. The translation primarily employs foreignization, preserving the scientific terminology and detailed descriptions. It uses a combination of literal translation, explicitation, transposition, and modulation to ensure the text is both accurate and readable in Russian. More precise and equivalent translations were made by language models, though the expression-**пытается взглянуть/попытка взглянуть** might be translated more correctly as- попытка заглянуть; the expression-**безпроблемной интеграции** with making mistake in the letter also could be translated as **плавная интеграция.**

5th Sample

Source Text-Acetylcholine (ACh) is one of the main neurotransmitters and modulators of the nervous system. ACh receptors – nicotinic (ionotropic) and muscarinic (metabotropic) ones – are expressed in a variety of tissues, from neuromuscular junctions and the parasympathetic nervous system to the cortical areas involved in cognitive functions such as learning and memory. Cholinergic agents, including allosteric modulators, are actively used in the treatment of various pathologies.

Target Text made by human translators

Ацетилхолин (АХ) – это один из основных нейромедиаторов и модуляторов нервной системы. Никотиновые (ионотропные) и мускариновые (метаботропные) холинорецепторы (ХР) экспрессируются в различных тканях: от нервно-мышечных синапсов (НМС) и парасимпатической нервной системы до областей коры, участвующих в когнитивных функциях, таких, как обучение и память, а холинергические агенты, в том числе аллостерические модуляторы, активно используются в лечении разного рода патологий.

ChatGPT-3.5

Ацетилхолин (АCh) является одним из основных нейромедиаторов и модуляторов нервной системы. Рецепторы АCh - никотиновые (ионотропные) и мускариновые (метаботропные) - экспрессируются в различных тканях, начиная от нейромышечных соединений и парасимпатической нервной системы и заканчивая кортикальными областями, участвующими в когнитивных функциях, таких как обучение и память. Холинергические агенты, включая аллостерические модуляторы, активно используются в лечении различных патологий.

Chatgpt4

Ацетилхолин (АCh) является одним из основных нейромедиаторов и модуляторов нервной системы. Рецепторы АCh – никотиновые (ионотропные) и мускариновые (метаботропные) – выражены в разнообразии тканей, от нейромышечных соединений и парасимпатической нервной системы до корковых областей, участвующих в когнитивных функциях, таких как обучение и память. Холинергические агенты, включая аллостерические модуляторы, активно используются в лечении различных патологий.

Discussion- In the given extract, the term "the cortical areas" relates to the equivalent meaning of "области коры головного мозга" in Russian, which was not adequately clarified by the human translator. Additionally, human translation employs calque technique, in such terms like "Acetylcholine (ACh)" to "Ацетилхолин (АХ)" and "nicotinic (ionotropic) and muscarinic (metabotropic) receptors" to "никотиновые (ионотропные) и мускариновые (метаботропные) холинорецепторы". The term "expressed in a variety of tissues" was translated by ChatGPT-4 as "в разнообразии тканей," which is not considered a correct expression in scientific texts. The correct translation should be "в разных/различных тканях."

6th Sample

Source text

1. The existence of a “blind spot” in the study of the genomes of the Medieval Russian nobility in many aspects is due to the complexity of personally identifying the remains of the Rurikids and other aristocratic families in the necropolises of the X–XIV centuries. It is well known that the names of the buried were not indicated in any way on funerary structures, sarcophagi, or tombstones until the beginning of the XV century. The location of princely burial grounds is established by annalistic messages, synodics of the XVI–XVII centuries, taking into account the later tradition of church veneration of many representatives of the princely family.

2. The reconstruction of the genome of Prince Dmitry Alexandrovich indicates the contribution of three ancestral components to his origin: (1) the early medieval population of the east of Scandinavia from the island of Oland, (2) representatives of the steppe nomadic peoples of the Eurasian steppes of the Iron Age or the early medieval population of central Europe (steppe nomads from the territory of Hungary), and (3) the ancient East-Eurasian component.

Target Text made by human translators

1. Существование «белого пятна» в изучении геномов древнерусской элиты во многом связано со сложностью персональной идентификации останков Рюриковичей и других аристократических родов в некрополях X–XIV вв. Хорошо известно, что имена погребенных до начала XV в. никак не обозначались на погребальных сооружениях, саркофагах и надгробиях. Местоположение княжеских погребений устанавливается по летописным сообщениям, синодикам XVI–XVII вв. с учетом позднейшей традиции церковного почитания многих представителей княжеского рода.

2. Моделирование генома князя Дмитрия Александровича указывает на вклад в его происхождение трех родовых компонентов: (1) популяции раннесредневекового населения востока Скандинавии с острова Эланд; (2) представителей степных кочевых народов евразийских степей железного века или раннесредневекового населения Центральной Европы (степные кочевники с территории Венгрии) и (3) древнего сибирского компонента.

ChatGPT-3.5

1.Существование **"темных пятен"** в изучении геномов средневековой русской знати во многом обусловлено сложностью личной идентификации останков Рюриковичей и других аристократических семей на некрополях X–XIV веков. Известно, что на похоронных сооружениях, саркофагах или надгробных плитах до начала XV века имена погребенных не указывались. Место княжеских захоронений **устанавливается по аннальным сообщениям**, синодикам XVI–XVII веков, с учетом поздней традиции церковного **поклонения** многим представителям княжеской семьи.

2.Реконструкция генома князя Дмитрия Александровича указывает на вклад **трех предковских компонентов** в его происхождение: (1) раннесредневекового населения востока Скандинавии с острова Оланд, (2) представителей степных кочевых народов евразийских степей железного века или раннесредневекового населения центральной Европы (степные кочевники с территории Венгрии), и (3) древневосточный евразийский компонент.

ChatGPT-4

1.Существование **«слепого пятна»** в изучении геномов средневековой русской знати во многих аспектах обусловлено сложностью личной идентификации останков Рюриковичей и других аристократических семей в некрополях X–XIV веков. Хорошо известно, что имена погребенных не указывались никак на похоронных структурах, саркофагах или надгробных плитах до начала XV века. Местоположение княжеских захоронений устанавливается по летописным сообщениям, синодикам XVI–XVII веков, с учетом поздней традиции церковного почитания многих представителей княжеской семьи.

2.Реконструкция генома князя Дмитрия Александровича указывает **на вклад трех предковых** компонентов в его происхождение: (1) раннесредневековое население востока Скандинавии с острова Эланд, (2) представители степных кочевых народов Евразийских степей железного века или раннесредневекового населения центральной Европы (степные кочевники с территории Венгрии), и (3) древний Восточно-Евразийский компонент.

Discussion- In the given extract, the both models of GPT could not come up with the translation of the idiomatic expression of a “blind spot” as well as with expression of – “three ancestral components to his origin”- **трех предковских компонентов/ на вклад трех предковых**. Also ChatGPT-3,5 misinterpreted the collocations: “annalistic messages” and “church veneration”

which have to be translated as “устанавливается по летописным сообщениям” and “церковного почитания”. While human translation also was incorrect in the terms of “Medieval Russian nobility” as “средневековая знать/дворянство” and “the ancient East-Eurasian component” as “древний Восточно-Евразийский компонент”, but using literal technique to create a text that is both accurate and readable, preserving the original meaning while ensuring it flows naturally in Russian.

7th Sample

Source Text

Establishment of the role of the polymorphism of the genes determining the activity of pro- and anti-inflammatory cytokines, including TGF- β 1, in the pathogenesis of various diseases in solid organ recipients will make it possible to both predict the risk of developing the disease and its severity and select a therapeutic approach for an individual patient.

Target Text made by human translators

Определение роли полиморфизма генов, определяющих активность про- и противовоспалительных цитокинов, в том числе и TGF- β 1, в патогенезе различных заболеваний у реципиентов солидных органов позволит, с одной стороны, прогнозировать риск развития патологии и тяжесть ее течения, а с другой – подбирать специфическую терапию для конкретного пациента.

ChatGPT-3.5

Установление роли полиморфизма генов, определяющих активность **противовоспалительных и противовоспалительных цитокинов, в транспатологии** различных заболеваний у **получателей твердых органов** позволит прогнозировать риск развития заболевания и его тяжесть и выбирать терапевтический подход для индивидуального пациента.

ChatGPT-4

Установление роли полиморфизма генов, определяющих активность про- и противовоспалительных цитокинов, включая TGF- β 1, в патогенезе различных заболеваний у

реципиентов **твердых органов, позволит как предсказать** риск развития заболевания и его тяжесть, так и выбрать терапевтический подход для индивидуального пациента.

Discussion-In the given extract the ChatGPT-3.5 made a repetition of the term “anti-inflammatory cytokines” as well as the mistranslation of the terms “the pathogenesis” as **“в транспатологии”** and “solid organ” considered the meaning as **“твердых органов”** as well as ChatGPT4 did. The human translator expanded translation by addition of “с одной стороны” and “а с другой” while we didn’t mention that phrases in the source text as might be equally – “from the one hand/from the other”. The correct translation could be more precise by using “как спрогнозировать, так и подобрать”.

8th Sample

Source text

Intranasal infection of mice with lethal doses of the cowpox virus or the ectromelia virus on day 60 post-immunization with the studied VACV variants demonstrated that the mutant LIVPΔati elicits a stronger protective response compared to the parent LIVP.

Target Text made by human translators

Интраназальное заражение мышей летальными дозами вируса оспы коров или вируса эктромелии на 60-й день после иммунизации изучаемыми вариантами VACV показало, что мутантный LIVPΔati обуславливает более выраженный **протективный эффект** по сравнению с родительским LIVP.

ChatGPT-3.5

Интраназальная инфекция мышей с смертельными дозами **вируса коровьего оспоподобного** вируса или вируса эктромелии на 60-й день после иммунизации изученными вариантами VACV показала, что мутантный LIVPΔati вызывает более сильный защитный ответ по сравнению с штаммом-родителем LIVP.

ChatGPT-4

Интраназальное заражение мышей летальными дозами вируса коровьей оспы или вируса эктромелии на 60-й день после иммунизации изученными вариантами VACV показало, что

мутант LIVP Δ ti вызывает более сильный защитный ответ по сравнению с родительским LIVP.

Discussion-In the given extract the term “protective response” have to be translated as “защитный ответ/реакция” instead of given translation made by human as “протективный эффект”. Moreover, ChatGPT-3.5 model couldn’t provide effective translation of the term “the cowpox virus” and translated as “вируса коровьего оспоподобного”.

9th Sample

Source Text

Microorganisms are found ubiquitously in the environment and play a crucial role in almost all ecosystems. Since many pathogens spread through the airborne route, including the SARS-CoV-2 coronavirus that has caused the current COVID-19 pandemic, it is especially relevant to study, monitor, and control the composition of outdoor and indoor air. Much data has been gained about the correlation between outdoor air pollution and the more severe course of COVID-19: for example, in India, a lower mortality rate from COVID-19 was observed in cities with better air quality.

Target Text made by human translators

Микроорганизмы повсеместно встречаются в окружающей среде и играют ключевую роль практически во всех экосистемах. В связи с распространением многих патогенных микроорганизмов воздушно-капельным путем, в том числе и коронавируса SARS-CoV-2, вызвавшего текущую пандемию COVID-19, изучение, мониторинг и регулирование состава воздуха как снаружи, так и внутри помещений приобрели особую актуальность. К настоящему времени уже накоплено много информации о корреляции между загрязнением наружного воздуха и более тяжелым течением COVID-19: так, в Индии более низкий уровень смертности от заболевания наблюдали в городах с лучшим качеством воздуха.

ChatGPT-3.5

Микроорганизмы обнаруживаются повсеместно в окружающей среде и играют ключевую роль в почти всех экосистемах. Поскольку многие патогены распространяются воздушным

путем, включая коронавирус SARS-CoV-2, который вызвал нынешнюю пандемию COVID-19, особенно важно изучать, **контролировать и контролировать состав открытого и закрытого воздуха.** Были получены многочисленные данные о взаимосвязи между загрязнением воздуха и более тяжелым течением COVID-19: например, в Индии было замечено более **низкое смертельное исхода** от COVID-19 в городах с лучшим качеством воздуха.

ChatGPT-4

Микроорганизмы встречаются повсеместно в окружающей среде и играют ключевую роль практически во всех экосистемах. Поскольку многие патогены распространяются по воздушному пути, включая коронавирус SARS-CoV-2, вызвавший текущую пандемию COVID-19, особенно актуально изучать, контролировать и мониторить состав наружного и **внутреннего** воздуха. Было получено много данных о корреляции между загрязнением наружного воздуха и более тяжёлым течением COVID-19: например, в Индии наблюдалась более низкая смертность от COVID-19 в городах с лучшим качеством воздуха.

Discussion- In the current extract the repetition provided by ChatGPT-3.5 mentioned in mistranslation of the part of the sentence “it is especially relevant to study, monitor, and control the composition of outdoor and indoor air” as **“контролировать и контролировать состав открытого и закрытого воздуха.”** The same problem we considered by ChatGPT-4 model’s translation in the collocation of “indoor air” as **“внутреннего воздуха”**. Human translator merely applied technique of calque to word “correlation” as “корреляции”. The better equivalent synonym for translation of the current word would be “соотношение, взаимосвязь, сопоставление”. Incorrect translation by ChatGPT-3.5 excluding grammatical rule of gender agreement in phrase **“более низкое смертельное исхода”**.

10th Sample

Source text

1. For many decades, diabetes mellitus (DM) has been one of the major health challenges in the world due to its associated increased morbidity, disability, and mortality.
2. The mechanisms of the most adequate and easily reproducible DM model, namely the streptozotocin-induced model of DM, were analyzed and discussed in.

3. There are considered to be two subtypes of T2DM: with mutations in individual genes (10–15%) and damage to a set of genes (85–90%) responsible for insulin binding to a cell receptor, internalization of the hormone–receptor complex, auto-phosphorylation of β receptors, and phosphorylation of other membrane protein components.

Target Text made by human translators

1. На протяжении многих десятилетий сахарный диабет (СД) был одной из основных проблем здравоохранения в мире из-за связанного с ним роста заболеваемости, инвалидности и смертности.

2. Проанализированы и обсуждены механизмы стрептозотоциновой модели СД, как наиболее адекватной и легко воспроизводимой.

3. Считается, что существуют два типа СД2: с повреждением отдельных генов (10–15%) и с повреждением с (85–90%), ответственных за связывание инсулина рецепторами клеток, а также за интернализацию гормон-рецепторного комплекса, за аутофосфорилирование бета-рецепторов или фосфорилирование других белковых компонентов мембран.

ChatGPT-3.5

1. На протяжении многих десятилетий сахарный диабет (СД) был одним из основных вызовов здравоохранения в мире из-за связанного с ним увеличения заболеваемости, инвалидности и смертности.

2. Механизмы наиболее адекватной и легко воспроизводимой модели СД, а именно модели СД, индуцированной стрептозотоцином, были проанализированы и обсуждались.

3. Считается, что существует два подтипа СД2: с мутациями в отдельных генах (10–15%) и повреждением набора генов (85–90%), ответственных за связывание инсулина с рецептором клетки, внутреннизацию комплекса гормон–рецептор, аутофосфорилирование β -рецепторов и фосфорилирование других компонентов мембраны белка.

ChatGPT-4

1. На протяжении многих десятилетий сахарный диабет (СД) оставался одной из основных проблем здравоохранения в мире из-за связанной с ним увеличенной заболеваемости, инвалидности и смертности.

2. Механизмы наиболее адекватной и легко воспроизводимой модели СД, а именно модели СД, индуцированной стрептозотоцином, были проанализированы и обсуждены.

3. Считается, что существуют два подтипа СД2: с мутациями в отдельных генах (10–15%) и повреждениями набора генов (85–90%), ответственных за связывание инсулина с клеточным рецептором, интернализацию комплекса гормон-рецептор, автофосфорилицию β -рецепторов и фосфорилицию других компонентов мембранного белка.

Discussion-In the given extract of scientific text the human translator made several mistakes in the term of “with mutations in” as **“с повреждением”** and in “a set of genes” mistranslated to an appropriate phrase **“повреждением множества ген”**. More concise translation of the mentioned terms should be “мутациями” and “набор ген”. Also, the term “of the most adequate” more suitable translated as “самая подходящая”. ChatGPT-3.5 also made several mistakes: the phrase “one of the major health challenges” translated as **“ОДНИМ ИЗ ОСНОВНЫХ ВЫЗОВОВ”** and “internalization” as **“внутреннизацию”**.

11th Sample

Source text

To date, more than 5,000 different microRNAs have been identified in the human genome (<http://www.mirbase.org>). The studied mechanism of microRNA action involves the implementation of RNA silencing. In the RNA-induced silencing complex (RISC), microRNA binds to the 3'-end of a complementary mRNA, leading to mRNA degradation and the prevention of protein translation. There are other mechanisms of expression regulation which involve microRNAs. Importantly, a single microRNA can bind to more than 200 different mRNAs, thus inducing shifts in the regulation of protein cascades. These changes may be the basis for various pathological processes, including those leading to neurodegeneration. In PD, shifts in blood microRNA levels may reflect, for the most part, the involvement of numerous organs and systems in the pathological process. It is the multiple organs pathology that is believed to cause the development of the non-motor (gastrointestinal, cardiac, etc.) manifestations that are so characteristic of PD.

Target Text made by human translators

На данный момент в геноме человека выявлено более 5000 различных микроРНК (<http://www.mirbase.org>). Изученный механизм действия микроРНК заключается в реализации РНК-сайленсинга. **В составе РНК-индуцируемого комплекса сайленсинга (RISC)** микроРНК связывается с 3'-концом комплементарной мРНК, что приводит к ее деградации и предотвращению трансляции белка. Описаны и другие механизмы регуляции экспрессии с участием микроРНК.

Немаловажно, что одна микроРНК может связываться более чем с 200 различными мРНК, индуцируя таким образом сдвиги в регуляции целых белковых ансамблей. Подобные изменения могут быть основой для развития самых разнообразных патологических процессов, в том числе, приводящих к нейродегенерации. При БП сдвиги концентраций микроРНК, определяемых в крови, могут в значительной степени отражать вовлечение в патологический процесс множества органов и систем – именно с полиорганной патологией связывают развитие столь характерных для БП немоторных проявлений (желудочно-кишечных, кардиальных и др.).

ChatGPT-4

На сегодняшний день в геноме человека идентифицировано более 5000 различных микроРНК (<http://www.mirbase.org>). Изученный механизм действия микроРНК включает реализацию РНК-опосредованного подавления. В комплексе РНК-индуцированного подавления (RISC) микроРНК связывается с 3'-концом комплементарной мРНК, приводя к деградации мРНК и предотвращению трансляции белка. Существуют и другие механизмы регуляции экспрессии, в которые вовлечены микроРНК. Важно, что одна микроРНК может связываться более чем с 200 различными мРНК, тем самым вызывая сдвиги в регуляции белковых каскадов. Эти изменения могут лежать в основе различных патологических процессов, включая нейродегенерацию. В БП сдвиги уровней микроРНК в крови могут отражать, в первую очередь, вовлечение множества органов и систем в патологический процесс. Считается, что именно множественная патология органов вызывает развитие немоторных (гастроинтестинальных, кардиальных и т. д.) проявлений, так характерных для БП.

ChatGPT-3.5

На сегодняшний день в геноме человека выявлено более 5000 различных микроРНК (<http://www.mirbase.org>). Изученный механизм действия микроРНК включает реализацию **рибонуклеинового затухания**. В комплексе **рибонуклеинового затухания** (RISC) микроРНК связывается с 3'-концом комплементарного мРНК, что приводит к деградации мРНК и предотвращению трансляции белка. Существуют и другие механизмы регуляции экспрессии, включающие микроРНК. Важно отметить, что одна микроРНК может связываться с более чем 200 различными мРНК, вызывая изменения в регуляции белковых каскадов. Эти изменения могут быть основой для различных патологических процессов, включая те, которые приводят к нейродегенерации.

При БП изменения уровня микроРНК в крови могут отражать, в большей части, вовлеченность многих органов и систем в патологический процесс. Именно патология множественных органов, по предположению, вызывает развитие немоторных (желудочно-кишечных, сердечных и т. д.) проявлений, которые так характерны для БП.

Discussion-in current extract we noted repetition of words made by ChatGPT-3.5. It seems the model faced with difficult predicting the new fragment of sequence and made addition expressed randomly. Shortly we can count it for hallucination. Human translator expressed the term “In the RNA-induced silencing complex (RISC)” as **“В составе РНК-индуцируемого комплекса сайленсинга”**, while ChatGPT4 model provided more clear translation by using “В комплексе РНК-индуцированного подавления (RISC)”.

12th Sample

Source Text

These platforms have already demonstrated their significant worth: in the last two years, five new genomes have been read – in addition to the reference genome sequence that was determined by several leading research groups at institutes in the U.S., United Kingdom, and Canada in the course of 10 years and at a cost of 3 bln dollars – such as the genomes of outstanding biologists of our times, that of a nigerian man, a chinese, and a Korean , not to mention thousands of other eukaryotic and prokaryotic species sequenced.

Target Text made by human translators

Эти платформы уже продемонстрировали свою состоятельность – за последние два года к уже известному геному человека, последовательность которого была определена рядом ведущих институтов Соединенных Штатов (США), Великобритании и Канады в течение 10 лет при общей стоимости проекта 3 млрд долларов, добавились пять новых: геномы выдающихся биологов современности, африканца нигерийского происхождения, китайца и корейца, **не считая геномов других эукариотических и прокариотических организмов.**

ChatGPT-4

Эти платформы уже доказали свою значительную ценность: за последние два года было прочитано пять новых геномов – помимо референсной последовательности генома, которая была определена несколькими ведущими исследовательскими группами в институтах США, Великобритании и Канады в течение 10 лет и стоимостью 3 млрд долларов – таких как геномы выдающихся биологов нашего времени, а также геном нигерийца, китайца и корейца, не говоря уже о тысячах других эукариотических и прокариотических видов, секвенированных.

ChatGPT-3.5

Эти платформы уже продемонстрировали свою значительную ценность: за последние два года было **прочитано** пять новых геномов - **помимо опорного геномного последовательности,** определенной несколькими ведущими научными группами институтов в США, Великобритании и Канаде в течение 10 лет и за сумму в 3 миллиарда долларов - таких как геномы выдающихся биологов нашего времени, геном нигерийца, китайца и корейца, а также тысячи других эукариотических и прокариотических видов.

Discussion: In the current extract we noted the avoidance of translating the word “sequenced” and ChatGPT-3.5 made the error in grammatical rule of gender agreement in the phrase **“помимо опорного геномного последовательности”**.

13th Sample

Source Text

1.The major criteria for scholarly journals selection to be covered by the Web of Science are discussed. These criteria include the compliance of a journal with world standards, the international diversity, citation score of authors and editorial board members, the journal impact factor and

journals self-citation. The demand for the unified transliteration of the authors name and the use of the unified English-language name and address of a research organization, as well as for including the information on funding organization (grant number and agency name) is emphasized.

2. The governments of a number of countries, including the developing ones, consider science and technology to be the key factor of economic growth and development and thus have confronted the challenge of establishing a knowledge-intensive economy.

3. The detailed pattern of the resulting significant changes in the number of students, PhD students, postdoctoral fellows, and researchers both in the U.S. and in other industrial and developing countries was presented in the report of the US national Science Foundation titled Science & engineering Indicators-2012 (SeI), published in January 2012, www.nsf.gov.

Target Text made by human translators

1. Рассмотрены основные критерии отбора научных журналов для включения их в информационную систему Web of Science (WoS) компании Thomson Reuters. К этим критериям относятся: соответствие журнала мировым стандартам, международный состав редколлегии, цитируемость авторов и членов редколлегии, импакт-фактор журнала и доля самоцитируемости в нем. Отмечается необходимость единообразной транслитерации фамилии автора и использования унифицированного англоязычного названия и адреса организации, а также включения информации о финансирующей организации (номер гранта и название фонда).

2. Правительства многих стран, в том числе развивающихся, рассматривая науку и технику как важнейший фактор экономического роста и развития, поставили задачу построить наукоемкую экономику.

3. В опубликованном в январе 2012 г. Отчете Национального научного фонда США «Показатели науки и техники» – Science & engineering Indicators-2012 (SeI) www.nsf.gov – представлена детальная картина значительных изменений в численности студентов, аспирантов, постдоков-1(ссылка объяснение) и исследователей как в США, так и в других индустриальных и развивающихся странах, приведена впечатляющая статистика об инвестициях в науку в США, Европейском союзе, Китае, России и других ведущих странах мира.

ChatGPT-4

1. Обсуждаются основные критерии отбора научных журналов для индексации в базе данных Web of Science. Эти критерии включают соответствие журнала мировым стандартам, международное разнообразие, цитируемость авторов и членов редакционного совета, импакт-фактор журнала и самоцитирование журналов. Подчеркивается необходимость единой транслитерации имен авторов, использования единого англоязычного названия и адреса исследовательской организации, а также включения информации о финансирующей организации (номер гранта и название агентства).

2. **Правительства ряда стран, включая развивающиеся, считают науку и технологии ключевым фактором экономического роста и развития и, таким образом, сталкиваются с задачей создания знаниевой экономики.**

3. В докладе Национального научного фонда США под названием "Индикаторы науки и инженерии-2012" (опубликован в январе 2012 года на сайте www.nsf.gov) была представлена подробная картина значительных изменений в числе студентов, аспирантов, **постдоков** и исследователей как в США, так и в других индустриальных и развивающихся странах.

ChatGPT-3.5

1. Обсуждаются основные критерии **выбора** научных журналов **для покрытия** в Web of Science. Эти критерии включают соответствие журнала мировым стандартам, международное разнообразие, цитирование авторов и членов редакционной коллегии, индекс цитирования журнала и самоцитирование журналов. Особое внимание уделяется единой транслитерации **имени автора** и использованию единого англоязычного названия и адреса исследовательской организации, а также **включению** информации о финансирующей организации (номер гранта и название агентства).

2. Правительства ряда стран, включая развивающиеся, считают науку и технологии ключевым фактором экономического роста и развития, и, **следовательно, столкнулись с вызовом установления знаниевой экономики.**

3. Подробная картина значительных изменений в численности студентов, аспирантов, **постдоков** и исследователей как в США, так и в других индустриальных и развивающихся

странах была представлена в отчете Национального научного фонда США под названием "Индикаторы науки и инженерии - 2012" (SeI), опубликованном в январе 2012 года, www.nsf.gov.

Discussion- in the current extract the both models struggled with problems to find an appropriate equal translation. For ChatGPT-4 the terms like “a knowledge-intensive economy” as **“знаниевой экономики”** as well as for ChatGPT-3.5- **“столкнулись с вызовом установления знаниевой экономики”**. The terms like “selection” and “covered” by the Web of Science were translated **“выбора научных журналов для покрытия”** incorrect by ChatGPT-3.5. The term of “PhD students” was expressed by 3 target texts as **“постдоков”**.

14th Sample

Source Text

1. In this paper, we compare the results of a sequencing of the aDNA libraries of a Bronze Age sample from burials of the Caucasian burial ground Klady, prepared using three different approaches: (1) shotgun sequencing, (2) strategies for selecting target genomic regions, and (3) strategies for selecting target genomic regions, including DNA pretreatment with a mixture of uracil-DNA glycosylase (UDG) and endonuclease VIII.

2. It was shown that preparation of genomic libraries without the use of UDG can result in distorted statistical data due to postmortem chemical modifications of the aDNA.

Target Text made by human translators

1. Нами проведено сравнение результатов секвенирования библиотек дДНК образца эпохи бронзы из погребений кавказского могильника Клады. Библиотеки были приготовлены с применением трех различных подходов: метода дробовика (shotgun sequencing), отбора целевых районов генома и отбора целевых районов генома с предварительной обработкой ДНК смесью урацил-ДНК-гликозилазы (UDG) и эндонуклеазы VIII.

2. Показано, что при приготовлении геномных библиотек без использования урацил-ДНК-гликозилазы возможно искажение результатов статистической обработки, связанное с посмертными химическими модификациями дДНК.

ChatGPT-4

1. В данной работе мы сравниваем результаты секвенирования библиотек aDNA образца из бронзового века с захоронений кавказского могильника Клады, подготовленных с использованием трех разных подходов: (1) секвенирование всего генома, (2) стратегии выбора целевых геномных регионов, и (3) стратегии выбора целевых геномных регионов, включая предварительную обработку ДНК смесью урацил-ДНК гликозилазы (UDG) и эндонуклеазы VIII.

2. Было показано, что подготовка геномных библиотек без использования UDG может привести к искажению статистических данных из-за посмертных химических модификаций aDNA.

ChatGPT-3.5

1. В этой статье мы сравниваем результаты секвенирования библиотек древней ДНК образца бронзового века из захоронений Кавказского кладбища Клады, подготовленных с использованием трех различных подходов: (1) секвенирование с использованием метода shotgun, (2) стратегии выбора целевых геномных областей и (3) стратегии выбора целевых геномных областей, включая предварительную обработку ДНК смесью урацил-ДНК-гликозилазы (UDG) и эндонуклеазы VIII.

2. Было показано, что подготовка геномных библиотек без использования UDG может привести к искажению статистических данных из-за **постморальных** химических модификаций древней ДНК.

Discussion-in the current extract the term “postmortem” was translated as **“постморальных”** by ChatGPT-3, while more precise term could be translated as “посмертных”.

15th Sample

Source text

All these functions are regulated through the interaction between stem cells and the microenvironment in the tissue cell niches that control these processes through direct cell-cell interactions, production of the extracellular matrix, release of extracellular vesicles, and secretion of growth factors, cytokines, chemokines, and proteases.

Target Text made by human translators

Регуляция всех этих функций осуществляется через взаимодействие стволовых клеток с микроокружением в тканевых клеточных нишах, контролирующим эти процессы через прямые межклеточные контакты, продукцию межклеточного матрикса, высвобождение внеклеточных везикул и секрецию факторов роста, цитокинов, хемокинов и протеаз.

ChatGPT-4

Все эти функции регулируются через взаимодействие между стволовыми клетками и микроокружением в клеточных нишах тканей, которые контролируют эти процессы через прямые клеточно-клеточные взаимодействия, производство внеклеточного матрикса, высвобождение экстрацеллюлярных везикул и секрецию факторов роста, цитокинов, хемокинов и протеаз.

ChatGPT-3.5

Все эти функции регулируются через взаимодействие между стволовыми клетками и микроокружением в клеточных нишах тканей, которые контролируют эти процессы через прямые клетка-клетка взаимодействия, производство внеклеточного матрикса, высвобождение внеклеточных везикул и выделение факторов роста, цитокинов, хемокинов и протеаз.

Discussion- In the given extract the term “direct cell-cell interactions” incorrect translated by the both models as “клеточно-клеточные взаимодействия”, “через прямые клетка-клетка взаимодействия”, while more precise and appropriate term is “прямые межклеточные взаимодействия”

16th Sample

Source text

It should also be noted that each malignant tumor is characterized by a large variety of mutations that are different in each individual and can have a different impact on the development of the tumor and emergence of resistance to therapy. The clonal evolution model and CSCs model are not mutually exclusive, since the evolution of CSCs most likely also follows the laws of clonal evolution.

Target Text made by human translators

Следует еще заметить, что каждая злокачественная опухоль характеризуется большим разнообразием мутаций, которые различаются у разных индивидов и способны по-разному **направлять развитие опухоли** и возникновение устойчивости к терапии. Модели клональной эволюции и РСК не являются взаимоисключающими, так как эволюция РСК, скорее всего, также осуществляется по законам клональной эволюции.

ChatGPT-4

Также следует отметить, что каждая злокачественная опухоль характеризуется большим разнообразием мутаций, которые различаются у каждого индивидуума и могут оказывать различное влияние на развитие опухоли и возникновение устойчивости к терапии. Модель клоновой эволюции и модель РСК **не исключают друг друга**, поскольку эволюция РСК, скорее всего, также следует законам клоновой эволюции.

ChatGPT-3.5

Следует также отметить, что каждая злокачественная опухоль характеризуется большим разнообразием мутаций, которые различны у каждого индивида и могут оказывать различное воздействие на развитие опухоли и возникновение **сопротивления терапии**. Модель клонической эволюции и модель **CSCs не исключают друг друга**, поскольку эволюция **CSCs**, скорее всего, также следует законам клонической эволюции.

Discussion-in the current extract the translation made by human couldn't capture the sense of phrase "can have a different impact on the development of the tumor", it was translated as "способны по-разному **направлять развитие опухоли**" but would be more precise expressed as "могут по-разному влиять на развитие опухоли. Additionally, the phrase "models are not mutually exclusive" was mistranslated by both language models as **"модель CSCs не исключают друг друга"**. The abbreviation "CSCs" was also left untranslated by ChatGPT-3.5. However, the most suitable translation for this phrase should be "модели не являются взаимоисключающими."

17th Sample

Original text

1. These treatments, in combination with various cellular and additive techniques, bring hope that most spinal cord injuries will be curable in the near future.

2. Since the second half of the last century, techniques for the prevention of consequences arising from spinal cord injury have been the subject of systematic studies in various animals, including rats, mice, cats, dogs, and minipigs.

3. When choosing the optimal animal model for solving specific research problems, it is necessary to take into account many factors: the type, age, size, and gender of animals and the possibility of using visualization techniques and a functional assessment of their condition.

Target Text made by human translators

1. В сочетании с различными клеточными и аддитивными технологиями эти методы дают надежду на то, что в скором времени большинство травм спинного мозга будут излечимы.

Начиная со второй половины прошлого века способы предотвращения последствий травмы спинного мозга стали предметом систематического исследования на различных животных, включая крыс, мышей, кошек, собак, **миниpigов**.

При выборе оптимальной животной модели для решения конкретных исследовательских задач необходимо учитывать множество факторов: вид, возраст, размер, пол животных, возможность применения методов визуализации и функциональной оценки их состояния.

ChatGPT-3.5

1. Эти методы, в сочетании с различными клеточными и добавочными техниками, приносят надежду на то, что большинство повреждений спинного **мозга будут лечимы** в ближайшем будущем.

2. С середины прошлого века методики предотвращения последствий, возникающих в результате повреждения спинного мозга, являются предметом систематических **исследований в различных** животных, включая крыс, мышей, кошек, собак и **миниpigов**.

3. При выборе оптимальной животной модели для решения конкретных исследовательских задач необходимо учитывать множество факторов: тип, возраст, размер и пол животных, а также возможность использования **визуализационных** техник и функциональной оценки их состояния.

ChatGPT-4

1. Эти методы лечения в сочетании с различными клеточными и добавочными техниками дают надежду, что в ближайшем будущем большинство травм спинного мозга будут излечимы.

2. Со второй половины прошлого века методы предотвращения последствий травмы спинного мозга стали предметом систематических исследований у различных животных, включая крыс, мышей, кошек, собак и **минипиггов**.

3. При выборе оптимальной животной модели для решения конкретных исследовательских задач необходимо учитывать множество факторов: тип, возраст, размер и пол животных, а также возможность использования техник визуализации и функциональной оценки их состояния.

Discussion-in the current extract the way of translation word “minipigs” by human translator can be expressed more clearly as “карликовая домашняя свинья” but in Russian language we also use the term –minipigs, as well as the term “аддитивными технологиями”, which was translated by both language models as “добавочными техниками”. However, the meaning **“добавочными техниками”** may not carry the same technical weight in the context. The similar problems and errors in grammatical rule of gender agreement in the phrases as we considered before was mentioned in translation by ChatGPT3.5 **“мозга будут лечимы”, “визуализационных техник”, “исследований в различных животных”**.

18th Sample

Source Text

The retinal pigment epithelium is a monolayer of pigmented, hexagonal cells connected by tight junctions. These cells compose part of the outer blood-retina barrier, protect the eye from excessive light, have important secretory functions, and support the function of photoreceptors, ensuring the coordination of a variety of regulatory mechanisms.

It is the degeneration of the pigment epithelium that is the root cause of many retinal degenerative diseases. The search for reliable cell sources for the transplantation of retinal pigment epithelium is of extreme urgency.

When light has been captured, retinal isomerizes from 11-cis-retinal into trans-retinal and then converts to trans-retinol.

There is no doubt, that the RPE plays an important role in...

Target Text made by human translators

Пигментный эпителий сетчатки представляет собой монослой пигментированных гексагональных клеток, соединенных плотными контактами. Эти клетки составляют часть внешнего барьера между кровью и сетчаткой, защищают глаз от излишнего света, обладают важными секреторными функциями, а также поддерживают работу фоторецепторов, обеспечивая координацию множества регуляторных механизмов. Именно дегенерация пигментного эпителия является первопричиной многих дегенеративных заболеваний сетчатки глаза. Поиск надежных источников клеток для трансплантации пигментного эпителия сетчатки представляет крайне актуальную задачу.

При поглощении света ретиналь изомеризуется из 11-цис-ретинала в трансретиналь, а затем превращается в трансретинол.

Несомненно, ПЭС играет важнейшую роль в...

ChatGPT-3.5

Эпителий пигмента сетчатки представляет собой однослойный пигментированный гексагональный эпителий, соединенный тесными соединениями. Эти клетки составляют часть внешней **крововетиновой барьеры**, защищают глаз от **избыточного света**, выполняют важные секреторные функции и поддерживают работу фоторецепторов, обеспечивая координацию различных регуляторных механизмов. Именно дегенерация пигментного эпителия является корневой причиной многих дегенеративных заболеваний сетчатки. Поиск надежных источников клеток для трансплантации эпителия пигмента сетчатки **крайне срочен**.

При захвате света ретиналь изомеризуется из 11-цис-ретинала в транс-ретиналь, а затем превращается в транс-ретинол.

Нет сомнения, что RPE играет важную роль в...

ChatGPT-4

Пигментный эпителий сетчатки представляет собой однослойный слой пигментированных, (шестигранных may be) **шестиугольных клеток**, соединенных плотными контактами. Эти

клетки составляют часть внешнего кровеносно-сетчаточного барьера, защищают глаз от **избыточного света**, выполняют важные секреторные функции и поддерживают функцию фоторецепторов, обеспечивая координацию различных регуляторных механизмов. Дегенерация пигментного эпителия является корневой причиной многих дегенеративных заболеваний сетчатки. Поиск надежных источников клеток для трансплантации пигментного эпителия сетчатки имеет **чрезвычайную важность**.

Когда свет пойман, ретиналь изомеризуется из 11-цис-ретинала в транс-ретиаль, а затем превращается в транс-ретинол.

Безусловно, ПЭС играет важную роль в.....

Discussion- the terms “blood-retina barrier”, “protect the eye from excessive light”, “There is no doubt”, “is of extreme urgency”, “hexagonal cells”, “When light has been captured” were mistranslated by both models, which we consider as rude mistakes, providing translation of scientific texts. The clearer and precise understandable terms in Russian language have to be translated as “гематоэнцефалический барьер”, “защита глаз от чрезмерного освещения/излишнего света”, “несомненно”, “крайне важно”, “гексагональные клетки”, “при поглощении света”.

19th Sample

Source Text

The Contribution of Major Histocompatibility Complex Class II Genes to an Association with Autoimmune Diseases.

This review dwells on the most relevant aspects of this problem: namely, the correlation between carriage of certain MHC II alleles and an increased (positively associated allele) or reduced (negatively associated allele) probability of developing the most common autoimmune diseases, such as type 1 diabetes, rheumatoid arthritis, multiple sclerosis, systemic lupus erythematosus, autoimmune thyroiditis, etc.

This nomenclature has evolved through historical sequences of discovery of HLAs: they were named using Roman numerals and English alphabet letters as they were progressively discovered.

Target Text made by human translators

Вклад генов главного комплекса гистосовместимости класса II в предрасположенность к аутоиммунным заболеваниям.

В обзоре рассмотрены наиболее значимые современные представления о взаимосвязи носительства определенных аллелей МНС II с повышенной (положительно ассоциированные аллели) и пониженной (отрицательно ассоциированные аллели) вероятностью развития наиболее распространенных аутоиммунных заболеваний, таких, как сахарный диабет типа 1, ревматоидный артрит, рассеянный склероз, системная красная волчанка, аутоиммунный тиреоидит и др.

Такая номенклатура возникла вследствие исторической очередности открытия антигенов HLA – их называли римскими цифрами и буквами алфавита по мере открытия.

ChatGPT-3.5

Вклад генов основного комплекса гистосовместимости класса II **в ассоциацию с** аутоиммунными заболеваниями.

Этот обзор затрагивает наиболее актуальные аспекты этой проблемы: взаимосвязь между наличием определенных аллелей МНС II и увеличенной (положительно ассоциированный аллель) или сниженной (отрицательно ассоциированный аллель) вероятностью развития наиболее распространенных аутоиммунных заболеваний, таких как сахарный диабет 1-го типа, ревматоидный артрит, рассеянный склероз, системная красная волчанка, аутоиммунный тиреоидит и др.

Эта номенклатура **развивалась через** исторические последовательности открытия HLA: они назывались с использованием римских цифр и английских букв алфавита по мере их последовательного обнаружения.

ChatGPT-4

Вклад генов главного комплекса гистосовместимости класса II **в ассоциацию с** аутоиммунными заболеваниями.

Этот обзор касается наиболее актуальных аспектов этой проблемы: а именно, корреляции между носительством определённых аллелей МНС II и увеличенной (позитивно ассоциированные аллели) или уменьшенной (негативно ассоциированные аллели)

вероятностью развития наиболее распространенных аутоиммунных заболеваний, таких как диабет 1-го типа, ревматоидный артрит, множественный склероз, системная красная волчанка, аутоиммунный тиреоидит и др.

Эта номенклатура развивалась в ходе исторической последовательности открытий HLA: они были названы с использованием римских цифр и букв английского алфавита по мере их постепенного обнаружения.

Discussion- in the 1st sentence of mentioned above extract the term “to an Association” was mistranslated by both GPT models and directly translated as “в ассоциацию”, while the meaning of the term should be considered as “взаимосвязь”. Thus the term “multiple sclerosis” translated by ChatGPT4 model as “множественный склероз” also unappropriated. The collocation “развивалась через” mistranslated by ChatGPT-3.5 have to be corrected as “возникла вследствие исторической очередности/развивалась в ходе исторической последовательности” better capturing the sense of historical progression. Such errors lead to misunderstandings.

20th Sample

Source text

In this study, we evaluated catecholamine concentration changes in the tear fluid of untreated PD patients as biomarkers.

Therefore, the challenge is to develop an early (pre-clinical) diagnosis of PD, which would enable detection of the disease before the appearance of the first motor symptoms and early start of preventive neuroprotective therapy.

The conjunctiva lining the inner surface of the eyelids and conjunctival glands display sympathetic innervation; dysfunction of the conjunctiva occurs in PD as part of a multisystem degeneration affecting both the central and peripheral parts of the autonomic nervous system.

Clinically, PD manifests itself many years after the disease onset, after most of the nigral dopaminergic neurons have died, which explains the limitations of the current PD pharmacotherapy.

Target Text made by human translators

Нами изучена возможность использования в качестве маркеров изменений концентрации катехоламинов в слезной жидкости больных с БП.

Отсюда следует, что важнейшей задачей является разработка ранней (доклинической) диагностики БП, которая позволит выявить заболевание до появления первых двигательных симптомов и использовать превентивную нейропротекторную терапию.

Конъюнктивa, выстилающая внутреннюю поверхность век, и находящиеся в ней железы имеют симпатическую иннервацию, дисфункция которой отмечается при БП в рамках мультисистемной дегенерации, затрагивающей как центральные, так и периферические отделы автономной нервной системы.

Клинически БП проявляется через много лет после начала патологического процесса при гибели большей части дофаминергических нейронов черной субстанции, что объясняет существующие ограничения фармакотерапии заболевания.

ChatGPT-4

В данном исследовании мы оценили изменения концентрации катехоламинов в слезной жидкости **необработанных пациентов** с PD в качестве биомаркеров.

Поэтому задача состоит в разработке ранней (предклинической) диагностики PD, которая позволила бы обнаружить заболевание до появления первых **моторных симптомов** и раннего начала профилактической нейропротективной терапии.

Конъюнктивa, выстилающая внутреннюю поверхность век и конъюнктивальные железы, демонстрируют симпатическую иннервацию; дисфункция конъюнктивы возникает при PD как часть мультисистемной дегенерации, затрагивающей как центральные, так и периферические части автономной нервной системы.

Клинически PD проявляется много лет после начала заболевания, после гибели большинства дофаминергических нейронов нигры, что объясняет ограничения существующей фармакотерапии PD.

ChatGPT-3.5

В этом исследовании мы оценили изменения концентрации катехоламинов в слезной жидкости **у необслуженных пациентов** с БП в качестве биомаркеров.

Поэтому вызов состоит в разработке ранней (предклинической) диагностики БП, которая позволила бы обнаруживать заболевание до появления первых двигательных симптомов и раннего начала профилактической нейропротективной терапии.

Конъюнктивит, выстилающая внутреннюю поверхность век, и конъюнктивальные железы демонстрируют симпатическую иннервацию; дисфункция конъюнктивы возникает при БП как часть **мультиорганных** дегенеративных изменений, затрагивающих как центральные, так и периферические части вегетативной нервной системы.

Клинически БП проявляется много лет после начала заболевания, когда большинство nigralных допаминергических нейронов погибли, что объясняет ограничения текущей фармакотерапии БП.

Discussion- in the given extract, the term “of untreated PD patients” was mistranslated by both GPT models as **“необработанных пациентов”, “необслуженных пациентов”** have to be corrected as “у пациентов, не получавших лечение”. The next term “the first motor symptoms”- was identified by ChatGPT-4 as **“моторных симптомов”** have to be translated as “первых двигательных симптомов” in order to make more clear and readable meaning of the term. The collocation of “the challenge is to develop” mistranslated by ChatGPT-3.5 as **(Поэтому вызов состоит)** and have to be expressed as **“Поэтому задача состоит в”**. The term “the nigral dopaminergic neurons” expressed in human translation as “допаминергических нейронов черной субстанции” translated by both GPT models as “допаминергических нейронов нигры” and “нигральных допаминергических нейронов”, which considered as incorrect translation. Finally, the term “a multisystem degeneration” mistranslated by ChatGPT-3.5 as **“мультиорганных** дегенеративных изменений” and have to be corrected to the “мультисистемной дегенерации”.

21st Sample

Source text

1.The number of people that got infected during the most recent Ebola hemorrhagic fever outbreak exceeded 28,000, and more than 11,000 deaths were reported.

2.Thus, intravenous injection of inhibitors of blood coagulation, such as recombinant human-activated protein C, increases the survival rate in patients by 18%.

3. Meanwhile, G.G. Olinger Jr. et al. demonstrated in primate experiments that, unlike nonspecific antiviral therapeutics, passive immunization with antibodies injected 24 h post-infection has a therapeutic effect.

4. Administration of the study drug in a larger patient sample was limited, because the trial was conducted at the final stage of the epidemic, when the number of newly infected patients was rather small, making it difficult to recruit Ebola virus carriers.

5. The aim of this study was to design full-length recombinant chimeric antibodies against EBOV GP based on murine mAbs and to investigate their immunochemical properties: to determine the authenticity, specificity, and immunoreactivity of full-length chimeric antibodies and to measure the dissociation constants and the epitope specificity.

Target Text made by human translators

1. Во время последней эпидемии геморрагической лихорадки Эбола количество зараженных превысило 28 тысяч и зафиксировано более 11 тысяч летальных исходов.

2. Так, внутривенное введение ингибиторов коагуляции крови, таких, как рекомбинантный активированный белок С человека, **повышает выживаемость больных на 18%.**

3. Вместе с тем, в опытах на приматах показано, что, в отличие от неспецифических средств противовирусной терапии, пассивная иммунизация антителами, введенными через 24 ч после инфицирования, позволяет добиться терапевтического эффекта.

4. Применение препарата на большем количестве пациентов было ограничено возможностью привлечения носителей вируса, так как проводилось на финальной стадии эпидемии, когда регистрировалось уже незначительное число новых инфицированных пациентов.

5. Цель настоящей работы состояла в создании на основе мышиных мАТ полноразмерных рекомбинантных химерных антител к EBOVGP, а также в изучении их иммунохимических свойств: установлении подлинности, специфичности и иммунореактивности полноразмерных химерных антител, определении констант диссоциации и эпитопной специфичности.

ChatGPT-4

- 1.Количество людей, заразившихся в ходе последней вспышки лихорадки Эбола, превысило 28 000, и было зарегистрировано более 11 000 смертей.
- 2.Так, внутривенное введение ингибиторов **кровосвертывания**, таких как рекомбинантный активированный белок С человека, увеличивает выживаемость пациентов на 18%.
- 3.Тем не менее, G.G. Олингер мл. и соавторы показали в экспериментах на приматах, что в отличие от неспецифических противовирусных терапевтических средств, пассивная иммунизация с введением антител через 24 часа после заражения имеет терапевтический эффект.
- 4.Применение исследуемого препарата на **большой выборке пациентов** было ограничено, так как испытание проводилось на заключительном этапе эпидемии, когда количество новых случаев заражения было невелико, что затрудняло набор носителей вируса Эбола.
5. Целью данного исследования было разработать полноразмерные рекомбинантные химерные антитела против EBOV GP на основе мышинных мАт и изучить их иммунохимические свойства: определить подлинность, специфичность и иммунореактивность полноразмерных химерных антител, а также измерить диссоциационные константы и специфичность эпитопов.

ChatGPT-3.5

- 1.Число зараженных во время последнего **вспышечного** острого вирусного геморрагического лихорадочного заболевания Эбола превысило 28 000, и было сообщено о более чем 11 000 смертях.
- 2.Так, внутривенное введение ингибиторов **кровяных свертывающих факторов**, таких как рекомбинантный активированный белок С человека, увеличивает выживаемость пациентов на 18%.
- 3.Тем временем Г.Г. Олингер мл. и др. продемонстрировали на **приматовых экспериментах**, что в отличие от неспецифических антивирусных терапевтических средств, пассивная иммунизация антителами, введенными через 24 часа после инфекции, оказывает терапевтический эффект.

4. Введение исследуемого препарата в более **крупную выборку пациентов** было ограничено, потому что исследование проводилось на финальном этапе эпидемии, когда количество новых инфицированных пациентов было достаточно мало, что затрудняло набор носителей вируса Эбола.

5. Целью данного исследования было создание полноразмерных рекомбинантных химерных антител против EBOV GP на основе **муриных МАб** и исследование их иммунохимических свойств: определение подлинности, специфичности и иммунореактивности полноразмерных химерных антител, а также измерение констант диссоциации и специфичности эпитопов.

Discussion- in the current extract the translator used mission technique and didn't not specify the reference of G.G. Olinger Jr. et al. In the 2nd sentence the phrase "increases the survival rate in patients by 18%" translated as **"повышает выживаемость больных на 18%"** while can be expressed as "увеличивает коэффициент выживаемости у пациентов на 18%". The term "blood coagulation" misinterpreted by ChatGPT-4 as **"кровосвертывания,** and **"кровяных свертывающих факторов"** by ChatGPT-3.5, while more correct term would be expressed as "свертывания крови". Additionally, in the 4th sentence the phrase "Administration of the study drug in a larger patient" as **"большой выборке пациентов"** and **"крупную выборку пациентов"** was incorrect translated by both language models. Many significant errors lead to misunderstandings in translation of the terms such "outbreak"- **"вспышечного"**, "in primate experiments"- **"приматовых экспериментах"**, "based on murine mAbs"- **"муриных МАб"** by ChatGPT-3.5. The precise translation of the terms has to be expressed as following- "вспышка", "в опытах на приматах", "основе мышиных МАТ".

22nd Sample

Source text

The most important factors associated with experiments in conventional monolayer cell cultures (2D) include the selection of a specific cell phenotype (adapted for growth on a plastic surface) from an initially very heterogeneous tumor cell population, abnormal cell polarization resulting from limited exposure of the cell surface to the culture medium, a drastic reduction in the number of cell-cell contacts, and a lack of cell-matrix interactions and metabolic gradients.

Target Text made by human translators

К числу важнейших факторов, сопровождающих эксперименты на традиционных культурах клеток в монослое (2D), относятся селекция специфического, адаптированного к росту на культуральном пластике фенотипа клеток из первоначально крайне гетерогенной опухолевой клеточной популяции, аномальная поляризация клеток, связанная с ограниченной экспозицией поверхности клеток к культуральной среде, резкое сокращение **клеточно-клеточных** контактов, отсутствие клеточно-матриксных взаимодействий и метаболических градиентов

ChatGPT-4

Важнейшими факторами, связанными с экспериментами в обычных однослойных клеточных культурах (2D), являются выбор специфического клеточного фенотипа (адаптированного для роста на пластиковой поверхности) из первоначально очень гетерогенной популяции опухолевых клеток, аномальная поляризация клеток в результате ограниченного воздействия поверхности клеток на культуральную среду, резкое снижение числа контактов **клетка-клетка** и отсутствие взаимодействий клетка-матрица и метаболических градиентов.

ChatGPT-3.5

Самые важные факторы, связанные с экспериментами в обычных монослоевых клеточных культурах (2D), включают выбор конкретного фенотипа клеток (адаптированных для роста на пластиковой поверхности) из изначально очень гетерогенной популяции опухолевых клеток, нарушение поляризации клеток из-за ограниченного воздействия среды культуры на поверхность клетки, радикальное сокращение контактов между клетками и отсутствие взаимодействия клеток с матрицей и метаболических градиентов.

Discussion-in the given extract the term “cell-cell contacts” translated irrelevant by translator as **“клеточно-клеточных** контактов” instead of **“межклеточных”** as well as ChatGPT-4 translated.

23rd Sample

Source text

Ischemic lesions initiate the development of a systemic inflammatory response (SIR) and endothelial dysfunction, which is the main cause of the multiple organ failure in PE.

The cells of the invasive (extravillous) trophoblast acquire the properties of pseudoneoplastic cells with a high proliferative, invasive and migratory potential, as well as specific expression of surface markers during placentation, which enables FPS formation and promotes the phenomenon of nonrejection.

Target Text made by human translators

Ишемические повреждения инициируют развитие системного воспалительного ответа (СВО) и эндотелиальной дисфункции, основных причин развития полиорганной недостаточности при ПЭ.

Инвазивный (вневорсинчатый) трофобласт в период плацентации приобретает свойства псевдоопухолевых клеток с высоким пролиферативным, инвазивным и миграционным потенциалом и особенностью экспрессии поверхностных маркеров, что обеспечивает формирование ФПС и способствует феномену неотторжения.

ChatGPT-4

Ишемические поражения инициируют развитие системного воспалительного ответа (СВО) и эндотелиальной дисфункции, которая является основной причиной **множественной органной недостаточности** при ПЭ.

Клетки инвазивного (экстравиллозного) трофобласта приобретают свойства псевдонеопластических клеток с высоким пролиферативным, инвазивным и миграционным потенциалом, а также специфической экспрессией поверхностных маркеров во время плацентации, что способствует формированию ФПС и поддерживает **явление непринятия.**

ChatGPT-3.5

Ишемические поражения инициируют развитие системного воспалительного ответа (SIR) и эндотелиальной дисфункции, которая является основной причиной **множественной органной** недостаточности при РЕ.

Клетки инвазивного (экстрахорионического) трофобласта приобретают свойства псевдонеопластических клеток с высоким потенциалом пролиферации, инвазии и миграции, а также специфической экспрессией поверхностных маркеров во время плацентации, что обеспечивает формирование FPS и способствует **явлению непринятия.**

Discussion-in the current fragment of text the term “the multiple organ failure” mistranslated by both models as **“множественной органной недостаточности”** while the correct translation would be “полиорганной недостаточности”. The term “promotes the phenomenon of nonrejection” has precise translation made by human as “способствует феномену неотторжения” while both language models misinterpreted as **“явление неприятия”**.

24th Sample

Source text

1. In the nearest future, if a preserved fossils brain is found, it will be possible to identify the evolutionary changes in the higher nervous system associated with epigenetic differences.

2. At that stage, researchers encountered a host of difficulties related to the quality of the ancient DNA, its contamination with foreign DNA and others, and these problems were resolved, in particular, by means of improved whole genome sequencing methods.

3. Until recently, it was considered impossible to extract information on the transcriptional activity of genes from DNA that had been isolated long after the death of an individual.

Target Text made by human translators

В дальнейшем **при нахождении хорошо сохранившихся тканей** появляется перспектива обнаружения эволюционных изменений, связанных с эпигенетическими различиями в высшей нервной деятельности современного и древнего человека.

На этом этапе исследователи сталкивались с множеством трудностей, связанных с качеством древней ДНК, ее контаминацией чужеродной ДНК и пр. Эти сложности удалось преодолеть, в частности, совершенствованием методов полногеномного секвенирования.

До недавнего времени считалось невозможным получить информацию о транскрипционной активности генов на основе ДНК, выделенной через длительное время после смерти индивида.

ChatGPT-4

В ближайшем будущем, если будет найден хорошо сохранившийся мозг ископаемого, станет возможным идентифицировать эволюционные изменения в высших нервных системах, связанные с эпигенетическими различиями.

На этом этапе исследователи столкнулись с множеством трудностей, связанных с качеством древней ДНК, её загрязнением чужеродной ДНК и другими, и эти проблемы были решены, в частности, с помощью улучшенных методов секвенирования всего генома.

До недавнего времени считалось невозможным извлечь информацию о транскрипционной активности генов из ДНК, выделенной давно после смерти индивидуума.

ChatGPT-3.5

В ближайшем будущем, если будет найдено сохраненное мозговое окаменелое останки, будет возможно выявить эволюционные изменения в высшей нервной системе, связанные с эпигенетическими различиями.

На этом этапе исследователи столкнулись с рядом трудностей, связанных с качеством древней ДНК, ее загрязнением чужеродной ДНК и другими, и эти проблемы были решены, в частности, с помощью усовершенствованных методов секвенирования всего генома.

До недавнего времени считалось невозможным извлечь информацию о транскрипционной активности генов из ДНК, изолированной долгое время после смерти индивида.

Discussion- in the 1st sentence of the given extract the human translator used the generalization technique and the phrase “if a preserved fossils brain is found” translated as “при нахождении хорошо сохранившихся тканей”. ChatGPT-3.5 translated incorrect, avoiding grammatical rules of gender agreement in the sentences as in “мозговое окаменелое останки”. The collocation “contamination with foreign DNA” have to be translated as “контаминацией чужеродной ДНК” because this translation more appropriate and digestible for readers of scientific texts. Both GPT models translated the same phrase less clear as “загрязнением чужеродной ДНК”. At the last sentence “had been isolated long after the death of an individual” misinterpreted by ChatGPT-3.5 as “изолированной долгое время” but appropriate translation has to be used as “выделенной”.

25th Sample

Source text

In the past few decades, with the rapid development and commercialization of reproductive technologies, preimplantation human embryos have become the focus of close attention. Ethical

norms and legal constraints have put limitations on all but non-invasive methods in the study of conceptus; i.e. only microscopic observations.

Comparison of the genealogy of blastomeres with the sequence of their divisions allowed us to assume the role of cleavage as predetermining the normal course of the events accompanying embryo compaction and cavitation.

The first (meridional) cleavage furrow of the polarized zygote of placental mammals begins at the animal pole (in the immediate vicinity of the polar body) and extends from the animal to the vegetal pole.

Target Text made by human translators

Последние десятилетия в связи с бурным развитием и коммерциализацией репродуктивных технологий, предимплантационные исследования человеческих эмбрионов стали объектом пристального внимания. Этические нормы и юридические ограничения позволяют использовать только неинвазивных методов в изучении концептуса, т.е. только микроскопические наблюдения.

Сопоставление генеалогии бластомеров с последовательностью их делений позволило предположить роль дробления, предопределяющую нормальное развитие событий компактизации и кавитации эмбрионов.

Первая (меридиональная) борозда дробления поляризованной зиготы плацентарных млекопитающих врезается на анимальном полюсе (в непосредственной близости к полярному тельцу) и распространяется от анимального к вегетативному полюсу.

ChatGPT-4

В последние десятилетия, с быстрым развитием и коммерциализацией репродуктивных технологий, преимплантационные человеческие эмбрионы стали объектом пристального внимания.

Этические нормы и правовые ограничения **наложили ограничения на все,** кроме неинвазивных методов в изучении концептуса, т.е. только микроскопические наблюдения.

Сравнение генеалогии бластомеров с последовательностью их делений позволило предположить роль **расщепления** как предопределяющего нормальный ход событий, сопровождающих уплотнение и кавитацию эмбриона.

Первая (меридиональная) борозда расщепления поляризованной зиготы плацентарных млекопитающих начинается у животного полюса (**непосредственно у полярного тельца**) и тянется от животного к вегетативному полюсу

ChatGPT-3.5

В последние десятилетия, с быстрым развитием и коммерциализацией репродуктивных технологий, человеческие эмбрионы перед имплантацией стали объектом **тесного внимания**. Этические нормы и юридические **ограничения ограничили все**, кроме неинвазивных методов в изучении концепции; то есть только микроскопические наблюдения.

Сравнение генеалогии бластомеров с последовательностью их делений позволило нам предположить роль разделения как предопределяющего нормальный ход событий, сопровождающих **компакцию и кавитацию эмбриона**.

Первая (меридиональная) **шовное борозда** поляризованного зиготы плацентарных млекопитающих начинается у животного полюса (в непосредственной близости от полюсного тела) и простирается от животного к вегетативному полюсу.

Discussion- In the extract, ChatGPT-3.5 repeated the phrase **“ограничения ограничили все”** for “have put limitations on” which should be more precisely expressed as “наложены ограничения”. The same error made by ChatGPT-4 **“наложили ограничения на все”**. The term “the focus of close attention” mistranslated as **“тесного внимания”** instead of the correct Russian expression “пристального внимания”. The term “accompanying embryo compaction and cavitation” the correct translation is “сопутствующее уплотнение и кавитация эмбриона” and “The first (meridional) cleavage” the correct translation is “первая (меридиональная) борозда” mistranslated by ChatGPT-3.5 as “шовное борозда” and “компакцию и кавитацию эмбриона”

26th Sample

Source text

The hepatitis C virus (HCV) triggers a chronic disease that is often accompanied by a spectrum of liver pathologies and metabolic alterations. The oxidative stress that occurs in the infected cells is considered as one of the mechanisms of HCV pathogenesis.

The hepatitis C virus (HCV) is a widespread and dangerous pathogen that infects the human liver. In most cases, HCV infection leads to chronic hepatitis, during which there is a high risk of liver fibrosis and cirrhosis, hepatocellular carcinoma, and various metabolic disorders (steatosis, type 2 diabetes mellitus, altered iron metabolism and other pathologies). In the absence of stress, the Nrf2 factor is located in the cytoplasm in a complex with its partner protein Keap1.

Target Text made by human translators

Инфицирование вирусом гепатита С (ВГС) вызывает хроническое заболевание печени, которое сопровождается развитием различных патологий и метаболических нарушений. Одним из механизмов патогенеза ВГС считается возникновение в инфицированных клетках окислительного стресса, вызываемого белками вирусного капсида и NS5A.

Вирус гепатита С (ВГС) вызывает широко распространенное и опасное заболевание печени человека. В большинстве случаев заражение ВГС приводит к хроническому гепатиту, при котором велик риск развития фиброза и цирроза печени, а также гепатоцеллюлярной карциномы и различных метаболических нарушений (стеатоз, сахарный диабет типа 2, нарушения метаболизма железа и другие патологии).

В отсутствие стресса фактор Nrf2 находится в цитоплазме в комплексе с белком-партнером Keap1.

ChatGPT-4

Вирус гепатита С (HCV) вызывает хроническое заболевание, которое часто сопровождается рядом патологий печени и метаболических нарушений. Окислительный стресс в инфицированных клетках рассматривается как один из механизмов патогенеза HCV. **Он вызывает** белками ядра и NS5A вируса.

Вирус гепатита С (HCV) является широко распространенным и опасным патогеном, поражающим **человеческую печень**.

В большинстве случаев инфекция HCV приводит к хроническому гепатиту, во время которого существует высокий риск развития фиброза печени и цирроза, гепатоцеллюлярной карциномы и различных **метаболических расстройств** (стеатоз, сахарный диабет 2 типа, нарушенный метаболизм железа и другие патологии).

В отсутствие стресса фактор Nrf2 находится в цитоплазме в комплексе с его **партнерским белком** Keap1.

ChatGPT-3.5

Вирус гепатита С (HCV) вызывает хроническое заболевание, часто сопровождаемое спектром патологий печени и метаболических изменений. Окислительный стресс, возникающий в зараженных клетках, считается одним из механизмов патогенеза HCV. Его вызывают вирусные белки ядра и NS5A.

Вирус гепатита С (HCV) - это распространенный и опасный патоген, инфицирующий печень человека. В большинстве случаев инфекция HCV приводит к хроническому гепатиту, во время которого высок риск фиброза и цирроза печени, гепатоцеллюлярного карциномы и различных метаболических нарушений (стеатоз, сахарный диабет типа 2, **изменение обмена железа** и другие патологии).

В отсутствие стресса фактор Nrf2 располагается в цитоплазме в комплексе **с его партнерским белком** Keap1.

Discussion- in this extract the term “infects the human liver” less clear translated as **“человеческую печень.”**by ChatGPT-4 instead of “печень человека”. The term “with its partner protein Keap1” was translated word-for-word by both models as **“с его партнерским белком** Keap1” as well as in term “, altered iron metabolism”- **“изменение обмена железа”** have to be translated as “нарушенный метаболизм железа”, the same problem in translation of the term “various metabolic disorders” as **“метаболических расстройств”**- “метаболических нарушений”.

27th Sample

Source text

Cadherins are transmembrane proteins consisting of an extracellular, a transmembrane, and cytoplasmic domain.

The extracellular calcium-binding site is formed by five domains; the transmembrane region consists of a single chain of glycoprotein repeats.

The phenomenon of epithelial–mesenchymal transition was first described in the early 1980s in Elizabeth Hay’s laboratory, in both embryonic notochord and lens epithelial cells isolated from chicken embryos, and in differentiated chicken lens epithelial cells.

Target Text made by human translators

Кадгерини – это трансмембранные белки, состоящие из внеклеточной кальций связывающей части, образованной пятью доменами, трансмембранной части, представленной одиночной цепью гликопротеиновых повторов, и цитоплазматической части, соединенной с β -катенином и белком p120, стабилизирующим кадгерин на поверхности клетки.

Явление эпителиально-мезенхимального перехода впервые было описано в начале 1980-х в лаборатории Elizabeth Hay, как на эмбриональных эпителиальных клетках ното хорда и хрусталика, выделенных из куриных эмбрионов, так и на дифференцированных эпителиальных клетках хрусталика.

ChatGPT-4

Кадгерини —это трансмембранные белки, состоящие из экстрацеллюлярного, трансмембранного и цитоплазматического доменов. Экстрацеллюлярное место связывания кальция формируется пятью доменами; трансмембранный регион состоит из одной цепи гликопротеиновых повторов.

Явление эпителиально-мезенхимального перехода впервые было описано в начале 1980-х годов в лаборатории Элизабет Хей, как в эмбриональных ното хордах, так и в изолированных эпителиальных клетках хрусталика куриного эмбриона, и в дифференцированных куриных эпителиальных клетках хрусталика.

ChatGPT-3.5

Кадгеринины - это трансмембранные белки, состоящие из внеклеточной, трансмембранной и цитоплазматической областей. **Сайт связывания кальция** образуется пятью доменами; трансмембранный регион состоит из одной цепочки гликопротеиновых повторов.

Феномен эпителиально-мезенхимального перехода был впервые описан в начале 1980-х годов в лаборатории Элизабет Хей, как в эмбриональных ното хордах, так и в **линзовых эпителиальных клетках, выделенных из куриных эмбрионов, а также в дифференцированных куриных линзовых эпителиальных клетках.**

Discussion- in the current extract the second sentence of the source text mistranslated by ChatGPT-3.5 as well as term “calcium-binding site is formed by five domains” to “**Сайт связывания кальция**”. Additionally, ChatGPT-3.5 made mistake in the phrase “в эмбриональных ното хордах”-using the plural "нотохордах". In embryonic development, "notochord" typically refers to a singular structure. Also translating the term "линзовых эпителиальных клетках” would be more precise interpreted correctly as "эпителиальные клетки хрусталика" (lens epithelial cells). However, ChatGPT-4 demonstrates clear, precise translation of the entire extract.

28th Sample

Source text

This method utilizes bitransgenic Fos-Cre-eGFP mice in which a green fluorescent protein is expressed upon tamoxifen-induced Cre-recombination only in the cells where immediate early gene c-fos expression takes place due to the new experience.

We used the classical fear conditioning model to show that ex vivo microscopy of the eGFP protein in Fos-Cre-eGFP mice enables mapping of the neurons of the various brain regions that undergo Cre-recombination during acquisition of a new experience.

The development of new experimental methods for cell-resolution visualization of the neural substrates of cognitive processes throughout the whole brain plays a key role in addressing these problems.

Target Text made by human translators

Метод основан на использовании двойных трансгенных мышей Fos-Cre-eGFP, у которых в результате управляемой тамоксифеном Cre-рекомбинации зеленый флуоресцентный белок

экспрессируется только в тех клетках, в которых в результате получения животным нового опыта происходила экспрессия непосредственного раннего гена c-fos.

Решение этих задач требует создания новых экспериментальных методов, позволяющих визуализировать нейронные субстраты когнитивных процессов в масштабе целого мозга с клеточным разрешением.

ChatGPT-3.5

Этот метод использует битрансгенных мышей Fos-Cre-eGFP, в которых зеленый флуоресцентный белок экспрессируется только в клетках, где происходит тамоксифен-индуцированная Cre-рекомбинация, только в тех клетках, где происходит экспрессия гена немедленного реагирования c-fos в результате нового опыта.

Мы использовали классическую модель условного страха, чтобы показать, что экс виво микроскопия белка eGFP в мышах Fos-Cre-eGFP позволяет картографировать нейроны различных областей мозга, подвергающихся Cre-рекомбинации во время приобретения нового опыта. Развитие новых экспериментальных методов для визуализации на клеточном уровне нейронных подстратов когнитивных процессов во всем мозгу играет ключевую роль в решении этих проблем.

ChatGPT-4

Этот метод использует битрансгенные мыши Fos-Cre-eGFP, в которых зеленый флуоресцентный белок экспрессируется только в клетках, где происходит экспрессия немедленного раннего гена c-fos благодаря CRE-рекомбинации, индуцированной тамоксифеном.

Мы использовали классическую модель страхового условия для демонстрации того, что экс виво микроскопия белка eGFP у мышей Fos-Cre-eGFP позволяет картографировать нейроны различных областей мозга, которые проходят CRE-рекомбинацию во время приобретения нового опыта.

Разработка новых экспериментальных методов для визуализации клеточного разрешения нейронных субстратов когнитивных процессов во всем мозге играет ключевую роль в решении этих проблем.

Discussion- in the given extract many errors were found made by both language models, for the terms “the neural substrates”- **“подстратов”**, processes throughout the whole brain- **“во всем мозгу”**; “we used the classical fear conditioning model”- **“классическую модель страхового условия”**, “bitransgenic Fos-Cre-eGFP mice”- **“битрансгенные”**. Such errors lead to misunderstandings. The second sentence in the source text regarding the classical fear conditioning model is omitted entirely by human translator, which is a significant omission leading to loss of important information. The human translator used a foreignization strategy for "Cre-рекомбинация," retaining the foreign term to ensure scientific accuracy. For "экс-виво микроскопия," a calque was used, directly translating "ex vivo microscopy" while maintaining the structure of the original term. There was also an omission in the second sentence, which is a basic transformation in the translation process but leads to a loss of important information.

29th Sample

Source text

The study revealed the ability of GK-2 to ameliorate hyperglycemia induced by streptozotocine (STZ 100 mg/kg i.p.) in C57Bl/6 mice, to restore learning ability in the Morris Water Maze test, and to overcome depression after both intraperitoneal (0.5 mg/kg) and peroral (5 mg/kg) long-term administration.

It has been suggested that the reduced level of NGF in blood serum should be regarded as a biomarker for major depression. Such a reduction is also observed in patients with bipolar disorder and senile depressions.

Target Text made by human translators

Показано, что ГК-2 как при внутрибрюшинном (в дозе 0.5 мг/кг), так и пероральном (в дозе 5 мг/кг) введении устраняет гипергликемию, вызванную стрептозотоцином (100 мг/кг), восстанавливает число (%) животных, обучившихся в водном лабиринте Морриса, и ослабляет выраженность депрессивно-подобного состояния.

Снижение содержания NGF в сыворотке крови предложено рассматривать как биомаркер большой депрессии.

Подобное снижение наблюдается и при маниакально-депрессивном психозе, и при депрессиях позднего возраста.

ChatGPT-3.5

Исследование показало способность GK-2 улучшать гипергликемию, вызванную стрептозотоцином (STZ 100 мг/кг в/в) у мышей C57Bl/6, восстанавливать способность к обучению в тесте "Лабиринт Морриса", и преодолевать депрессию **после как внутрибрюшинного** (0,5 мг/кг), так и перорального (5 мг/кг) длительного применения.

Предполагается, что сниженный уровень NGF в сыворотке крови следует рассматривать как биомаркер для **крупной депрессии**. Такое снижение также наблюдается у пациентов с биполярным расстройством и **сенильной депрессией**.

ChatGPT-4

Исследование показало способность GK-2 улучшать гипергликемию, вызванную стрептозотоцином (STZ 100 мг/кг в/в) у мышей C57Bl/6, восстанавливать способность к обучению в тесте Морриса в водном лабиринте и преодолевать депрессию после длительного внутрибрюшинного (0,5 мг/кг) и перорального (5 мг/кг) введения.

Предполагается, что сниженный уровень NGF в сыворотке крови следует рассматривать как биомаркер основной депрессии.

Такое снижение наблюдается также у пациентов с биполярным расстройством и сенильными депрессиями.

Discussion-the given extract provides the ability of GK-2 to ameliorate hyperglycemia induced by streptozotocine, where both language models mistranslated the terms as “major depression and senile depressions” to **“сенильной депрессией”** and **“крупной депрессии.”** The term “overcome depression after both intraperitoneal” to “преодолевать депрессию **после как внутрибрюшинного**” mistranslated by ChatGPT-3.5. No mistakes were found in translation by ChatGPT-4.

30th Sample

Source text

The 95% confidence intervals for TREC and KREC levels (expressed in the number of cDNA copies per 105 cells) were established for different gestational groups.

The importance of studying immune system development in newborns is informed by the discovered dependence of the level of naive markers on the gestational stage in the early neonatal period. This approach has other advantages, including high sensitivity, high throughput capacity, relatively low cost, and the possibility of using DNA isolated from the minimum volume of a blood sample collected using Guthrie cards. However, in order to characterize IEI patients, one has to know the state of these immunity markers in a healthy individual, especially with taking into account his/her age and sex.

Target Text made by human translators

Определены 95% интервалы уровней TREC и KREC (выраженные в числе копий на 105клеток) в различных возрастных группах. Обнаружено, что количество маркеров наивных лимфоцитов зависит от гестационной стадии развития в ранний неонатальный период.

Этот анализ имеет и другие преимущества, включая высокую чувствительность, высокую пропускную способность, относительно низкую цену и возможность использования ДНК, выделенной из минимального объема образцов крови, собранных на Гатри-картах. Однако, чтобы говорить о характеристике пациентов с ВОИ, необходимо знать состояние данных маркеров иммунитета у здоровых индивидов, учитывая, в частности, половозрастные особенности каждого.

ChatGPT-3.5

Для различных групп по гестационному возрасту были установлены 95% доверительные интервалы для уровней TREC и KREC (выраженные в количестве копий cDNA на 105 клеток). Важность изучения развития иммунной системы у новорожденных обусловлена обнаруженной зависимостью уровня наивных маркеров от гестационного возраста в раннем неонатальном периоде.

Этот подход имеет другие преимущества, включая высокую чувствительность, высокую производительность, относительно низкие затраты и возможность использования ДНК, изолированной из минимального объема образца крови, взятого с помощью карточек Гутри. Однако для характеристики пациентов с ВОИ необходимо знать состояние этих маркеров иммунитета у здорового человека, особенно учитывая его возраст и пол.

ChatGPT-4

Для различных гестационных групп были установлены 95% доверительные интервалы для уровней TREC и KREC (выраженные в числе копий cDNA на 10^5 клеток).

Важность изучения развития иммунной системы у новорожденных подчеркивается обнаруженной зависимостью уровня наивных маркеров от гестационного этапа в раннем неонатальном периоде.

Этот подход имеет другие преимущества, включая высокую чувствительность, большую пропускную способность, относительно низкую стоимость, а также возможность использования ДНК, выделенной из минимального объема крови, собранной с использованием **карт Гутри.**

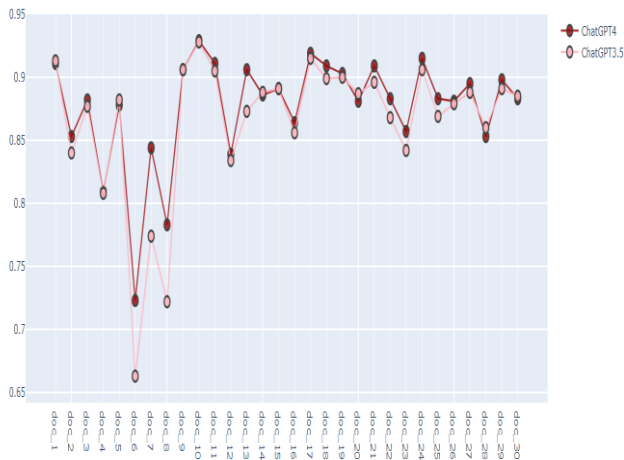
Однако для **характеризации** пациентов с ВОИ необходимо знать состояние этих иммунных маркеров у здорового человека, особенно с учетом его возраста и пола.

Discussion- in the given extract, the term “The 95% confidence intervals” correct translated as **“доверительные интервалы”** by both models but omitted in human translation text. The term “gestational groups” in the first sentence, translated as “возрастных группах” by the human translator, should be “гестационных группах” for accuracy. By employing literal translation, the calque technique—as in the example where “Guthrie cards” was translated as “Гатри-картах”—and the explicitation technique—adding details to clarify the meaning and make implicit information explicit, such as in the example where “in a healthy individual, especially with taking into account his/her age and sex” was translated as “у здоровых индивидов, учитывая, в частности, половозрастные особенности каждого,” adding “в частности” to make the reference to age and sex more explicit—the human translator ensured that the translation was both accurate and readable. The term “collected using Guthrie cards” mistranslated by both models as **“с помощью карточек Гутри”** and **“использованием карт Гутри.”**, while the correct term is “собранных на Гатри-картах”. The term “to characterize IEI patients” mistranslated as **“характеризации”** by ChatGPT-4, correctly have to be used as “характеристики пациентов”. The term “isolated from the minimum volume of a blood” mistranslated by ChatGT3.5 as **“изолированной”**, while more appropriate translation is “выделенной из минимального объема образцов крови”.

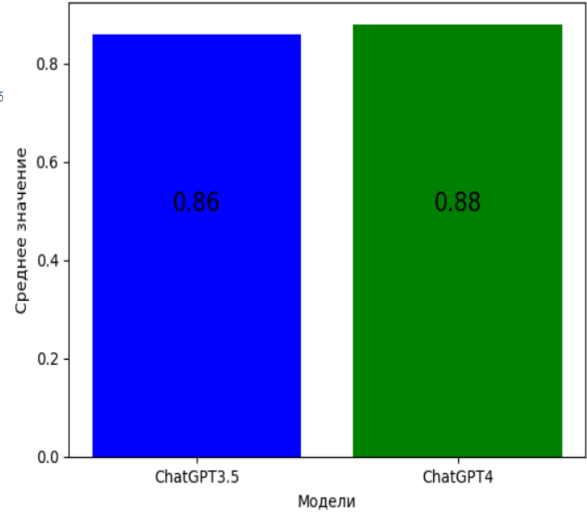
Quantitative Results by neural-based machine translation evaluation metric COMET-22

	ChatGPT3.5	ChatGPT4
doc_1	0,913	0,911
doc_2	0,840	0,853
doc_3	0,877	0,882
doc_4	0,808	0,809
doc_5	0,882	0,878
doc_6	0,663	0,723
doc_7	0,774	0,844
doc_8	0,722	0,782
doc_9	0,906	0,906
doc_10	0,928	0,929
doc_11	0,905	0,911
doc_12	0,834	0,839
doc_13	0,873	0,906
doc_14	0,888	0,886
doc_15	0,891	0,891
doc_16	0,856	0,864
doc_17	0,915	0,919
doc_18	0,899	0,909
doc_19	0,900	0,903
doc_20	0,887	0,881
doc_21	0,896	0,909
doc_22	0,868	0,883
doc_23	0,842	0,857
doc_24	0,906	0,915
doc_25	0,869	0,883
doc_26	0,879	0,881
doc_27	0,888	0,895
doc_28	0,860	0,853
doc_29	0,891	0,898
doc_30	0,885	0,883

Сравнение переводов по метрике Comet-22

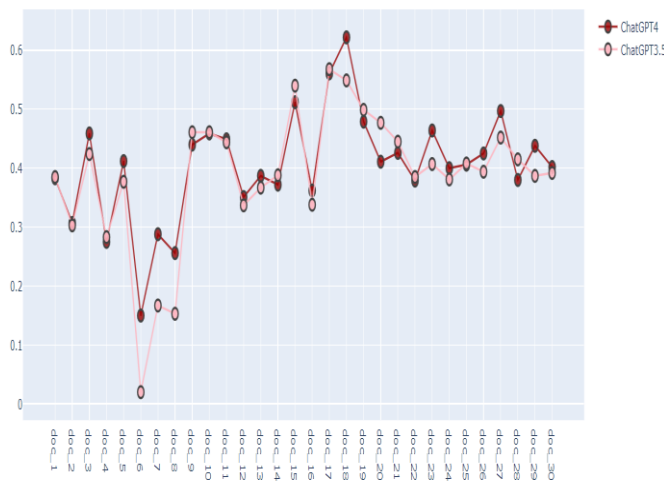


Средние значения Comet-22 на ChatGPT4 и ChatGPT3.5

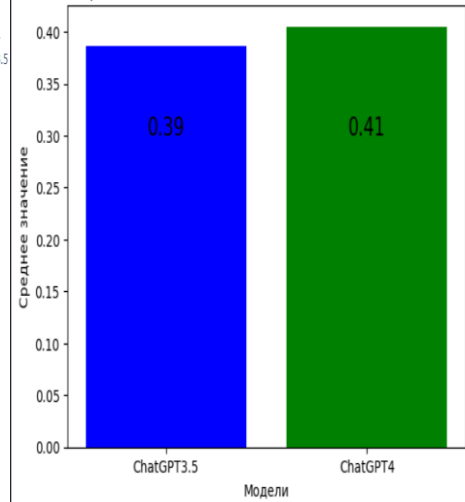


Quantitative Results by the neural-based machine translation evaluation metric BLEURT-20

Сравнение переводов по метрике BLEURT



Средние значения BLEURT на ChatGPT4 и ChatGPT3.5



Quantitative Results by the neural-based machine translation evaluation metric BLEURT-20

	ChatGPT3.5	ChatGPT4
doc_1	0,385	0,383
doc_2	0,303	0,307
doc_3	0,424	0,459
doc_4	0,283	0,275
doc_5	0,377	0,412
doc_6	0,02	0,15
doc_7	0,167	0,288
doc_8	0,153	0,256
doc_9	0,461	0,440
doc_10	0,461	0,459
doc_11	0,444	0,449
doc_12	0,337	0,351
doc_13	0,367	0,387
doc_14	0,388	0,372
doc_15	0,540	0,512
doc_16	0,338	0,362
doc_17	0,568	0,561
doc_18	0,549	0,622
doc_19	0,499	0,479
doc_20	0,477	0,411
doc_21	0,445	0,426
doc_22	0,385	0,379
doc_23	0,407	0,464
doc_24	0,381	0,400
doc_25	0,408	0,406
doc_26	0,394	0,425
doc_27	0,452	0,497
doc_28	0,415	0,380
doc_29	0,387	0,438
doc_30	0,392	0,402

CONCLUSION

We have analyzed 30 scientific texts translated by GPT models and human translators.

In assessing the efficiency of automated translation systems ChatGPT-4 and ChatGPT-3.5 in comparison with human translation capabilities, the performance of ChatGPT-4, as quantified by the COMET-22 metric achieves an accuracy of 92% and ChatGPT-3.5 achieves an accuracy of 87%. In the BLEURT-20 metric the performances of ChatGPT-4 and ChatGPT-3.5 models are 62% and 54% respectively.

It should be underlined, that COMET-22 neural fine-tuned metric is a more powerful metric, trained on a wider and more diverse datasets and focusing on the comprehensive semantic similarity between source text and translation. As a result, it is less sensitive to minor errors or deviations in translation. The indicators shown in the schemes above are higher, than those of BLEURT-20.

BLEURT-20 in its turn, is more appropriate for tasks where a high degree of linguistic knowledge and stylistic accuracy is required.

In general, ChatGPT-4 demonstrates relatively high-quality translations in the meaning and choice of a word in simple extended sentences, and complex sentences, in translation of simple terms as well.

However, the primary errors identified in the comparison between human translated and machine translated texts involve incorrect translations and the repetition, non-equivalent and misinterpreted idiomatic expressions, terms, avoiding grammatical rules of gender agreement in the sentences, which translators, specialized in a specific field, can easily overcome such errors.

Due to the progress in computational linguistics and machine learning, these innovative systems i.e. ChatGPT-3.5 and ChatGPT-4 can interpret and translate the complex nuances of language.

Nevertheless, it's important to remember that automated systems often face rather difficult idiomatic expressions, cultural references, and the deep contexts that typically demand human intervention because of intuitively specific cultural knowledge.

Still, it's worth mentioning that the 92% accuracy rate of ChatGPT-4 demonstrates strong general translation abilities, similar to the rate score presented in "The AI Index 2024 Annual Report" (Maslej et.al, 2024).

We have to highlight the limitations where human oversight and intervention are still necessary in the translation process. But we hope that performance capabilities of neural machine translation systems will considerably improve in the nearest future.

P.S. When we finally finished the current thesis, the ChatGPT-4o launched.

Definition of Key Terms

Artificial Intelligence (AI): Artificial Intelligence refers to the simulation of human intelligence in machines that can perform tasks that typically require human intelligence, such as learning, problem-solving, perception, and decision-making.

AI-assisted Translation: AI-assisted translation involves the use of artificial intelligence tools, such as ChatGPT-4, to aid human translators in the translation process. The AI model assists in generating translations, which are then reviewed and refined by human translators.

An assistant tool, also known as an AI assistant or virtual assistant, is a software program or application that uses artificial intelligence (AI) and natural language processing (NLP) technologies to interact with users and perform various tasks. Assistant tools are designed to understand and respond to human language, making them capable of handling voice commands or text-based inputs.

An API, or application programming interface, is a set of defined rules that enable different applications to communicate with each other. It acts as an intermediary layer that processes data transfers between systems.

Cultural References: Cultural references are elements in language that are tied to specific cultural contexts, customs, traditions, or historical events. Understanding cultural references is crucial for accurate translation and effective communication across cultures.

COMET a specific model and framework designed for evaluating and fine-tuning machine translation systems. COMET stands for "Cross-lingual Optimized Metric for Evaluation of Translation." It is primarily used to assess the quality of machine translation output by comparing it to human-generated translations and providing a metric for evaluation. It works by training a neural network-based model on large-scale parallel corpora, which are collections of texts in two or more languages with corresponding translations. This model learns to rank different translations based on their quality. It can evaluate the performance of machine translation systems across

various language pairs, domains, and text types. It can be used for fine-tuning machine translation models. By comparing the model's translations to high-quality reference translations, it helps optimize the model's parameters to improve translation quality. COMET is a valuable tool in the field of machine translation, aiding researchers and developers in assessing and enhancing the performance of translation systems.

Chatbot is an AI-based conversational agent that can interact with users in natural language. GPT can be used as the underlying model for Chatbot's, enabling more human-like and contextually relevant responses.

Domain-Specific Terminologies: Domain-specific terminologies are specialized terms and jargon that are specific to a particular field or domain, such as scientific, medical, legal, or technical terminologies.

DeepL is a powerful AI-based translation tool that offers users a wide range of features and options.

Data Security: involves the protection of sensitive information from unauthorized access, disclosure, or alteration, ensuring the confidentiality and integrity of data used in AI-assisted translation.

Dalle-E3- text-to-image models developed by OpenAI using deep learning methodologies to generate digital images from natural language descriptions, called "prompts."

Ethical Considerations: Ethical considerations in the context of AI-driven translation involve the ethical implications of using AI language models in sensitive areas, data privacy, potential biases in translations, and ensuring responsible use of AI technologies.

"Fine-tuning" is a process in machine learning and neural network training where a pre-trained model is further trained on a specific task or dataset to improve its performance and adapt it to the nuances of the target domain. This process involves adjusting the model's parameters based on the new data, allowing it to specialize and refine its predictions for the particular task at hand. Fine-tuning builds upon the knowledge gained during the initial pre-training phase and tailors the model's capabilities to make more accurate predictions or classifications in a specific context.

FLORES-101 evaluation benchmark, consisting of 3001 sentences extracted from English Wikipedia and covering a variety of different topics and domains. These sentences have been translated in 101 languages by professional translators through a carefully controlled process.

High resource languages are languages for which many data resources exist, making possible the development of machine-learning based systems for these languages.

InstructGPT a framework that allows for instruction fine-tuning of a pre-trained language model based on Reinforcement Learning from Human Feedback (RLHF)

Low-resource languages are languages without a lot of training data, with none or very few resources available. This is the case for some extinct or near-extinct languages and many local dialects.

LLMs- Open-sourced large language models refer to the practice of making large language models, such as GPT-3, GPT-2, and similar models, publicly available for use, research, and development by the broader community. These models are trained on vast amounts of text data and are capable of performing a variety of natural language processing tasks, including text generation, translation, summarization, and more.

Machine Translation (MT): Machine Translation is the process of using computer algorithms to automatically translate text or speech from one language to another. It aims to bridge language barriers and enable communication between people who speak different languages.

Natural Language Processing (NLP): Natural Language Processing is a branch of artificial intelligence that focuses on enabling computers to understand, interpret, and generate human language. It involves the use of algorithms and computational linguistics to analyze and process natural language data.

Neural network is a type of artificial intelligence model inspired by the structure and function of the human brain's interconnected neurons. It consists of layers of interconnected nodes, known as neurons, organized in different layers, such as input, hidden, and output layers. Each neuron performs simple computations on its input data and passes the results to other neurons.

PARROT-Protein Analysis using Recurrent neural networks On Training data. PARROT is designed to be a general framework for training machine learning networks on large protein datasets, then using the trained network to make predictions on new protein sequences.

Pivot prompting is a technique used in natural language processing (NLP) and machine learning, particularly in the context of language models like GPT (Generative Pre-Trained Transformer), to guide the generation of text outputs. In pivot prompting, an initial prompt is used to generate

intermediate or "pivot" text, which is then used as a context or input for generating the final desired output. The purpose of pivot prompting is to guide the language model's generation process in a more controlled and specific manner. By providing an intermediate step, researchers and users can fine-tune the context and steer the model towards generating text that aligns with their desired outcome.

Pivot prompting is often used to address challenges such as generating accurate and contextually appropriate responses in conversation or achieving higher-quality translations in machine translation tasks. It allows users to guide the language model's output more effectively by breaking down the generation process into multiple steps and using the intermediate pivot text as a bridge between the source and target outputs.

RLHF-Reinforcement Learning from Human Feedback

Stylistic and Semantic Quality: Stylistic quality refers to the coherence, consistency, and fluency of the translated text, ensuring it reads naturally and professionally. Semantic quality refers to the accuracy and precision in conveying the intended meaning and preserving the original content's context.

Translation Quality Assurance (QA) refers to the systematic processes and measures implemented to ensure that translated content meets predefined quality standards and accurately conveys the intended meaning of the source text. QA play a crucial role in ensuring linguistic accuracy and cultural relevance. QA is a critical step in the translation process to verify the accuracy, consistency, and overall quality of the translated material. It involves various strategies and techniques to identify and rectify errors or issues that may arise during translation.

The WMT19 Bio test set refers to a specific dataset used in the context of the Workshop on Machine Translation (WMT) 2019.

The transformer architecture is a type of neural network architecture that has been highly successful in various natural language processing tasks, including language generation, translation, and understanding.

WMT- is an annual conference and shared task series that focuses on evaluating and advancing the field of machine translation. It provides researchers with benchmark datasets and a platform to compare different machine translation systems.

REFERENCES

1. Abbasbeyli, E. (2021). The role of the translation bureau (Tercümə Odası) in preparing translators, interpreters, and diplomats of the Ottoman Empire. CITI Science, CITI University (Ulaanbaatar, Mongolia), 1, 16-22. Retrieved from "[THE ROLE OF THE TRANSLATION BUREAU \(TERCÜME ODASI\) IN PREPARING TRANSLATORS, INTERPRETERS, AND DIPLOMATS OF THE OTTOMAN EMPIRE - PDF Free Download \(docplayer.net\)](#)
2. Acta Naturae. (2009). Acta Naturae is a new international journal on life sciences based in Moscow, Russia (published since April 2009).[Acta Naturae](#)
3. Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Xu, Y., & Fung, P. (2023). A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity. arXiv. Retrieved from [2302.04023 \(arxiv.org\)](#)
4. Berger, A. L., Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., GiUett, J. R., Lafferty, J. D., Mercer, R. L., Printz, H., Urei, L. (IBM Thomas J. Watson Research Center). (n.d.). The Candide System for Machine Translation.[H94-1028.pdf - Yandex Documents](#)
5. Bishop, J. M. (2014). History and Philosophy of Neural Networks. Retrieved from [Bishop-2014.pdf - Yandex Documents](#)
6. Cabré, M. T. (1998). Terminology: Theory, Methods and Applications. Terminology and Lexicography Research and Practice. Epdf.tips. Retrieved from [Terminology: Theory, Methods, and Applications \(Terminology and Lexicography Research and Practice\) - PDF Free Download \(epdf.tips\)](#)
7. Freitag, M., Kocmi, T., et al. (2022). Proceedings of the Seventh Conference on Machine Translation (WMT), 46-68. [2022.wmt-1.2.pdf - Yandex Documents](#)
8. Gill, S. S., & Kaur, R. (2023). ChatGPT: Vision and Challenges. Preprint submitted to IoTCPS Elsevier. Retrieved from [Gill ChatGPT: Vision and challenges 2023 Published.pdf?sequence=2 \(yandex.ru\)](#)
9. Goodfellow, I., Bengio, Y., & Courville, A. (2016). The Deep Learning textbook. Retrieved from [DeepLearning \(deeplearningbook.org\)](#)
10. Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural Computation. Retrieved from [A fast learning algorithm for deep belief nets.pdf - Yandex Documents](#)

11. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9, 1735–1780. Retrieved from [1997-hochreiter.pdf - Yandex Documents](#)
12. Holmes, J. S. (1979). The Name and Nature of Translation Studies. [Holmes.pdf - Яндекс Документы \(yandex.ru\)](#)
13. Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79, 2554–2558. Retrieved from [Hopfield82.pdf - Яндекс Документы \(yandex.ru\)](#)
14. Hopfield, J. J. (1986). Simple Neural Optimization Networks: An A/D Converter, Signal Decision Circuit, and a Linear Programming Circuit. *IEEE Transactions on Circuits and Systems*, 33(5), 533-541. Retrieved from [\(PDF\) Hopfield, J.J.: Simple Neural Optimization Networks: An A/D Converter, Signal Decision Circuit, and a Linear Programming Circuit. IEEE Transactions on Circuits and Systems 33\(5\), 533-541 \(researchgate.net\)](#)
15. Hugging Face. (n.d.). [Hugging Face – The AI community building the future.](#)
16. Hutchins, J. (2003). Two precursors of machine translation: A forgotten pioneer of mechanical translation. *Machine Translation*, 15(3), 187–221. Retrieved from [IJT-2004-Hutchins.pdf - Yandex Documents](#)
17. Hutchins, W. J. (2004). The Georgetown-IBM experiment demonstrated in January 1954. [AMTA-2004-Hutchins.pdf - Yandex Documents](#)
18. Işım, Ç., & Balcıoğlu, Y. S. (2023). ChatGPT: Performance of Translate. [CHATGPT-PERFORMANCE-OF-TRANSLATE.pdf \(researchgate.net\)](#)
19. Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z. (2023). Is ChatGPT a Good Translator? A Preliminary Study. *arXiv*. Retrieved from [tech_chatgpt_arxiv.pdf - Yandex Documents](#)
20. Kalchbrenner, N., Danihelka, I., & Graves, A. (2015). Grid Long Short-Term Memory. *arXiv*. Retrieved from [1507.01526v1 \(arxiv.org\)](#)
21. Khoshafah, F. (2023). ChatGPT for Arabic-English Translation: Evaluating the Accuracy. [v1_covered.pdf - Yandex Documents](#)
22. Kocmi, T., & Federmann, C. (2023). Large Language Models Are State-of-the-Art Evaluators of Translation Quality. *arXiv*. Retrieved from [2302.14520v1 \(arxiv.org\)](#)
23. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *NeurIPS*. Retrieved from [ImageNet Classification with Deep Convolutional Neural Networks \(neurips.cc\)](#)

24. Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models. arXiv. Retrieved from [[2304.01852v1](#)] [Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models \(arxiv.org\)](#)
25. Martin, C. D. (1995). ENIAC: The Press Conference That Shook the World. IEEE Technology and Society Magazine. Retrieved from [eniac2.pdf - Yandex Documents](#)
26. Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J. C., Shoham, Y., Wald, R., & Clark, J. (2024). The AI Index 2024 Annual Report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. Retrieved from [HAI AI-Index-Report-2024.pdf \(stanford.edu\)](#)
27. McCulloch, W. S., & Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. Retrieved from [mcculloch.logical.calculus.ideas.1943.pdf - Yandex Documents](#)
28. McKenna, N., Li, T., Cheng, L., Hosseini, M. J., Johnson, M., & Steedman, M. (2023). Sources of Hallucination by Large Language Models on Inference Tasks. Findings of the Association for Computational Linguistics: EMNLP 2023, 2758-2774. Retrieved from [2023.findings-emnlp.182.pdf - Yandex Documents](#)
29. Nida, E. A. (1964). Toward a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating. Adler's Foreign Books Inc. Retrieved from [\(PDF\) Eugene a. Nida-Toward a Science of Translating With Special Reference to Principles and Procedures Involved in Bible Translating-Adler's Foreign Books Inc \(1964\) - PDFSLIDE.US](#)
30. Nielsen, M. A. (2015). Neural Networks and Deep Learning. Retrieved from [neuralnetworksanddeeplearning.pdf - Yandex Documents](#)
31. OpenAI. (n.d.). Retrieved from <https://chat.openai.com/>
32. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askill, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. NeurIPS. Retrieved from [b1efde53be364a73914f58805a001731-Paper-Conference.pdf - Yandex Documents](#)
33. Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (n.d.). BLEU: a Method for Automatic Evaluation of Machine Translation. IBM T. J. Watson Research Center. Retrieved from [P02-1040.pdf - Yandex Documents](#)

34. Pickens, E., & Batjargal, B. (2020). Hidden Markov Models. [hmm_draft.pdf - Yandex Documents](#)
35. Pigott, I. M. (1986). Machine translation as an integral part of the electronic office environment. In C. Picken (Ed.), *Translating and the Computer 7*. Proceedings of a conference, 14-15 November. London: Aslib. Retrieved from [Machine translation as an integral part of the electronic office environment \(aclanthology.org\)](#)
36. Pollack, J. B. (1988). Book Review: Marvin L. Minsky and Seymour A. Papert, *Perceptrons: An Introduction to Computational Geometry, Expanded Edition*. Cambridge, MA: MIT Press, 292 pp. [perceptron.pdf - Yandex Documents](#)
37. Rei, R., et al. (2022). Unbabel-IST 2022 Submission for the Metrics Shared Task. Proceedings of the Seventh Conference on Machine Translation (WMT), 578–585. Retrieved from [2022.wmt-1.52.pdf \(aclanthology.org\)](#)
38. Rosenblatt, F. (1958). [Rosenblatt_Perceptron.pdf - Yandex Documents](#)
39. Schwab, K. (2016). *The 4th Industrial Revolution*. Geneva: World Economic Forum. Retrieved from [Schwab-Klaus-The-Fourth-Industrial-Revolution-2016.pdf - Yandex Documents](#)
40. Sellam, T., Das, D., & Parikh, A. P. (2020). BLEURT: Learning Robust Metrics for Text Generation. arXiv. Retrieved from [2004.04696.pdf - Yandex Documents](#)
41. Siu, S. C. (2023). ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation. ResearchGate.net. Retrieved from [\(PDF\) ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation \(researchgate.net\)](#)
42. Suleyman, M., & Bhaskar, M. (2023). *The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma*. Crown Publishing Group. Retrieved from [\[PDF\] \[EPUB\] The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma Download \(oceanofpdf.com\)](#)
43. Terna, P., D'Acunto, G., & Caselle, M. (2016). A Deep Learning Model to Forecast Financial Time-Series. Retrieved from [d'acunto.pdf - Yandex Documents](#)
44. The Automatic Language Processing Advisory Committee (ALPAC). (1966). [ALPAC-1966.pdf - Yandex Documents](#)
45. Tuomi, I. (2018). The Impact of Artificial Intelligence on Learning, Teaching, and Education. In M. Cabrera Giraldez, R. Vuorikari, & Y. Punie (Eds.), EUR 29442 EN. Publications Office of the European Union. Retrieved from [JRC Publications Repository - The Impact of Artificial Intelligence on Learning, Teaching, and Education \(europa.eu\)](#)

46. Weaver, W. (1949). Retrieved from [Weaver-1949.pdf - Yandex Documents](#)
47. Кочергин, В. И. (2008). Англо-Русский Толковый Научно-Технический Словарь в двух томах. Томск: Яндекс Документы. Retrieved from [2017-0178.pdf - Яндекс Документы \(yandex.ru\)](#)
48. Лейбниц, Г. В. (2005). Письма и Эссе о Китайской философии и двоичной системе исчисления (Предисловие, переводы и примечания В. М. Яковлева). Retrieved from [86200644-Лейбниц-Г-В-Письма-и-эссе-о-китайской-философии-и-двоичной-системе-исчисления-2005.pdf \(vk.com\)](#)
49. Пройдаков, Э., & Теплицкий, Л. (2021). Англо-русский толковый словарь терминов открытым данным и программному обеспечению с открытыми кодом. Retrieved from [ODD-glossary_EN-RU_preview.pdf \(infoculture.ru\)](#)
50. Циммерман, М., & Веденева, К. (2003). Русско-английский научно-технический Словарь Переводчика. Retrieved from [tsymmerman_rus-en_tech.pdf - Яндекс Документы \(yandex.ru\)](#)

References of samples, sourced from Acta Naturae

1. [The Pharmaceutical Industry in Russia: Reality and Prospects - PDF \(English\) - Gordeev - Acta Naturae](#)
2. [The Pharmaceutical Industry in Russia: Reality and Prospects - PDF \(Russian\) - Gordeev - Acta Naturae](#)
3. [RIAC :: ‘Three Poles of the Cold’: Myth or Reality? \(russiancouncil.ru\)](#)
4. [РСМД :: Концепция «Трех полюсов холода»: миф или реальность? \(russiancouncil.ru\)](#)
5. [Животные модели митохондриальных заболеваний, вызванных мутациями в ядерных генах - PDF \(Английский\) - Аверина - Acta Naturae](#)
6. [Животные модели митохондриальных заболеваний, вызванных мутациями в ядерных генах - PDF \(Русский\) - Аверина - Acta Naturae](#)
7. [Том 15, № 4 \(2023\) - PDF \(Английский\) - Acta Naturae](#)
8. [Soft X-ray Microscopy in Cell Biology: Current Status, Contributions and Prospects - PDF \(Russian\) - Golyshev - Acta Naturae](#)
9. [Том 15, № 4 \(2023\) - PDF \(Английский\) - Acta Naturae](#)
10. [Том 15, № 4 \(2023\) - PDF \(Русский\) - Acta Naturae](#)
11. [Рюриковичи: первый опыт реконструкции генетического облика правящего рода средневековой Руси по данным палеогеномики - PDF \(Английский\) - Жур - Acta Naturae](#)

12. Рюриковичи: первый опыт реконструкции генетического облика правящего рода средневековой Руси по данным палеогеномики - PDF (Русский) - Жур - Acta Naturae
13. Анализ ассоциаций гаплотипов гена Tgfb1 с болезнями печени у детей - PDF (Английский) - Курабекова - Acta Naturae
14. Анализ ассоциаций гаплотипов гена Tgfb1 с болезнями печени у детей - PDF (Русский) - Курабекова - Acta Naturae
15. Влияние делеции гена at1 на патогенные и иммуногенные свойства вируса осповакцины - PDF (Английский) - Якубицкий - Acta Naturae
16. Влияние делеции гена at1 на патогенные и иммуногенные свойства вируса осповакцины - PDF (Русский) - Якубицкий - Acta Naturae
17. О биоразнообразии микробиома воздуха - PDF (Английский) - Наумова - Acta Naturae
18. О биоразнообразии микробиома воздуха - PDF (Русский) - Наумова - Acta Naturae
19. Сахарный диабет 2 типа: особенности патогенеза и экспериментальные модели на грызунах - PDF (Английский) - Гвазава - Acta Naturae
20. Сахарный диабет 2 типа: особенности патогенеза и экспериментальные модели на грызунах - PDF (Русский) - Гвазава - Acta Naturae
21. Изменение профиля экспрессии микроРНК в лейкоцитах пациентов с болезнью Паркинсона - PDF (Английский) - Ардаширова - Acta Naturae
22. Изменение профиля экспрессии микроРНК в лейкоцитах пациентов с болезнью Паркинсона - PDF (Русский) - Ардаширова - Acta Naturae
23. Combining Two Technologies for Full Genome Sequencing of Human - PDF (Английский) - Skryabin - Acta Naturae
24. Combining Two Technologies for Full Genome Sequencing of Human - PDF (Русский) - Skryabin - Acta Naturae
25. Качество научных журналов и основные критерии для включения в информационную систему Web of Science компании Thomson Reuters - PDF (Английский) - Маркусова - Acta Naturae
26. Качество научных журналов и основные критерии для включения в информационную систему Web of Science компании Thomson Reuters - PDF (Русский) - Маркусова - Acta Naturae

27. Искажение популяционной статистики как результат различных методических подходов к приготовлению геномных библиотек древней ДНК - PDF (Английский) - Шарко - Acta Naturae
28. Искажение популяционной статистики как результат различных методических подходов к приготовлению геномных библиотек древней ДНК - PDF (Русский) - Шарко - Acta Naturae
29. Роль урокиназной системы в многоуровневой регуляции ниш стволовых клеток - PDF (Английский) - Дергилев - Acta Naturae
30. Роль урокиназной системы в многоуровневой регуляции ниш стволовых клеток - PDF (Русский) - Дергилев - Acta Naturae
31. Раковые стволовые клетки: пластичность против терапии - PDF (Английский) - Виноградова - Acta Naturae
32. Раковые стволовые клетки: пластичность против терапии - PDF (Русский) - Виноградова - Acta Naturae
33. Экспериментальное моделирование травмы спинного мозга у лабораторных крыс - PDF (Английский) - Минаков - Acta Naturae
34. Экспериментальное моделирование травмы спинного мозга у лабораторных крыс - PDF (Русский) - Минаков - Acta Naturae
35. Возможности использования плюрипотентных стволовых клеток для восстановления поврежденного пигментного эпителия сетчатки глаза - PDF (Английский) - Харитонов - Acta Naturae
36. Возможности использования плюрипотентных стволовых клеток для восстановления поврежденного пигментного эпителия сетчатки глаза - PDF (Русский) - Харитонов - Acta Naturae
37. Вклад генов главного комплекса гистосовместимости класса II в предрасположенность к аутоиммунным заболеваниям - PDF (Английский) - Захарова - Acta Naturae
38. Вклад генов главного комплекса гистосовместимости класса II в предрасположенность к аутоиммунным заболеваниям - PDF (Русский) - Захарова - Acta Naturae
39. Катехоламины в слезной жидкости как маркеры болезни Паркинсона: клиническое и экспериментальное исследование - PDF (Английский) - Ким - Acta Naturae
40. Катехоламины в слезной жидкости как маркеры болезни Паркинсона: клиническое и экспериментальное исследование - PDF (Русский) - Ким - Acta Naturae

41. Рекомбинантные антитела к гликопротеину вируса Эбола - PDF (Английский) - Панина - Acta Naturae
42. Рекомбинантные антитела к гликопротеину вируса Эбола - PDF (Русский) - Панина - Acta Naturae.
43. Bioreactor-Based Tumor Tissue Engineering - PDF (English) - Guller - Acta Naturae
44. Bioreactor-Based Tumor Tissue Engineering - PDF (Russian) - Guller - Acta Naturae
45. Hyaluronic Acid in Vascular and Immune Homeostasis during Normal Pregnancy and Preeclampsia - PDF (English) - Ziganshina - Acta Naturae
46. Hyaluronic Acid in Vascular and Immune Homeostasis during Normal Pregnancy and Preeclampsia - PDF (Russian) - Ziganshina - Acta Naturae
47. Epigenetics of Ancient DNA - PDF (English) - Zhenilo - Acta Naturae
48. Epigenetics of Ancient DNA - PDF (Russian) - Zhenilo - Acta Naturae
49. Cleavage of Human Embryos: Options and Diversity - PDF (English) - Doronin - Acta Naturae
50. Cleavage of Human Embryos: Options and Diversity - PDF (Russian) - Doronin - Acta Naturae
51. Analysis of the Domains of Hepatitis C Virus Core and NS5A Proteins that Activate the Nrf2/ARE Cascade - PDF (English) - Smirnova - Acta Naturae
52. Analysis of the Domains of Hepatitis C Virus Core and NS5A Proteins that Activate the Nrf2/ARE Cascade - PDF (Russian) - Smirnova - Acta Naturae
53. Epithelial–mesenchymal transition: role in cancer progression and the perspectives of antitumor treatment - PDF (English) - Gaponova - Acta Naturae
54. Epithelial–mesenchymal transition: role in cancer progression and the perspectives of antitumor treatment - PDF (Russian) - Gaponova - Acta Naturae
55. Cognitive Tagging of Neurons: CREMediated Genetic Labeling and Characterization of the Cells Involved in Learning and Memory - PDF (English) - Ivashkina - Acta Naturae
56. Cognitive Tagging of Neurons: CREMediated Genetic Labeling and Characterization of the Cells Involved in Learning and Memory - PDF (Russian) - Ivashkina - Acta Naturae
57. Low-Molecular-Weight NGF Mimetic Corrects the Cognitive Deficit and Depression-like Behavior in Experimental Diabetes - PDF (English) - Ostrovskaya - Acta Naturae
58. Low-Molecular-Weight NGF Mimetic Corrects the Cognitive Deficit and Depression-like Behavior in Experimental Diabetes - PDF (Russian) - Ostrovskaya - Acta Naturae
59. Analysis of the TREC and KREC Levels in the Dried Blood Spots of Healthy Newborns with Different Gestational Ages and Weights - PDF (English) - Cheremokhin - Acta Naturae

60. Analysis of the TREC and KREC Levels in the Dried Blood Spots of Healthy Newborns with Different Gestational Ages and Weights - PDF (Russian) - Cheremokhin - Acta Naturae
Evaluation by metrics- Код Comet.ipynb - Colab (google.com)

Appendix

Acknowledgements

I really deeply appreciate and am profoundly grateful to my supervisor, Prof. Dr. Huseynagha Rzayev, for his important contribution, whose expertise and insightful guidance have been instrumental in shaping both this research and my growth.

Special thanks also go to Dr. Milana Abbasova, for her patience, understanding our challenges and doing every effort in helping each of us in overcoming those difficulties.

I'm also thankful to our instructors in the MA program: Prof. Yaser Hadidi, Dr. Davud Kuhi, PhD candidate Ilham Taghiyev, PhD candidate Zamira Gurbanova and all those who were helpful this or that way in completing this research.

Many thanks to my Department of Graduate School of Science, Art and Technology for providing environment conducive to academic research and personal development.

I also owe a great deal of gratitude to my classmates, whose collaborative spirit have enriched this journey.

Last but not least, I must express my deepest appreciation to my family, who have been my best supporters and friends, and especially to my friend Milena, who has provided constant support throughout my academic pursuits. This achievement is not only mine but also yours.

To all whom I love with the deepest respect and gratitude.

ABSTRACT

You may not believe but we all have already surrounded by different technologies and have close connection with AI (Artificial Intelligence). For instance, we follow the needed destination using the voice-guided navigation of car's GPS (Global Positioning System) or ask Alice or Siri voice assistants to dial number from the friends' list. In various fields, we try to create automated advanced systems and applications to reduce our work. We have created robotic vacuum cleaners, wash machines, "smart house" systems, which distantly could be controlled by phone or PC, and we go forward, working on 3D printed Animal and Human Prosthetics to make happier disabled people.

Undoubtedly, the Translation Science also undergoes significant changes and we are all witnesses to the inventions that people have waited for and tried to achieve for several decades. In 2022, the world appreciated for the large language model ChatGPT-3.5, one year later for ChatGPT-4, introduced and intensively developed by OpenAI company. These Generative Pretrained Transformer (GPT) language models trained on a huge network dataset, can generate texts, the latest language models translate myriad languages, solve math problems, answer questions, create imagines and just can be a good interlocutor, carrying on a dialogue.

This paper aims to examine and compare translations of scientific texts from English to Russian performed by ChatGPT-3.5, ChatGPT-4, and human translators. The study conducts a comparative quality analysis of these translations and utilizes the neural-based machine translation evaluation metrics COMET-22 and BLEURT-20.

A key goal of this paper is to identify errors, and shed the light on the strengths and limitations primarily of ChatGPT-4 model.

Additionally, the paper finds out improvements of ChatGPT-4 capabilities and difference with its predecessor ChatGPT-3.5, and estimates their performance in comparison to human translation.

Using mixed-method empirical research, this study analyzed 30 scientific articles, assessing the efficiency and capability of GPT models.

Key words: AI (Artificial Intelligence), neural networks, deep learning, natural language processing (NLP), large language model(LLM), COMET-22, BLEURT-20 metrics, ChatGPT-4, ChatGPT-3.5, machine translation, human translation, scientific texts

XÜLASƏ

Əslində inanmasaq da, artıq hamımız müxtəlif texnologiyaların əhatəsində və Süni İntellekt (AI) ilə yaxın əlaqədəyik. Məsələn, avtomobilin GPS (Qlobal Mövqeləşdirmə Sistemi) vasitəsilə səsli naviqasiyadan istifadə edərək lazım olan istiqaməti izləyir, həmçinin Alice və ya Siri səsli köməkçilərindən dostlarımızın siyahısından nömrə yığmağı xahiş edirik. Müxtəlif sahələrdə işimizi azaltmaq üçün avtomatlaşdırılmış qabaqcıl sistemlər və tətbiqlər yaratmağa çalışırıq. Biz robot tozsoranlar, paltaryuyan maşınlar yaratmışıq, telefon və ya PC vasitəsilə uzaqdan idarə oluna bilən “ağıllı ev” sistemləri yaratmışıq və əlilləri daha xoşbəxt etmək üçün 3D çap olunmuş Heyvan və İnsan Protezləri üzərində işləyirik.

Şübhəsiz ki, Tərcümə Elmi də mühüm dəyişikliklərə məruz qalır və biz hamımız insanların on illər boyu gözlədikləri və əldə etməyə çalışdıqları ixtiraların şahidiyik. 2022-ci ildə dünya böyük dil modeli olan ChatGPT-3.5-i, bir il sonra isə OpenAI şirkəti tərəfindən təqdim olunan və intensiv inkişaf etdirilən ChatGPT-4-ü yüksək qiymətləndirdi. Bu Generative Pretrained Transformer (GPT) dil modelləri nəhəng şəbəkə verilənlər bazasında öyrədilmişdir, mətnlər yarada bilər, son dil modelləri saysız-hesabsız dillərə tərcümə edir, riyazi problemləri həll edir, suallara cavab verir, təsvirlər yaradır, hətta dialoq aparən yaxşı bir həmsöhbət ola bilər.

Bu məqalənin məqsədi ChatGPT-3.5, ChatGPT-4 və insan tərcüməçilər tərəfindən ingilis dilindən rus dilinə edilən elmi mətnlərin tərcümələrini araşdırmaq və müqayisə etməkdir. Tədqiqat bu tərcümələrin müqayisəli keyfiyyət analizini aparır və neyron əsaslı maşın tərcümə qiymətləndirmə metrikləri COMET-22 və BLEURT-20-dən istifadə edir.

Bu məqalənin əsas məqsədi səhvləri müəyyən etmək və ilk növbədə ChatGPT-4 modelinin güclü və zəif tərəflərinə diqqət çəkməkdir.

Bundan əlavə, məqalə ChatGPT-4 imkanlarının təkmilləşdirilməsini və onun sələfi ChatGPT-3.5 ilə fərqi aşkar edir və onların insan tərcüməsi ilə müqayisədə performansını qiymətləndirir.

Qarışıq metoddan istifadə edərək aparılan empirik tədqiqat, GPT modellərinin səmərəliliyini və imkanlarını qiymətləndirərək 30 elmi məqaləni təhlil edir.

Açar sözlər: Süni İntellekt (AI), neyron şəbəkələr, dərin öyrənmə, təbii dil işlənməsi (NLP), böyük dil modeli (LLM), COMET-22, BLEURT-20 metrikləri, ChatGPT-4, ChatGPT-3.5, maşın tərcüməsi, insan tərcüməsi, elmi mətnlər