

Leyla Vaqif qızı ƏLİZADƏ

Azərbaycan Turizm və Menecment Universiteti, filologiya ü.f.d., Bakı, Azərbaycan

E-mail: leyla.alizada@atmu.edu.az

ORCID ID: 0000-0002-8360-973X

SÜNİ İNTELLEKT ƏSASLI AVTOMATİK QIYMƏTLƏNDİRMƏ SİSTEMLƏRİ: YAZILI VƏ ŞİFAHİ TƏRCÜMƏNİN KEYFİYYƏTİNİN ÖLÇÜLMƏSİ ÜZRƏ YANAŞMALAR, ETİBARLILIQ VƏ TƏTBİQ PERSPEKTİVLƏRİ

Xülasə

Bu tədqiqatın məqsədi süni intellekt əsaslı avtomatik qiymətləndirmə sistemlərinin yazılı və şifahi tərcümənin keyfiyyətinin ölçülməsində rolunu, onların metodoloji əsaslarını və etibarlılıq səviyyəsini təhlil etməkdir. Tədqiqatda müqayisəli və analitik metodlardan, eləcə də mövcud elmi ədəbiyyatın sistemli icmalından istifadə olunmuşdur. Araşdırma BLEU, METEOR, TER, BERTScore və COMET kimi klassik və neyron əsaslı metrikaların, həmçinin LLM əsaslı qiymətləndirmə yanaşmalarının imkanlarını və məhdudiyyətlərini əhatə edir. Nəticələr göstərir ki, süni intellekt modelləri tərcümə keyfiyyətinin qiymətləndirilməsində yüksək potensiala malik olsa da, onların tam obyektivliyi hələ də insan ekspertizasını tam əvəz etmir. Xüsusilə şifahi tərcümədə multimodal amillərin nəzərə alınması qiymətləndirmənin daha mürəkkəb olduğunu göstərir.

Açar sözlər: süni intellekt, tərcümə keyfiyyəti, avtomatik qiymətləndirmə, BLEU, şifahi tərcümə.

UOT: 004.8:81'25:004.912

DOI: <https://doi.org/10.54414/IKNF5453>

Giriş

Rəqəmsal texnologiyaların sürətli inkişafı və süni intellektin müxtəlif sahələrə inteqrasiyası tərcümə prosesinin mahiyyətini əhəmiyyətli dərəcədə dəyişdirmişdir. Müasir dövrdə tərcümə yalnız mətnin bir dildən digərinə ötürülməsi kimi deyil, həm də kompleks informasiya emalı prosesi kimi qəbul edilir [1, s.45]. Bu prosesdə maşın tərcüməsi, post-redaktə və avtomatik qiymətləndirmə kimi mərhələlər getdikcə daha mühüm rol oynamağa başlamışdır [2, s.112].

Tərcümə keyfiyyətinin qiymətləndirilməsi uzun müddət ərzində əsasən insan ekspertlərinin subyektiv qərarlarına əsaslanmışdır. Bu yanaşma semantik uyğunluq, üslubi adekvatlıq, terminoloji düzgünlük və kommunikativ məqsəd kimi meyarların nəzərə alınmasını tələb etsə də, qiymətləndirmə nəticələrinin fərqliliyi və yüksək resurs sərfi bu metodun məhdudiyyətlərini ortaya çıxarmışdır [3, s.67]. Xüsusilə böyük həcmli tərcümə korpuslarının qiymətləndirilməsi zamanı insan

əsaslı yanaşmaların səmərəliliyi ciddi şəkildə azalır [4, s.98].

Son illərdə süni intellekt əsaslı avtomatik qiymətləndirmə sistemlərinin inkişafı bu sahədə yeni mərhələnin başlanmasına səbəb olmuşdur. Bu sistemlər tərcümə mətnlərinin keyfiyyətini statistik, semantik və neyron şəbəkə əsaslı modellər vasitəsilə qiymətləndirir. Ənənəvi yanaşmalardan fərqli olaraq, bu modellər yalnız səthi uyğunluqları deyil, həm də dərin məna əlaqələrini analiz etməyə çalışır [5, s.134].

Avtomatik qiymətləndirmə sahəsində ilk geniş yayılmış göstəricilərdən biri BLEU metrikası olmuşdur. Bu metod namizəd tərcümə ilə istinad tərcüməsi arasındakı n-gram uyğunluğunu ölçərək keyfiyyəti müəyyən edir. Lakin BLEU-nun əsas çatışmazlığı onun semantik məzmunu tam əks etdirməməsi və sinonim ifadələri nəzərə almamasıdır [6, s.201]. Bu səbəbdən sonrakı illərdə METEOR, TER və chrF kimi alternativ metrikalar



hazırlanmışdır ki, bunlar müəyyən dərəcədə semantik yaxınlığı da nəzərə alır [7, s.89].

Neyron şəbəkələrə əsaslanan yeni nəsil qiymətləndirmə modelləri bu sahədə ciddi irəliləyiş yaratmışdır. Xüsusilə BERTScore kimi metodlar mətnlər arasında semantik oxşarlığı kontekstual embeddinglər vasitəsilə qiymətləndirərək insan qiymətləndirmələrinə daha yaxın nəticələr təqdim edir [8, s.156]. Bu yanaşma tərcümənin yalnız formal deyil, həm də məna səviyyəsində qiymətləndirilməsinə imkan yaradır.

Daha inkişaf etmiş modellərdən biri olan COMET sistemi çoxdillli neyron şəbəkələrdən istifadə edərək tərcümə, mənbə mətni və istinad tərcüməsi arasında dərin əlaqələri analiz edir. Aparılmış tədqiqatlar göstərir ki, COMET insan ekspertlərinin verdiyi qiymətləndirmələrlə yüksək uyğunluq nümayiş etdirir və müxtəlif dil cütlərində sabit nəticələr verir [9, s.77]. Bu xüsusiyyət onu müasir maşın tərcüməsi sistemlərinin qiymətləndirilməsində əsas alətlərdən birinə çevirmişdir.

Son dövrlərdə generativ süni intellekt modellərinin (GPT, Claude, Gemini və s.) inkişafı avtomatik qiymətləndirmə prosesinə yeni yanaşmalar gətirmişdir. Bu modellər yalnız tərcümənin keyfiyyətini ölçmək deyil, həm də qiymətləndirmənin səbəblərini izah etmək qabiliyyətinə malikdir [10, s.214]. Bu isə qiymətləndirməni daha şəffaf və interpretasiya oluna bilən hala gətirir.

Eyni zamanda "LLM-as-a-Judge" yanaşması çərçivəsində böyük dil modelləri insan ekspertlərinin rolunu qismən əvəz etməyə başlamışdır. Lakin aparılan tədqiqatlar göstərir ki, bu modellər bəzən kontekstual səhvləri və terminoloji incəlikləri tam dəqiqliklə müəyyən edə bilmir və nəticələr istifadə olunan prompt strukturundan asılı olaraq dəyişə bilər [11, s.143]. Buna görə də müasir elmi yanaşma bu texnologiyaların insan ekspertizasını əvəz etməsi deyil, onu tamamlaması fikrini dəstəkləyir [12, s.59].

Beləliklə, süni intellekt əsaslı avtomatik qiymətləndirmə sistemləri tərcümə və şifahi tərcümə sahəsində yeni elmi istiqamət formalaşdırmışdır və bu sahənin gələcək inkişafı insan-Sİ əməkdaşlığı modelinin daha da güclənəcəyini göstərir.

Əsas hissə

Süni intellekt əsaslı avtomatik qiymətləndirmə sistemlərinin nəzəri əsasları.

Tərcümə keyfiyyətinin qiymətləndirilməsi tərcüməşünaslığın ən mürəkkəb və çoxsahəli istiqamətlərindən biridir. Tərcümənin keyfiyyəti yalnız məzmunun ötürülməsi ilə deyil, həm də semantik ekvivalentlik, funksional uyğunluq, üslubi adekvatlıq və terminoloji sabitlik kimi meyarlarla müəyyən edilir [1, s.45]. Bu yanaşma uzun illər ərzində əsasən insan ekspertlərinin qiymətləndirməsinə əsaslanmışdır.

Ənənəvi qiymətləndirmə metodları tərcümə keyfiyyətinin daha dəqiq və kontekstual şəkildə təhlilinə imkan yaratsa da, onların tətbiqi ciddi resurs tələb edir. Xüsusilə böyük həcmli tərcümə layihələrində ekspert qiymətləndirməsi həm zaman baxımından, həm də iqtisadi baxımdan səmərəli hesab edilmir [2, s.112]. Bundan əlavə, fərqli ekspertlərin eyni mətnə müxtəlif qiymətlər verməsi subyektivlik problemini artırır [3, s.67].

Bu problemlər avtomatik qiymətləndirmə sistemlərinin inkişafına təkan vermişdir. Avtomatik qiymətləndirmə yanaşmaları ilkin mərhələdə statistik modellərə əsaslanmış və əsas məqsəd maşın tərcüməsi nəticələrinin referans mətnlə müqayisəsi olmuşdur [4, s.89]. Bu mərhələdə əsas məqsəd insan müdaxiləsi olmadan sürətli və standartlaşdırılmış qiymətləndirmə mexanizminin yaradılması idi.

Avtomatik qiymətləndirmə sahəsində ən geniş istifadə olunan metrikalardan biri BLEU sistemidir. BLEU modeli n-gram uyğunluğuna əsaslanaraq tərcümə keyfiyyətini ölçür və uzun illər standart qiymətləndirmə göstəricisi kimi istifadə edilmişdir [5, s.201]. Lakin bu metod semantik ekvivalentliyi tam əks etdirmir və sinonim ifadələri nəzərə almır, bu da bəzi hallarda düzgün olmayan nəticələrə gətirib çıxarır [6, s.134].

Bu çatışmazlıqları aradan qaldırmaq məqsədilə METEOR və TER kimi alternativ metrikalar hazırlanmışdır. METEOR modeli sinonimlik və morfoloji uyğunluğu nəzərə alaraq daha insanabənzər nəticələr təqdim edir [7, s.88]. TER isə tərcümənin istinad mətninə uyğunlaşdırılması üçün tələb olunan redaktə əməliyyatlarının sayını hesablayır və post-



redaktə proseslərinin qiymətləndirilməsində geniş istifadə olunur [8, s.73].

Sonrakı mərhələdə neyron şəbəkələrə əsaslanan qiymətləndirmə modelləri inkişaf etdirilmişdir. Bu modellər artıq yalnız leksik uyğunluq deyil, həm də semantik oxşarlığı nəzərə almağa başlamışdır. BERTScore kimi metodlar kontekstual embeddinglərdən istifadə edərək mətnlər arasındakı dərin semantik əlaqəni ölçür və insan qiymətləndirmələrinə daha yaxın nəticələr verir [9, s.156].

Daha inkişaf etmiş yanaşmalardan biri COMET modelidir. COMET sistemi çoxdillli neyron şəbəkələr vasitəsilə mənbə mətn, tərcümə və istinad mətn arasında kompleks əlaqələri analiz edir. Aparılan tədqiqatlar göstərir ki, COMET insan ekspertlərinin verdiyi qiymətləndirmələrlə yüksək korrelyasiya nümayiş etdirir və müxtəlif dil cütlərində sabit nəticələr verir [10, s.77].

Son illərdə generativ süni intellekt sistemlərinin inkişafı qiymətləndirmə prosesinə yeni imkanlar gətirmişdir. GPT və digər böyük dil modelləri yalnız nəticəni qiymətləndirmir, həm də qiymətləndirmənin səbəblərini izah edə bilir [11, s.214]. Bu, avtomatik qiymətləndirməni daha interpretasiya oluna bilən və şəffaf edir.

Eyni zamanda “LLM-as-a-Judge” yanaşması çərçivəsində böyük dil modelləri insan ekspertlərinin funksiyalarını qismən yerinə yetirməyə başlamışdır. Lakin aparılan tədqiqatlar göstərir ki, bu modellər bəzən terminoloji incəlikləri və kontekstual səhvləri tam dəqiqliklə müəyyən edə bilmir [12, s.143]. Buna görə də müasir yanaşmalar insan ekspertizasının tam əvəz olunmasını deyil, Sİ ilə birgə işləməsini daha məqsədəuyğun hesab edir [13, s.59].

Beləliklə, süni intellekt əsaslı avtomatik qiymətləndirmə sistemləri tərcümə keyfiyyətinin ölçülməsində yeni elmi istiqamət formalaşdırmış və klassik qiymətləndirmə modellərini əhəmiyyətli dərəcədə təkmilləşdirmişdir.

Yazılı və şifahi tərcümənin süni intellekt vasitəsilə avtomatik qiymətləndirilməsi: mövcud təcrübə və problemlər

Süni intellekt əsaslı qiymətləndirmə sistemlərinin inkişafı yalnız yazılı tərcümə ilə

məhdudlaşmış, eyni zamanda şifahi tərcümə və real vaxt nitq tərcüməsi sahələrini də əhatə etməyə başlamışdır. Xüsusilə avtomatik nitq tanıma (ASR – Automatic Speech Recognition) və təbii dil emalı (NLP – Natural Language Processing) texnologiyalarının inkişafı bu istiqamətdə yeni imkanlar yaratmışdır [1, s.58].

Şifahi tərcümənin qiymətləndirilməsi yazılı tərcümə ilə müqayisədə daha mürəkkəb hesab olunur, çünki burada yalnız mətn deyil, həm də intonasiya, pauza, nitq sürəti, akustik xüsusiyyətlər və kommunikativ davranışlar nəzərə alınmalıdır [2, s.101]. Bu səbəbdən ənənəvi qiymətləndirmə metodları şifahi tərcümənin bütün xüsusiyyətlərini tam əhatə edə bilmir.

Son illərdə ASR əsaslı sistemlər şifahi tərcümənin avtomatik transkripsiyası və ilkin qiymətləndirilməsi üçün geniş istifadə olunur. Bu sistemlər nitqi mətnə çevirərək sonrakı analiz mərhələsinə hazırlayır. Lakin ASR texnologiyalarının dəqiqliyi səs keyfiyyəti, aksent, danışq sürəti və fon səs-küyündən ciddi şəkildə asılıdır [3, s.77].

Şifahi tərcümənin avtomatik qiymətləndirilməsində LLM (Large Language Models) əsaslı yanaşmalar da geniş tətbiq olunmağa başlamışdır. GPT, Claude və Gemini kimi modellər artıq yalnız yazılı mətnləri deyil, transkripsiya olunmuş nitqi də qiymətləndirə bilir. Bu modellər semantik uyğunluq, məna bütövlüyü və terminoloji düzgünlük kimi meyarlar üzrə kompleks analiz aparır [4, s.134].

Bununla yanaşı, bu sistemlərin əsas üstünlüklərindən biri qiymətləndirmə nəticələrinin izah oluna bilməsidir. Yəni model yalnız “yaxşı” və ya “zəif” qiymət vermir, həm də qərarın səbəblərini izah edir. Bu xüsusiyyət tərcümə təhsilində və peşəkar hazırlıq prosesində xüsusi əhəmiyyət kəsb edir [5, s.92]. Lakin aparılan tədqiqatlar göstərir ki, LLM əsaslı qiymətləndirmə sistemləri hələ də tam etibarlı hesab edilmir. Onların nəticələri istifadə olunan prompt strukturundan, model versiyasından və kontekstin təqdim olunma formasından asılı olaraq dəyişə bilər [6, s.145].



Bu işə qiymətləndirmənin sabitliyi və təkrarlanırlılığı ilə bağlı müəyyən problemlər yaradır.

Şifahi tərcümənin avtomatik qiymətləndirilməsində mühüm yanaşmalardan biri də multimodal analiz sistemləridir. Bu sistemlər yalnız mətn deyil, həm də səs signalları, prosodiya və hətta vizual məlumatları (məsələn, bədən dili) nəzərə alaraq daha kompleks qiymətləndirmə aparır [7, s.110]. Multimodal yanaşma xüsusilə konfrans tərcüməsi və real vaxt interpretasiya sahəsində daha dəqiq nəticələr əldə etməyə imkan verir.

Bununla belə, bu texnologiyaların tətbiqi müəyyən çətinliklərlə müşayiət olunur. İlk növbədə, müxtəlif dillər və dialektlər üzrə məlumat bazalarının məhdudluğu sistemlərin öyrədilməsini çətinləşdirir. Digər tərəfdən, mədəni və situativ kontekstlərin avtomatik sistemlər tərəfindən düzgün interpretasiya olunmaması qiymətləndirmə nəticələrinə təsir göstərir [8, s.67].

Yazılı və şifahi tərcümənin Sİ əsaslı qiymətləndirilməsi arasında mühüm fərqlər mövcuddur. Yazılı tərcümədə əsas diqqət mətnin strukturuna və semantik uyğunluğa yönəldiyi halda, şifahi tərcümədə zaman faktoru, nitqin axıcılığı və kommunikativ effektivlik əsas rol oynayır. Bu fərq avtomatik qiymətləndirmə modellərinin dizaynına da birbaşa təsir göstərir [9, s.119].

Müasir yanaşmalarda yazılı və şifahi tərcümə qiymətləndirməsinin inteqrasiyası istiqamətində də tədqiqatlar aparılır. Bu yanaşma “hybrid evaluation systems” adlandırılır və həm mətn əsaslı, həm də səs əsaslı analizlərin birləşdirilməsini nəzərdə tutur. Bu sistemlər daha obyektiv və kompleks nəticələr əldə etməyə imkan yaradır [10, s.88].

Bununla yanaşı, AI əsaslı qiymətləndirmə sistemlərinin etik problemləri də diqqət mərkəzindədir. Məlumatların məxfiliyi, alqoritmik qərəzlilik və insan ekspertlərinin rolunun azalması kimi məsələlər elmi ədəbiyyatda geniş müzakirə olunur [11, s.102]. Xüsusilə tərcümə təhsili sahəsində Sİ sistemlərinin həddindən artıq istifadəsi pedaqoji balansın pozulmasına səbəb ola bilər.

Beləliklə, yazılı və şifahi tərcümənin süni intellekt vasitəsilə qiymətləndirilməsi

sahəsi sürətlə inkişaf etsə də, hələ də bir sıra metodoloji və texniki problemlər mövcuddur. Bu problemlərin həlli gələcəkdə daha dəqiq, etibarlı və çoxölçülü qiymətləndirmə sistemlərinin yaradılmasına imkan verəcəkdir.

Tərcümə təhsilində Sİ əsaslı qiymətləndirmənin tətbiqi və gələcək perspektivlər. Süni intellekt əsaslı avtomatik qiymətləndirmə sistemlərinin tərcümə təhsilinə inteqrasiyası son illərdə mühüm elmi və pedaqoji istiqamət kimi formalaşmışdır. Müasir universitetlərdə tərcümə proqramlarının məqsədi yalnız dil bacarıqlarının inkişafı deyil, həm də texnoloji vasitələrdən səmərəli istifadə edə bilən peşəkar tərcüməçilərin hazırlanmasıdır [1, s.76].

Sİ əsaslı qiymətləndirmə sistemləri tələbələrin tərcümə bacarıqlarının daha obyektiv və sistemli şəkildə ölçülməsinə imkan yaradır. Bu sistemlər vasitəsilə tələbələrin etdiyi səhvlər avtomatik şəkildə təsnif edilir, semantik və qrammatik uyğunsuzluqlar müəyyən olunur və fərdi inkişaf trayektoriyası formalaşdırılır [2, s.98]. Bu yanaşma ənənəvi müəllim qiymətləndirməsinə əlavə dəstək kimi çıxış edir.

Tərcümə təhsilində süni intellekt texnologiyalarının istifadəsi həmçinin “formativ qiymətləndirmə” prosesini gücləndirir. Yəni tələbə yalnız yekun nəticəyə görə deyil, proses boyunca davamlı geribildirim alır. Bu işə öyrənmə prosesinin effektivliyini artırır və tələbənin öz səhvlərini daha tez aşkar etməsinə imkan yaradır [3, s.121].

Son illərdə aparılan tədqiqatlar göstərir ki, GPT əsaslı sistemlər tələbələrin tərcümə bacarıqlarının inkişafında mühüm rol oynaya bilər. Bu sistemlər yalnız səhvləri göstərmir, həm də alternativ tərcümə variantları təklif edərək tələbənin analitik düşünmə qabiliyyətini inkişaf etdirir [4, s.134].

Bununla belə, Sİ əsaslı qiymətləndirmə sistemlərinin təhsil prosesində geniş tətbiqi müəyyən etik və pedaqoji suallar doğurur. İlk növbədə, alqoritmik qərarların şəffaflığı məsələsi xüsusi əhəmiyyət kəsb edir. Tələbələrə verilən qiymətin hansı meyarlara əsaslandığı aydın şəkildə izah edilməlidir [5, s.87].



Digər mühüm məsələ akademik dürüstlük və insan ekspertinin roludur. Sİ sistemlərinin həddindən artıq istifadəsi müəllimin qiymətləndirmə funksiyasını zəiflədə bilər. Buna görə də müasir yanaşmalar Sİ-ni müəllimi əvəz edən deyil, onu dəstəkləyən vasitə kimi qəbul edir [6, s.102].

Tərcümə sahəsində Sİ texnologiyalarının tətbiqi iqtisadi baxımdan da əhəmiyyətlidir. Avtomatik qiymətləndirmə sistemləri böyük həcmli işlərin qısa zamanda emalına imkan verdiyi üçün universitetlərdə və tərcümə agentliklərində əməliyyat xərclərinin azalmasına səbəb olur [7, s.110]. Bu, xüsusilə böyük təhsil proqramlarında resursların optimallaşdırılması baxımından vacibdir.

Gələcək perspektivlər göstərir ki, süni intellekt əsaslı qiymətləndirmə sistemləri daha adaptiv və fərdiləşdirilmiş olacaqdır. Yeni sistemlər yalnız ümumi qiymət verməyəcək, həm də hər bir tələbənin zəif və güclü tərəflərini analiz edərək fərdi öyrənmə planları təklif edəcək [8, s.129].

Eyni zamanda multimodal qiymətləndirmə sistemlərinin inkişafı gözlənilir. Bu sistemlər yalnız mətn deyil, həm də səs, intonasiya, nitq sürəti və hətta mimika kimi elementləri nəzərə alaraq şifahi tərcümənin daha kompleks qiymətləndirilməsini təmin edəcək [9, s.141].

Azərbaycan kimi çoxdilli və multikultural mühitə malik ölkələrdə Sİ əsaslı tərcümə qiymətləndirmə sistemlərinin tətbiqi xüsusi əhəmiyyət daşıyır. Bu sistemlər tərcümə təhsilinin keyfiyyətini artırmaqla yanaşı, beynəlxalq elmi integrasiyanı da gücləndirə bilər [10, s.95].

Nəticə etibarilə, süni intellekt əsaslı avtomatik qiymətləndirmə sistemləri tərcümə və şifahi tərcümə sahəsində yeni paradigma formalaşdırır. Bu sistemlər insan ekspertizasını əvəz etmir, lakin onu tamamlayaraq daha obyektiv, sürətli və çoxölçülü qiymətləndirmə imkanları yaradır.

Nəticə

Təqdim olunan tədqiqatın elmi yeniliyi süni intellekt əsaslı avtomatik qiymətləndirmə sistemlərinin yalnız maşın tərcüməsinin deyil, həm də insan tərəfindən yerinə yetirilən yazılı

və şifahi tərcümənin keyfiyyətinin kompleks şəkildə qiymətləndirilməsi baxımından sistemləşdirilmiş təhlilində özünü göstərir. Araşdırmada BLEU, METEOR, TER, BERTScore və COMET kimi klassik və neyron əsaslı metrikaların yanaşı, böyük dil modellərinə (LLM) əsaslanan yeni "LLM-as-a-Judge" yanaşmasının imkanları və məhdudiyyətləri müqayisəli şəkildə qiymətləndirilmişdir. Bu yanaşma tərcümə keyfiyyətinin qiymətləndirilməsində insan ekspertizası ilə süni intellekt modelləri arasında qarşılıqlı integrasiyanın nəzəri əsaslarını formalaşdırır və sahədə yeni hibrid qiymətləndirmə modelinin inkişafına töhfə verir.

Tədqiqatın tətbiqi əhəmiyyəti onun tərcümə təhsili, peşəkar tərcümə fəaliyyəti və şifahi tərcümə hazırlığı proseslərində istifadə imkanları ilə bağlıdır. Süni intellekt əsaslı avtomatik qiymətləndirmə sistemləri tələbələrə tərcümə bacarıqlarının operativ və obyektiv qiymətləndirilməsinə, səhvlərin avtomatik identifikasiyasına və fərdi geribildirim mexanizminin qurulmasına imkan yaradır. Eyni zamanda bu sistemlər tərcümə agentliklərində keyfiyyətə nəzarət prosesini sürətləndirir və insan resurslarının iş yükünü azaldır. Şifahi tərcümə sahəsində isə ASR və NLP texnologiyalarının integrasiyası real vaxt rejimində performansın qiymətləndirilməsini mümkün edir.

Süni intellekt əsaslı qiymətləndirmə sistemlərinin tətbiqi iqtisadi baxımdan əhəmiyyətli qənaət imkanları yaradır. Ənənəvi insan ekspertizasına əsaslanan qiymətləndirmə sistemləri yüksək əmək və zaman xərcləri tələb etdiyi halda, avtomatik sistemlər böyük həcmli tərcümə materiallarını qısa müddətdə emal edə bilər. Bu isə həm təhsil müəssisələrində, həm də tərcümə sənayesində əməliyyat xərclərinin azalmasına, məhsuldarlığın artmasına və resursların daha səmərəli idarə olunmasına səbəb olur. Bundan əlavə, Sİ əsaslı sistemlərin istifadəsi beynəlxalq tərcümə bazarında rəqabət qabiliyyətinin yüksəlməsinə də şərait yaradır.

Beləliklə, süni intellekt əsaslı avtomatik qiymətləndirmə sistemləri tərcümə və şifahi tərcümə sahəsində yeni elmi-paradigmatik yanaşma formalaşdırır. Bu sistemlər insan



ekspertizasını əvəz etməsə də, onu tamamlayaraq daha obyektiv, sürətli və çoxölçülü qiymətləndirmə mexanizmlərinin tətbiqinə imkan verir. Gələcək tədqiqatlarda bu sistemlərin etibarlılığının artırılması və multimodal qiymətləndirmə modellərinin inkişaf etdirilməsi əsas istiqamətlərdən biri kimi qalacaqdır.

ƏDƏBİYYAT SİYAHISI

1. Baker M. In other words: a coursebook on translation. 3rd ed. London: Routledge, 2018.
2. House J. Translation quality assessment: past and present. London: Routledge, 2015.
3. Koehn P. Neural machine translation. Cambridge: Cambridge University Press, 2020.
4. Barrault, L., Biesialska M., Bojar O., Costa-jussà M.R., Federmann C., Graham Y., Grundkiewicz R., Haddow B., Huck M., Joanis E., Kocmi T., Koehn P., Lo C.K., Ljubešić N., Monz C., Morishita M., Nagata M., Nakazawa T., Pal S., Post M., Zampieri M. Findings of the 2020 Conference on Machine Translation (WMT20). In Proceedings of the Fifth Conference on Machine Translation. Association for Computational Linguistics. 2020; 1–55.
5. Papineni K., Roukos S., Ward T., Zhu W.J. BLEU: a method for automatic evaluation of machine translation. ACL Proceedings, 2002.
6. Banerjee S., Lavie A. METEOR: an automatic metric for MT evaluation with improved correlation with human judgments. ACL Workshop, 2005.
7. Snover M., Dorr B., Schwartz R., Micciulla L., Makhoul J. A study of translation edit rate with targeted human annotation. AMTA, 2006.
8. Zhang T., Kishore V., Wu F., Weinberger K., Artzi Y. BERTScore: evaluating text generation with BERT." ICLR, 2020.
9. Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020). COMET: A Neural Framework for MT Evaluation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 2685–2702). Association for Computational Linguistics.
10. Freitag M., Mathur N., Lo C.K., Avramidis E., Rei R., Thompson B., Kocmi T., Blain F., Deutsch D., Stewart C., Zerva C., Castilho S., Lavie A., Foster G. Results of WMT23 metrics shared task: metrics might be guilty but references are not innocent. In Proceedings of the Eighth Conference on Machine Translation. Association for Computational Linguistics. 2023; 578–628.
11. OpenAI. GPT-4 Technical Report. 2023. <https://arxiv.org/abs/2303.08774>
12. Zhang Y., Van Genabith J. LLM-based evaluation of machine translation: opportunities and risks. Machine Translation Journal, 2024.

Leyla Vagif ALIZADA

PhD in German Philology, Senior Lecturer of the Department of Languages
Azerbaijan Tourism and Management University, Baku, Azerbaijan

ARTIFICIAL INTELLIGENCE-BASED AUTOMATIC EVALUATION SYSTEMS: APPROACHES TO ASSESSING THE QUALITY OF WRITTEN AND INTERPRETING TRANSLATION, RELIABILITY, AND APPLICATION PERSPECTIVES

Summary

Purpose of this study is to analyze the role of artificial intelligence-based automatic evaluation systems in assessing the quality of written and interpreting translation, as well as to examine their methodological foundations and reliability. Design/method/approach is based on comparative and

analytical methods combined with a systematic review of existing scientific literature. The study covers both traditional and neural evaluation metrics such as BLEU, METEOR, TER, BERTScore, and COMET, as well as large language model (LLM)-based evaluation approaches. Findings (Results) indicate that AI-based systems have strong potential in translation quality assessment; however, they still cannot fully replace human expertise, particularly in interpreting, where multimodal factors play a critical role.

Keywords: artificial intelligence, translation quality, automatic evaluation, BLEU, interpreting.

Лейла Вагиф АЛИЗАДЕ

PhD в области германской филологии, старший преподаватель кафедры языков
Азербайджанский Университет Туризма и Менеджмента, Баку, Азербайджан

**СИСТЕМЫ АВТОМАТИЧЕСКОЙ ОЦЕНКИ НА ОСНОВЕ ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА: ПОДХОДЫ К ИЗМЕРЕНИЮ КАЧЕСТВА ПИСЬМЕННОГО И
УСТНОГО ПЕРЕВОДА, НАДЕЖНОСТЬ И ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ**

Резюме

Цель исследования заключается в анализе роли систем автоматической оценки на основе искусственного интеллекта в измерении качества письменного и устного перевода, а также в изучении их методологических основ и уровня надежности. Метод исследования включает сравнительно-аналитический подход и систематический обзор научной литературы. В исследовании рассматриваются классические и нейронные метрики оценки перевода, такие как BLEU, METEOR, TER, BERTScore и COMET, а также подходы на основе больших языковых моделей (LLM). Результаты исследования показывают, что системы искусственного интеллекта обладают высоким потенциалом в автоматической оценке качества перевода, однако они пока не могут полностью заменить человеческую экспертизу, особенно в контексте устного перевода, где важны мультимодальные факторы.

Ключевые слова: искусственный интеллект, качество перевода, автоматическая оценка, BLEU, устный перевод.

Daxil olub: 01.07.2026