ISSN E: 2709-8273
ISSN P: 2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

اُردو ڈیجیٹل ترجمہ نگاری: موجودہ صورت حال اور امکانات

(URDU DIGITAL TRANSLATION: CURRENT STATUS AND FUTURE PROSPECTS)

ڈاکٹر قمر عباس
پوسٹ ڈاکٹریٹ فیلوشپ ریسرچ سکالر
شعبہ اردو، بین الاقوامی اسلامی یونیورسٹی اسلام آباد
ڈاکٹر سائرہ بتول
ایسوسی ایٹ پروفیسر / چیئرپرسن
شعبہ اردو، بین الاقوامی اسلامی یونیورسٹی اسلام آباد
Email: sairabatool@iiu.edu.pk

ABSTRACT:
This study examines the current state and future prospects of Urdu digital translation within the broader historical and technological development of machine translation. It begins by outlining the evolution of digital translation systems and reviewing their application across major world languages, followed by a critical analysis of existing Urdu translation platforms such as Google Translate, Bing Translate, ChatGPT, and other AI-based tools. The research identifies key linguistic and technical challenges that affect Urdu translation quality, including script directionality, morphological and syntactic complexity, polysemy, idiomatic expressions, cultural references, tokenization and parsing difficulties, and Unicode compatibility issues. By situating Urdu within the framework of Artificial Intelligence (AI) and Natural Language Processing (NLP), the study highlights the need for language-specific AI models, large-scale corpora, annotated treebanks, and domain-sensitive lexical resources to improve translation accuracy and contextual coherence. It further explores the applicability of advanced language models such as BERT, LLaMA, and generative AI systems in enhancing Urdu machine translation. In response to the identified limitations, the research proposes a corpus-driven, AI-integrated Urdu translation web application framework designed to provide context-aware, stylistically appropriate, and semantically accurate translations. The study contributes both analytically and practically by offering a comprehensive evaluation of Urdu digital translation and presenting a scalable model aimed at strengthening Urdu's position in the global digital and AI-driven linguistic landscape.

KEYWORDS: Urdu Digital Translation, Machine Translation, Neural Machine Translation, Computational Linguistics, Corpus Development, Multilingual Models, Translation Technology, Linguistic Challenges, AI-based Translation

---

ترجمے کا فن ہمیشہ انسانی ضرورت رہا ہے جس کے ذریعے قوموں کے باہمی روابط ممکن ہوئے ہیں۔ اس فن نے انسانی ارتقا کے ساتھ ترقی کی ہے۔ آج کا دور ڈیجیٹل دور ہے۔ چنانچہ اب ترجمہ بھی مشینی صورت اختیار کر چکا ہے اور مصنوعی ذہانت کی شمولیت کی وجہ سے اس میں نئے نئے امکانات پیدا ہو چکے ہیں۔ ڈیجیٹل ترجمے کی ابتدا دوسری جنگ عظیم کے ساتھ ہوئی جب عالمی زبانوں کی تفہیم کی ضرورت محسوس کی گئی۔ مشینی ترجمے پر تحقیق کا باقاعدہ آغاز ہوا۔ اس کا پہلا تجربہ 1954 میں امریکہ کی چارج ٹاؤن یونیورسٹی اور آئی بی ایم کمپنی کی مشترکہ کوشش سے کیا گیا۔ (1) اس تجربے کو جارج ٹاؤن آئی بی ایم ایکسپیریمنٹ کا نام دیا گیا۔ روسی زبان کے 60 جملوں کا انگریزی میں ترجمہ ہوا۔ اس محدود نوعیت کے تجربے نے مشینی ترجمے کی بنیادیں رکھ دیں۔

1960 اور 1970 کی دہائیوں میں مشینی ترجمے پر کافی توجہ دی گئی لیکن خاطر خواہ نتائج بر آمد نہ ہوئے۔ جس کی وجہ سے اس کو غیر عملی قرار دے کر امریکی حکومت نے فنڈنگ بند کر دی۔ اس ترجمے کی ناکامی کی بنیادی وجہ قواعد پر مبنی ترجمے کا نظام تھا۔ جس میں زبان کے قواعد، لغات اور نحوی ساخت کو پیش نظر رکھا جاتا۔ اس طرح یہ طریقہ کار غیر لچکدار ثابت ہوا۔

1990 کی دہائی میں آئی بی ایم نے ترجمے کا شماریاتی طریقہ کار متعارف کروایا جس میں ذولسانی جوڑوں کا تجزیہ کر کے ترجمے کے نئے امکانات تلاش کیے گئے۔ اسی شماریاتی ماڈل کو استعمال کرتے ہوئے 2006 میں گوگل نے گوگل ٹرانسلیٹ کی ابتدا کی۔ یہ ڈیجیٹل ترجمے میں انقلاب کی حیثیت رکھتا ہے۔ اگرچہ اس کا معیار ابھی تک

ISSN E: 2709-8273
ISSN P:2709-8265
JALT
JOURNAL OF APPLIED LINGUISTICS AND TESOL

غیر تسلی بخش تھا لیکن اس سے ترجمے کی رفتار میں غیر معمولی ترقی آگئی۔ بات یہیں تک رکی نہیں۔ مصنوعی ذہانت کی آمد نے اس میں نئے امکانات پیدا کر دیے۔ چنانچہ گوگل نے 2016 میں ایس ایم ٹی ماڈل کی جگہ نیورل مشین ٹرانسلیشن کا آغاز کر دیا۔(2) یہ خود کار سیکھنے کی صلاحیت رکھتا تھا اور سیاق وسباق کا لحاظ رکھتے ہوئے پورے جملے کا ترجمہ کر سکتا تھا۔ اس نے ترجمے کی دنیا میں انقلاب پیدا کر دیا اور آج بیشتر ترجمے کے ذرائع اسی ٹیکنالوجی پر انحصار کر رہے ہیں۔ اس سے قبل مشینی ترجمہ میں شماریاتی مشینی ترجمہ (Statistical Machine Translation) کے طریقے استعمال کیے جاتے تھے جن میں بڑے دولسانی کارپس کی مدد سے ترجمہ کے امکانات کا اندازہ لگایا جاتا تھا۔(3)

اُردو زبان کے دنیا کی کئی زبانوں کے ساتھ روابط ہیں اور مزید روابط بڑھانے کی ہمیشہ ضرورت رہی ہے۔ لیکن جب بات ڈیجیٹل ترجمے تک پہنچی تو اسے کافی مسائل کا سامنا کرنا پڑا۔ اُردو ایک انفرادی نوعیت کی زبان ہے جس کی نحوی ترتیب کافی لچکدار ہے۔ اس کے الفاظ کے معنی بھی سیاق وسباق کے مطابق تبدیل ہوتے رہتے ہیں۔ جونہی دیگر زبانوں نے ڈیجیٹل ترجمے کی دنیا میں قدم رکھا تو اُردو بھی اس دوڑ میں شامل ہو گئی۔ اگرچہ یہ دیگر ترقی یافتہ زبانوں کے مقابلے میں اس میدان میں بہت پیچھے ہے۔ لیکن پھر بھی یہ ان کے نقش قدم پر چلنے کی کوشش کر رہی ہے۔ اُردو کو یونی کوڈ کی طاقت ملنے سے اس کے لیے کافی آسانیاں پیدا ہوئیں۔ اس کے ڈیجیٹل ترجمے میں وقت کے ساتھ ساتھ بہتری پیدا ہو رہی ہے۔ اُردو کی موجودہ صورتحال جاننے کے لیے ضروری ہے کہ یہ دیکھا جائے کہ اس وقت دنیا کی ترقی یافتہ زبانیں ڈیجیٹل ترجمے کے حوالے سے کس مقام پر ہیں اور ان کے مقابلے میں اُردو کی صورتحال کیا ہے۔

جن زبانوں نے ابتدا ہی میں ڈیجیٹل دنیا کے ساتھ ہم آہنگی پیدا کی اور ڈیجیٹل ترجمے کو اپنایا، وہ اس میدان میں سب سے آگے نظر آتی ہیں۔ اس حوالے سے انگریزی، چینی، ہسپانوی، فرانسیسی، روسی، جرمن اور عربی جیسی بڑی زبانیں نمایاں ہے۔ ان زبانوں میں ڈیجیٹل ترجمے کی سہولیات حیران کن ہیں اور ترجمے کی روانی اور معیار کافی بہتر ہے۔ اس کے ساتھ ساتھ یہ زبانیں روز بروز اس حوالے سے بہتری کی صورتحال پیدا کر رہی ہیں۔

دنیا کی اہم زبانوں میں ڈیجیٹل ترجمہ:

ڈیجیٹل ترجمے کا سب سے وسیع پیمانے پر استعمال انگریزی زبان ہی میں نظر آتا ہے۔ یوں بھی کہہ سکتے ہیں کہ انگریزی ڈیجیٹل ترجمے کی بنیاد کی حیثیت رکھتی ہے۔ انگریزی کی گلوبل حیثیت اسے مقبول بناتی ہے اور یہی وجہ ہے کہ اس پر ڈیجیٹل دنیا میں بھی سب سے زیادہ توجہ دی گئی ہے۔ یہ زبان ڈیجیٹل دنیا میں سب سے زیادہ مستعمل زبان بن چکی ہے اور ڈیجیٹل ترجمے میں اپنا کلیدی کردار ادا کر رہی ہے۔ اس کی کئی وجوہات ہیں:

انگریزی زبان میں انٹرنیٹ پر مواد وافر مقدار میں میسر ہے جس کے ذریعے مشینی ترجمے کے ماڈلز کو تربیت دینے میں سہولت پیدا ہوتی ہے۔ ایک بڑی تعداد میں انگریزی زبان کے ڈیٹا کارپس موجود ہیں۔ مشینی ماڈلز انگریزی زبان کو ہی بنیادی ذریعہ کے طور پر استعمال کرتے ہیں جس کی وجہ سے انگریزی زبان مرکزی حیثیت حاصل کر لیتی ہے۔

اسی طرح ترجمے کے مختلف پلیٹ فارمز بھی انگریزی زبان کو مرکزی حیثیت دیتے ہیں۔ گوگل ٹرانسلیٹ انگریزی زبان کو ہی اہمیت دیتا ہے اور تمام زبانوں کے تراجم کے لیے اسے رابطے کی زبان کے طور پر استعمال کرتا ہے۔ مائیکروسافٹ ٹرانسلیٹر کے زیادہ تر ماڈلز انگریزی زبان پر مشتمل ہیں جن کے ذریعے دیگر زبانوں کے ساتھ تعلق زیادہ آسان ہو جاتا ہے۔ ڈیپ ایل ٹرانسلیٹ ایک جدید طاقتور ٹول ہے اور اس کی کامیابی کی بڑی وجہ بھی یہی ہے کہ اس نے انگریزی اور یورپی زبانوں کے تراجم کی طرف خصوصی توجہ دی ہے۔ انگریزی سے دوسری زبانوں میں ترجمے میں درستی کا تناسب بھی کافی بہتر ہے۔ این ایم ٹی ماڈلز جن کو گوگل نے اپنایا، ان کی ابتدائی تربیت انگریزی اور دیگر یورپی زبانوں کے باہمی تراجم پر رکھی گئی اور اس میں انگریزی زبان کو مرکزی حیثیت حاصل تھی۔ اس کی وجوہات یہ تھی کہ انگریزی جملوں کی ساخت دنیا کی دیگر بڑی زبانوں سے کافی مطابقت رکھتی ہے۔ انگریزی میں قواعد کی پیچیدگیاں کم ہوتی ہیں جس کی وجہ سے ماڈلز کی ٹریننگ آسان ہوتی ہے۔ انگریزی کی ساختیاتی سادگی بھی اس حوالے سے معاون ہوتی ہے۔

انگریزی زبان پر ڈیجیٹل ترجمے کے میدان میں کافی ترقی ہوئی ہے لیکن کچھ پیچیدگیاں ابھی بھی موجود ہیں۔ کئی الفاظ کثیر المعانی ہوتے ہیں جن کا ترجمہ سیاق وسباق کو سمجھے بغیر مناسب نہیں ہوتا۔ محاورے اور ثقافتی اظہار ترجمے میں صحیح معنویت بیان نہیں کر سکتا۔ انگریزی کارپس مغربی تہذیبی اثرات رکھتے ہیں اس لیے دیگر زبانوں کے ثقافتی اظہار میں کمزوری کا باعث بنتے ہیں۔ خاص علوم سے متعلقہ اصطلاحات کا ترجمہ بھی نیورل ماڈلز کے لیے کافی مشکل ہو جاتا ہے۔ ان مشکلات کے باوجود انگریزی زبان مشینی ترجمہ میں قائدانہ حیثیت رکھتی ہے اور نہ صرف یہ خود آگے بڑھ رہی ہے بلکہ دیگر زبانوں کو بھی سہولت فراہم کرنے کا باعث بن رہی ہے۔

چینی دنیا میں سب سے زیادہ بولی جانے والے زبانوں میں مرکزی حیثیت رکھتی ہے۔ چین کی آبادی، معاشی طاقت اور سائنس و ٹیکنالوجی میں ترقی نے اسے دنیا میں اہم مقام عطا کیا ہے۔ چینی قوم نے جہاں دیگر میدانوں میں ترقی کی ہے وہیں اپنی زبان پر بھی خصوصی توجہ دی ہے۔ خصوصاً ڈیجیٹل ترجمہ نگاری ان کی ترجیحات میں شامل ہے۔ چینی زبان جدید مشینی ماڈلز کو اپنا چکی ہے۔ اس کے ساتھ ہی اپنے اے آئی ماڈلز کو بھی اس مقصد کے لیے استعمال کیا گیا ہے۔ چینی حکومت نے ڈیجیٹل ترجمے کو قومی سطح پر اہم شمار کرتے ہوئے کئی اقدامات کیے ہیں۔ نیورل مشین ترجمہ اور لسانی ماڈلز پر تحقیقی ادارے قائم کیے گئے ہیں۔ یہ حکومتی ادارے خصوصی اقدامات کرتے ہیں۔ ان کی سرپرستی میں مشینی ترجمے کو نصابی ضروریات اور حکومتی ویب سائٹس میں استعمال کیا جا رہا ہے۔ حکومتی سطح پر اربوں یوان کی چینی کمپنیوں کو مالی امداد فراہم کی جاتی ہے تاکہ وہ ترجمہ کے نظام کو بہتر بنا کر دنیا کے ساتھ رابطے آسان کر سکیں۔ چینی ٹیکنالوجی کے بہت سے ادارے اس حوالے سے کام کر رہے ہیں۔ بیدو ٹرانسلیٹ چین کے سب سے بڑے سرچ انجن کی ترجمہ سروس ہے۔ یہ اُردو، چینی، انگریزی، ہسپانوی، جاپانی، کوریائی زبانوں سمیت دنیا کی دیگر زبانوں کے باہمی ترجمے کی سہولت فراہم کرتا ہے۔ یہ ماڈل سیاق و سباق کے حوالے سے درست ترجمہ کرنے کی کافی حد تک صلاحیت رکھتا ہے۔

ٹینسنٹ ٹرانسلیٹ چینی پلیٹ فارم چینی زبان کے ساتھ ساتھ دیگر ایشیائی زبانوں کے مشینی ترجمے کے منصوبے چلا رہا ہے۔ آئی فلائی ٹیک کمپنی خصوصی طور پر چینی زبان کے ترجمے کی مہارت رکھتی ہے اور اس کے تیار کردہ ترجمے کے آلات بین الاقوامی کانفرنسوں اور تعلیمی اداروں میں کام آتے ہیں۔

چینی زبان کی پیچیدہ لسانی ساخت کو اے آئی ماڈلز کے ساتھ ہم آہنگ کرنے کے لیے بہت سے اقدامات کیے گئے ہیں۔ اس حوالے سے ڈیٹا کارپس اور مصنوعی ذہانت کے خصوصی ماڈلز استعمال کیے گئے ہیں۔ اس سلسلے میں ٹرانسفارمر ماڈلز جیسے جی پی ٹی، برٹ کے ذریعے چینی زبان کے الفاظ کے سیاق و سباق کو سمجھنے میں مدد لی گئی ہے۔ ملٹی لینگوئل پری ٹرینڈ ماڈلز کئی زبانوں پر ٹرینڈ ہوتے ہیں ان میں چینی زبان کی شمولیت سے چینی میں ترجمے کی صلاحیت بڑھ گئی ہے۔ چینی حکومت کی ویب سائٹس، صنعتی رپورٹوں اور تعلیمی اداروں میں ترجمہ اب خودکار انداز میں ہوتا ہے، چینی ادبیات کے تراجم میں بھی اب ان ذرائع سے مدد لی جا رہی ہے۔ چینی کمپنیاں عالمی سطح پر اپنے کاروبار کو فرغ دے رہی ہیں ۔ جدید ڈیجیٹل تراجم کے بغیر یہ چیز ممکن نہیں۔ یہی وجہ ہے کہ نہ صرف حکومتی ادارے بلکہ نجی کمپنیاں بھی اس میدان میں خصوصی دلچسپی لے رہی ہیں جس کی وجہ سے چینی زبان اس وقت ڈیجیٹل ترجمہ نگاری میں اہم مقام حاصل کر چکی ہے۔

ہسپانوی زبان دنیا کی بڑی زبانوں میں شمار ہوتی ہے۔ یہ 20 سے زائد ممالک کی سرکاری زبان کا درجہ رکھتی ہے۔ انگریزی کے بعد سب سے زیادہ مشینی ترجمہ اسی زبان کا کیا جاتا ہے۔ اس کی ایک وجہ یہ ہے کہ اس کی ساخت ڈیجیٹل ترجمہ کے لیے نہایت موزوں ہے۔ اس کے علاوہ اس زبان کی عالمی اہمیت ڈیجیٹل ترجمہ نگاروں کو اپنی طرف متوجہ کرتی ہے۔ معروف ترجمہ کے پلیٹ فارمز بھی اس کو ترجیحی زبانوں میں شامل کرتے ہیں۔ نیورل مشین ترجمے کے ماڈلز اس زبان پر بہت موثر ثابت ہوتے ہیں۔ ہسپانوی زبان کی گرامر اور ساخت کافی لحاظ سے انگریزی زبان کے قریب تر ہے۔ انگریزی ہسپانوی ترجمہ کا معیار کافی بہتر ہوتا ہے۔ اس زبان کے مشین لرننگ کے نتائج کافی بہتر سامنے آتے ہیں۔ (4)

مختلف ڈیجیٹل پلیٹ فارمز ہسپانوی ترجمے میں حیران کن نتائج کا مظاہرہ کرتے ہیں۔ گوگل ٹرانسلیٹ پر ہسپانوی زبان کے تراجم سب سے زیادہ ہوتے ہیں۔ خاص طور پر انگریزی ہسپانوی ترجمہ 90 فیصد درست دیکھا گیا ہے۔ ڈیپ ایل ٹرانسلیٹ اس زبان کے جملوں اور الفاظ کو سیاق و سباق کے مطابق سمجھ کر کافی معیاری تراجم فراہم کرتا ہے۔ مائیکروسافٹ ٹرانسلیٹ بھی ہسپانوی زبان کو مرکزی اہمیت دیتا ہے۔ مصنوعی ذہانت اس زبان کو بہتر انداز میں سمجھنے لگی ہے۔ البتہ بعض شعبوں میں اس زبان میں تراجم کا معیار ابھی محدود ہے۔ خاص طور پر ثقافتی اور ادبی متون کے تراجم میں مسائل موجود ہیں۔ یہ زبان جس تیزی سے ترجمے کے میدان میں آگے بڑھ رہی ہے مستقبل قریب میں اس میں مزید بہتری کی توقع ہے۔

فرانسیسی زبان کا شمار دنیا کی ان زبانوں میں ہوتا ہے جنہوں نے ڈیجیٹل ترجمہ نگاری کو بخوبی اپنایا ہے۔ اقوام متحدہ، یورپی یونین، افریقی یونین اور دیگر اہم عالمی اداروں میں اس کو سرکاری حیثیت حاصل ہے۔ اس حوالے سے بھی اہم پلیٹ فارمز اس کو ترجیحی بنیادوں پر ترجمہ نگاری کی فہرست میں رکھتے ہیں۔ نیورل مشین ترجمہ نگاری کو بھی اپنایا ہے اور اس حوالے سے انقلابی اقدامات سامنے آئے ہیں۔ اس کا رومن رسم الخط اور ڈیجیٹل متون کی دستیابی مصنوعی ذہانت کے لیے مفید ہے اور اے آئی ماڈلز کو بخوبی ٹریننگ دی جا سکتی ہے۔ فرانسیسی انگریزی ترجمے کا معیار بہت زیادہ بہتر ہے۔ گوگل ٹرانسلیٹ فرانسیسی زبان کا ترجمہ نہایت عمدگی سے کرتا ہے جسے مختلف شعبوں میں استعمال کیا جا سکتا ہے۔ اسی طرح ڈیپ ایل ٹرانسلیٹ نے فرانسیسی ترجمہ نگاری میں بہت زیادہ کامیابی دکھائی ہے۔ اس پلیٹ فارم میں فرانسیسی تراجم کافی رواں اور فطری انداز رکھتے ہیں۔ انگریزی اور فرانسیسی کے الفاظ کا ایک بہت بڑا ذخیرہ مشترک ہے جس کی وجہ سے مشین لرننگ کے ماڈلز کو تربیت دینے میں کافی آسانی

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

ہوتی ہے۔ فرانسیسی ڈیجیٹل ترجمہ تعلیم، سیاحت، صحت، ساجیات، سرکاری اور عدالتی امور اور دیگر شعبوں میں استعمال ہو رہا ہے۔ اس میدان میں اس زبان کو کچھ چیلنجز کا بھی سامنا ہے۔ خاص طور پر ثقافتی تراجم، افریقی اور یورپی فرانسیسی کے درمیان لسانی فرق اور علاقائی اصطلاحات وغیرہ کے تغیرات مسائل کا باعث بنتے ہیں۔(5)

جرمن زبان کا شمار یورپی یونین کی سب سے زیادہ بولی جانے والی زبانوں میں ہوتا ہے۔ سائنسی، فنی اور ادبی میدان میں اس زبان کو عالمی حیثیت حاصل ہے۔ انگریزی اور چینی کے بعد جرمن زبان کو سمجھنے میں ڈیجیٹل ماڈلز نے کافی مہارت حاصل کی ہے۔ اس زبان کی پیچیدہ صرفی اور نحوی ساخت اگرچہ مشینی ترجمے کے لیے چیلنجز رکھتی ہے، تاہم مصنوعی ذہانت نے اس پر کافی حد تک قابو پالیا ہے۔ گوگل ٹرانسلیٹ نے جرمن زبان اور دیگر یورپی زبانوں میں ترجمے کی درستی میں نمایاں بہتری دکھائی ہے۔ ڈیپ ایل ٹرانسلیٹ کو جرمن زبان میں ترجمے کے حوالے سے سب سے زیادہ معتبر سمجھا جاتا ہے۔ ٹریڈاس اور سسٹران کے سافٹ ویئرز جرمن زبان میں بزنس، قانون اور ٹیکنیکل ڈاکومنٹیشن کے ترجمے کے حوالے سے زیادہ مقبول ہیں ۔

جرمنی تحقیق و ترقی پر سرمایہ کاری کرنے والا ایک بہت بڑا ملک ہے اور ان شعبوں میں ڈیجیٹل ترجمے کی معاونت اہم سمجھی جاتی ہے۔ تجارتی فیصلوں اور سرگرمیوں کی تیز ترسیل کے لیے بھی ڈیجیٹل ترجمے کی ضرورت اور اہمیت جرمن زبان کے لیے کلیدی کردار رکھتی ہے۔ مستقبل میں اس زبان کے لیے ڈیجیٹل ترجمہ مزید اہمیت اختیار کرے گا۔ جرمن حکومت اور تعلیمی ادارے اس مقصد کے لیے خصوصی کوششیں کر رہے ہیں اور اس سلسلے میں نیورل ماڈلز کی تربیت تیزی کے ساتھ جاری ہے۔

روسی زبان اقوام متحدہ کی سرکاری زبانوں میں شامل ہے اور یہ سب سے بڑی سلاوک زبان ہے جو بہت سے ملکوں میں بولی جاتی ہے۔ یہ عالمی سطح پر سائنس، سیاست اور ادبی اظہار کے لیے نمایاں حیثیت رکھتی ہے۔ ڈیجیٹل دنیا میں اس کے دیگر زبانوں سے روابط کے لیے مشینی ترجمے کی اہمیت مسلم ہے۔(6)

روسی زبان کے پیچیدہ گرامر کے نظام کو سمجھنے کے لیے مصنوعی ذہانت کے ماڈلز کو خصوصی تربیت کی ضرورت ہے۔ روس کی سب سے بڑی سرچ انجن کمپنی یانڈکس نے یانڈکس ٹرانسلیٹ کے ذریعے مشینی ترجمے کے میدان میں نمایاں ترقی کی ہے۔ یہ پلیٹ فارم اُردو اور دیگر زبانوں میں ترجمہ کی سہولت فراہم کرتا ہے۔ گوگل ٹرانسلیٹ بھی روسی ترجمہ کی سہولت فراہم کرتا ہے لیکن اس زبان میں یہ پلیٹ فارم سیاق وسباق سے ہٹ جاتا ہے۔ جی پی ٹی اور دیگر نیورل ماڈلز کی وجہ سے اس زبان کے ڈیجیٹل ترجمہ میں کافی بہتری آئی ہے۔ روسی حکومت نے تعلیمی اداروں، نجی کمپنیوں اور تحقیقی مراکز کو فنڈز جاری کیے ہیں تاکہ مشینی ترجمے کی بہتری کے لیے کام کیا جا سکے۔

روسی اور اُردو ترجمے میں معیار کے کافی مسائل ہیں۔ یہ ترجمہ براہ راست نہیں ہوتا بلکہ انگریزی زبان کو بطور لنک زبان استعمال کیا جاتا ہے۔ یعنی اُردو کو پہلے انگریزی اور پھر روسی زبان میں ترجمہ کیا جاتا ہے جس سے معیار میں کافی کمی آجاتی ہے۔ البتہ اب مصنوعی ذہانت کے مختلف ماڈلز کی وجہ سے براہ راست ترجمے کی راہ ہموار ہو رہی ہے جس سے بہتری کی توقع ہے۔

عربی زبان دنیا کی قدیم، کلاسیکی اور اثر انگیز زبانوں میں شمار ہوتی ہے۔ یہ زبان نہ صرف مذہبی لحاظ سے مسلمانوں کے لیے مقدس ہے بلکہ لسانی، ادبی، سائنسی اور قانونی متون میں بھی اس کا ایک بھرپور ذخیرہ موجود ہے۔ ڈیجیٹل ترجمے کے میدان میں عربی زبان کو ایک منفرد حیثیت حاصل ہے۔ یہ دائیں سے بائیں لکھے جانے والی عبارت اور پیچیدہ صرف ونحو کی وجہ سے ترجمے کے نظاموں کے لیے چیلنج بنی ہوئی ہے۔ جدید نیورل نیٹ ورکس اور مصنوعی ذہانت نے اس زبان کے ڈیجیٹل ترجمے کو کافی بہتر بنا دیا ہے۔ اس زبان کی بناوٹ دیگر زبانوں سے کافی مختلف ہے۔ اس میں جملوں کی ساخت، مذکر اور مونث کی پہچان، اعراب کی موجودگی مشینی ترجمے کے لیے چیلنج بن جاتی ہے۔ پھر یہ کہ ایک لفظ کے کئی کئی معنی جو سیاق وسباق کے مطابق تبدیل ہوتے ہیں ڈیجیٹل ترجمے کے معیار میں رکاوٹ بنتے ہیں۔ مصنوعی ذہانت کی شمولیت نے اس حوالے سے انقلابی سہولت پیدا کی۔ مائیکروسافٹ ٹرانسلیٹ اور ایمازون ٹرانسلیٹ نے عربی زبان کے لیے الگ ماڈلز تخلیق کیے جن میں sequence to sequence لرننگ کے عمل کو اپنایا گیا۔ اسی طرح گوگل ٹرانسلیٹ نے بھی عربی زبان کے لیے اقدامات کیے جس سے سیاق وسباق کے حوالے سے بہتر ترجمے کی سہولت ممکن ہوئی۔ ڈیپ ایل ٹرانسلیٹ اور پھر یانڈکس نے عربی زبان کو اپنے نظام میں شامل کیا جس سے دیگر زبانوں اور روسی زبان میں اس کے تراجم ممکن ہوئے۔(7)

عربی دنیا نے خود اپنی ضرورت کو محسوس کرتے ہوئے ڈیجیٹل ترجمے کی طرف توجہ دی۔ سعودی عرب، مصر، قطر، متحدہ عرب امارات وغیرہ نے اپنے تعلیمی اداروں اور سرکاری محکموں میں مشینی ترجمے کو فروغ دیا۔ اس مقصد کے لیے کارپس تیار کیے گئے اور سرمایہ کاری کی گئی۔ عربی کیونکہ قرآن کی زبان ہے اس لیے اس کے ڈیجیٹل ترجمے میں بے حد احتیاط کی ضرورت ہوتی ہے۔ اس لیے اس پر مکمل اعتبار ممکن نہیں کیونکہ اس طرح کا ترجمہ قابل بھروسہ نہیں ہوتا۔ البتہ اب کچھ مذہبی اداروں نے مصنوعی ذہانت پر مبنی قرآن فہمی کے تراجم کے نظام تیار کیے ہیں جن میں مخصوص اصولوں کو مد نظر رکھا گیا ہے۔ عربی زبان کے ڈیجیٹل ترجمے میں مسائل اس

ISSN E: 2709-8273
ISSN P:2709-8265
JALT
JOURNAL OF APPLIED LINGUISTICS AND TESOL

لیے بھی پیدا ہوتے ہیں کہ مختلف ممالک اور علاقوں میں عربی کے مختلف لہجے مستعمل ہیں جن کا فرق ڈیجیٹل نظاموں کو الجھن میں ڈال دیتا ہے۔ عربی میں محاورات اور ضرب الامثال کا ترجمہ بھی ترجمہ کے نظاموں کی فہم سے باہر ہو جاتا ہے۔ ثقافتی پس منظر تک رسائی بھی بعض اوقات ممکن نہیں ہوتی۔ جس تیزی سے اس زبان پر کام ہو رہا ہے امید کی جاتی ہے کہ مستقبل قریب میں عربی میں معیاری ڈیجیٹل ترجمے کی صورت پیدا ہو جائے گی۔

اُردو ڈیجیٹل ترجمے کے موجودہ ذرائع:

اُردو زبان میں ترجمے کی ایک مضبوط روایت موجود ہے۔ اُردو میں کافی معیاری ترجمے کیے جا رہے ہیں۔ ترجمہ نگاری جب ڈیجیٹل دور میں داخل ہوئی ہے تو اُردو کو اس حوالے سے نئے چیلنجز کا سامنا کرنا پڑا ہے۔ ابتدائی دور میں اُردو کی ڈیجیٹل ترجمہ نگاری لغت پر مبنی نظام رول بیسڈ ٹرانسلیشن تک محدود تھی جس میں الفاظ کے سادہ متبادلات فراہم کیے جاتے تھے۔ اس قسم کا ترجمہ غیر فطری اور ناقص ہوتا تھا۔ بعد میں شماریاتی ترجمہ اور پھر نیورل نیٹ ورکس پر مبنی ترجمہ کے نظام نے اُردو ترجمہ نگاری کے معیار میں نمایاں بہتری پیدا کر دی۔(8)

گوگل ٹرانسلیٹ، مائیکروسافٹ ٹرانسلیٹر اور دیگر آن لائن پلیٹ فارمز نے اُردو کو عالمی ڈیجیٹل منظر نامے میں شامل کرنے میں اہم کردار ادا کیا ہے۔ اگرچہ ان نظاموں میں ابھی بھی معنوی باریکیوں، شعری اظہار اور کلاسیکی نثر کے ترجمے میں مسائل موجود ہیں۔ تاہم روز بروز بہتری پیدا ہو رہی ہے۔ اُردو زبان دیگر بڑی زبانوں کے مقابلے میں ابھی اس میدان میں بہت پیچھے ہے اور اس کے لیے خصوصی کاوشوں کی ضرورت ہے۔ اس زبان کو ترجمہ نگاری کے بڑے پلیٹ فارمز کے ساتھ ہم آہنگ کرنے کے لیے اُردو دان طبقے کو دلچسپی لینے کی ضرورت ہے۔

اُردو زبان کو گوگل ٹرانسلیٹ نے اپنی ابتدائی زبانوں میں شامل نہیں کیا تھا لیکن جونہی اس نے اس زبان کی اہمیت کو تسلیم کیا تو اس کو اپنی سروس میں شامل کر لیا۔ اس وقت اُردو بھی باقاعدہ طور پر گوگل ٹرانسلیٹ کی ڈیجیٹل ترجمہ کی زبانوں میں شامل ہے۔

اُردو ایک پیچیدہ زبان ہے جس کی نحوی اور صرفی ساخت دیگر مغربی زبانوں یا انگریزی سے کافی مختلف ہے۔ اس میں مختلف زبانوں کے الفاظ کثیر تعداد میں موجود ہیں۔ ان وجوہات کی بنا پر اس کے ترجمے میں کافی مسائل ہیں۔ گوگل ٹرانسلیٹ ان مسائل پر قابو پانے کی کوشش کر رہا ہے۔ گوگل ٹرانسلیٹ کا موجودہ نظام نیورل نیٹ ورکس پر مشتمل ہے جو جملے کو سیاق وسباق کے مطابق مکمل طور پر سمجھنے کی کوشش کرتا ہے۔ اس سے اُردو ترجمے کے فطری بہاؤ میں کافی مدد مل رہی ہے۔ گوگل کی ترجمہ ایپلیکیشن کیونکہ آسانی سے دستیاب ہے اس لیے اس کا استعمال ہر کسی کے لیے میسر ہے۔ اس کی کثیر الجہتی خصوصیات بھی اس کے استعمال کو موثر بناتی ہیں۔ ٹیکسٹ ترجمہ، وائس آؤٹ پٹ، کیمرہ کے ذریعے ان پٹ کی سہولت، ہینڈ رائٹنگ فیچر وغیرہ اس کو مفید بناتے ہیں۔ گوگل ٹرانسلیٹ کے ذریعے تعلیمی اور تحقیقی سہولیات میسر آتی ہیں۔ گوگل اپنی ترجمہ کی سہولیات گوگل میٹ، جی میل اور دیگر پلیٹ فارمز میں بھی ریئل ٹائم سہولت کے ساتھ فراہم کرتا ہے جو اس کی اہمیت کو دوچند کر دیتا ہے۔(9)

اُردو کے حوالے سے گوگل ٹرانسلیٹ کو کچھ مسائل بھی درپیش ہیں۔ خاص طور پر سیاق وسباق کو مدنظر رکھتے ہوئے مختلف محاوروں، ضرب الامثال اور ثقافتی اشاروں کا یہ درست ترجمہ نہیں کر پاتا۔ اس کے علاوہ گرامر اور نحوی اغلاط بھی موجود ہوتی ہیں۔ ادبی اور تخلیقی متون کی تفہیم میں اسے دشواری کا سامنا کرنا پڑتا ہے اور بعض اوقات یہ لغوی ترجمے تک محدود رہتا ہے۔ پھر بھی گوگل ٹرانسلیٹ اُردو زبان کا دیگر زبانوں اور دیگر زبانوں کا اُردو میں ترجمہ کرنے کی حد تک کافی بہتر کارکردگی دکھانے والا پلیٹ فارم ہے۔ اگر اس کی تربیت اُردو کے مزید مواد پر کی گئی اور پھر اُردو کے لیے ڈیٹا کارپس مہیا کیے گئے تو معیار میں کافی بہتری آسکتی ہے۔

گوگل ٹرانسلیٹ کی طرح مائیکروسافٹ ٹرانسلیٹر بھی اُردو ترجمہ نگاری کے حوالے سے ایک اہم پلیٹ فارم ہے۔ یہ بھی نیورل مشین لرننگ کی ٹیکنیک کو اپناتا ہے اور تیز رفتار اور قابل فہم ترجمہ فراہم کرتا ہے۔ یہ اُردو جیسی ڈیجیٹل ترجمے کے حوالے سے پیچیدہ زبان کو بھی بخوبی سنبھال رہا ہے اور اس میں اس حوالے سے مزید بہتری آ رہی ہے۔ اگرچہ یہ گوگل ٹرانسلیٹ کے معیار تک نہیں پہنچا لیکن پھر بھی اس کی کچھ اپنی خصوصیات ہیں۔ یہ اُردو زبان کا ترجمہ پیش کرتے ہوئے نستعلیق فونٹ کو ظاہر کر سکتا ہے جس سے پڑھنے میں سہولت ہوتی ہے۔ اس کا ترجمہ بعض اوقات گوگل ٹرانسلیٹ سے زیادہ فصیح ہوتا ہے اور علمی نوعیت کے متون میں اس کی ساخت نسبتاً درست ہے۔ بعض اوقات یہ گوگل سے سیاق وسباق کے حوالے بہتر نتائج فراہم کرتا ہے۔

اس کی ایک خصوصیت یہ ہے کہ یہ مائیکروسافٹ ورڈ اور پاور پوائنٹ کے ساتھ ہم آہنگ ہے اور ایسے اہم سافٹ ویئرز میں یہ فوری ترجمہ فراہم کرتا ہے۔ تحقیقی مواد کے لیے بھی یہ بہترین انداز میں ترجمہ فراہم کرنے کی صلاحیت رکھتا ہے۔ بین الاقوامی سطح پر اُردو زبان کو سمجھنے اور اس کا فوری وائس ترجمہ کرنے میں بھی اس کا ثانی نہیں۔ یہ کاروباری ویب سائٹس اور ایپلیکیشنز میں بھی استعمال ہوتا ہے۔ نیوز ویب سائٹس اور بلاگز میں بھی اس کو عام طور پر استعمال کیا جاتا ہے۔ اس کے بعض

نقائص بھی ہیں۔ علمی اور ادبی متون کی سطح پر یہ ترجمہ نگاری میں اکثر اوقات غلطیاں کرتا ہے۔ اس کے مشینی ماڈلز گوگل کی نسبت زیادہ وسیع اور متنوع نہیں ہیں۔ پھر یہ کہ یہ مکمل آن لائن سروس ہے اور بغیر انٹرنیٹ کے یہ کام نہیں کرتا۔ اگرچہ ایک سے زیادہ لوگوں کے درمیان گفتگو کی سہولت فراہم کرتا ہے لیکن اُردو کے لیے یہ سہولت ابھی موجود نہیں۔ اُردو زبان کے حوالے سے بھی اس کی مقبولیت آہستہ آہستہ بڑھ رہی ہے اور مستقبل قریب میں یہ اپنی جگہ بنا لے گا۔

ڈیجیٹل ترجمہ نگاری کے میدان میں اے آئی پلیٹ فارمز نے بہت زیادہ ترقی کی ہے۔ ان میں ایمازون ٹرانسلیٹ، ڈیپ ایل اور اوپن اے آئی جیسے مصنوعی ذہانت کے ماڈلز شامل ہیں اُردو ترجمہ نگاری میں بھی یہ ماڈلز بہترین سہولیات مہیا کر رہے ہیں۔

اوپن اے آئی کا ماڈل جی پی ٹی 5.2 اس وقت مصنوعی ذہانت کی بہترین سہولیات مہیا کر رہا ہے۔ یہ اُردو زبان کی تفہیم کے حوالے سے بھی کافی مہارت رکھتا ہے۔ اسے نہ صرف اُردو زبان پر عبور حاصل ہے بلکہ ادبی اور تحقیقی اسلوب بیان کو سمجھنے اور اس کا ترجمہ کرنے کی بھی صلاحیت رکھتا ہے۔ اس کے علاوہ ثقافتی متون کے تراجم میں بھی یہ کافی بہتری دکھا رہا ہے۔ سیاق وسباق کو سمجھتے ہوئے درست ترجمہ پیش کرنے کی بھی صلاحیت رکھتا ہے۔ اسی لیے اساتذہ، بلاگرز، محققین اور مترجمین تحقیق وتدریس اور دیگر ضروریات کے لیے اس ٹول کو بہت زیادہ استعمال کر رہے ہیں۔ البتہ ابھی چند مسائل درپیش ہیں جس میں مواد کی کمی اور پھر یہ کہ اخلاقی اور ادبی مواد کو سمجھنے کے چیلنجز اس کی صلاحیتوں کے لیے مشکلات پیدا کرتے ہیں۔

ڈیپ ایل ٹرانسلیٹر ایک نیورل مشینی سروس ہے جو یورپی زبانوں میں ترجمہ کرنے کی بہترین صلاحیت رکھتی ہے۔ اگرچہ ابھی یہ براہ راست اُردو ترجمہ کرنے کے قابل نہیں لیکن پھر بھی یہ اے پی آئی اور دیگر پلیٹ فارمز کے تعاون سے اُردو متون کا ترجمہ کرنے کی سہولت فراہم کرتی ہے۔ اس کا ترجمہ کیونکہ نیورل نیٹ ورکس کی طاقت سے بھرپور ہوتا ہے اس لیے اس میں فصاحت موجود ہوتی ہے اور زبان کا قدرتی پن قائم رہتا ہے۔ البتہ اس کی صلاحیت اس حوالے سے کافی محدود ہے۔

ایمازون ٹرانسلیٹ ایک مشینی ترجمہ پلیٹ فارم ہے جو کاروباری معاملات اور ویب سائٹس کے ترجمے کے لیے تیار کیا گیا ہے اور اس میں اُردو ترجمے کی سہولت بھی موجود ہے۔ اس کی خصوصیت یہ ہے کہ یہ اے پی آئی فراہم کرتا ہے جو مختلف پلیٹ فارمز اور ویب سائٹس کے ساتھ منسلک کر کے براہ راست اُردو کے ترجمے کی سہولت فراہم کرتی ہیں۔ اس کے ذریعے اُردو زبان کا تیز ترجمہ حاصل کرنا ممکن ہوتا ہے۔ یہ خاص طور پر ای کامرس اور ٹیکنیکل مضامین کے تراجم کے لیے کافی اہم ہے۔ البتہ اس میں بھی ادبی، ثقافتی اور محاوراتی تراجم کے حوالے سے کمزوری موجود ہے اور اس کا ایک بار انسانی سطح پر جائزہ لینے کی ضرورت ہوتی ہے۔

Meta AI's SeamlessM4T اے آئی ماڈل تقریر سے تقریر، متن سے تقریر اور متن سے متن کا ترجمہ فراہم کرنے میں مہارت رکھتا ہے اور کئی زبانوں کا بیک وقت ترجمہ ہینڈل کرنے کی صلاحیت رکھتا ہے جن میں اُردو زبان بھی شامل ہے۔ آواز سے آواز کا ترجمہ رکھنے کی صلاحیت اس کو بے مثال بناتی ہے اور یہ اُردو کے ساتھ ساتھ اس سے منسلک علاقائی زبانوں کی پہچان بھی فراہم کرتا ہے۔ اسی طرح دیگر پلیٹ فارمز میں یانڈیکس شامل ہے جو روسی پلیٹ فارم ہے اور اُردو ترجمے کی سہولت فراہم کرتا ہے۔ آئی ٹرانسلیٹ ایک موبائل ایپلیکیشن ہے جو اُردو کا دیگر زبانوں کے ساتھ ترجمہ فراہم کرنے کی سہولت دیتی ہے۔(10)

یہ جدید اے آئی ٹولز اور پلیٹ فارمز اُردو ترجمے کے نئے امکانات ہمارے سامنے لا رہے ہیں اور اُردو جیسی ڈیجیٹل میدان میں کم ترقی یافتہ زبان کو بھی دیگر زبانوں کے ہم پلہ لانے میں معاونت کر رہے ہیں۔

موبائل ایپلیکیشنز بھی اُردو ترجمے کے سلسلے میں کافی خدمات سرانجام دے رہی ہیں۔ ان میں گوگل ٹرانسلیٹ ایپ، آئی ٹرانسلیٹ جیسی ایپلیکیشنز شامل ہیں۔ اس طرح میٹ ٹرانسلیٹ، گوگل ڈکشنری بھی ترجمے کی سہولت فراہم کرتی ہیں۔ ان ایپلیکیشنز کے بہت سے مثبت پہلو ہیں۔ یہ صوتی ترجمہ بھی فراہم کرتی ہیں، تصویری ترجمے کے ذریعے بھی یہ بصری وضاحت فراہم کرتی ہیں۔ البتہ ان ایپلیکیشنز کا ترجمہ غیر فصیح ہوتا ہے اور زیادہ تر یہ لغوی ترجمے پر انحصار کرتی ہیں۔

مختلف آن لائن ڈکشنریز بھی اُردو زبان کے حوالے سے موجود ہیں مثلاً اُردو پوائنٹ، ریختہ ڈکشنری، آکسفورڈ اُردو انگلش ڈکشنری۔ یہ ڈکشنریز بھی ترجمے کے حوالے سے معاونت کرتی ہیں۔ ان میں الفاظ کے معنی، تلفظ، مترادفات، استعمالات وغیرہ کا تعارف کرایا جاتا ہے۔ اسی طرح علمی ترجمے میں بھی ان لغات کا حوالہ قابل اعتبار ٹھہرتا ہے۔ البتہ ان لغات کی کچھ خامیاں بھی ہیں۔ یہ جملوں کی سطح پر ترجمہ فراہم نہیں کرتیں، سست رفتار ہیں اور محدود ذخیرہ الفاظ رکھتی ہیں۔

اُردو ترجمے کے تکنیکی ولسانی چیلنجز:

اُردو زبان اپنی تاریخ، ساخت اور لسانی تنوع کے اعتبار سے دنیا کی ایک نہایت پیچیدہ اور کثیر الجہتی زبان ہے۔ برصغیر کی مختلف زبانوں کے ملاپ سے جنم لینے والی اس زبان میں فارسی، عربی، ترکی اور مقامی بولیوں کے اثرات نمایاں ہیں۔(11) یہی تنوع اسے معنوی وسعت بخشتا ہے لیکن جب بات ڈیجیٹل یا مشینی ترجمے کی آتی

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

ہے تو یہی تنوع ایک بڑی رکاوٹ بھی بن جاتا ہے۔ دوسری زبانوں خصوصاً انگریزی یا یورپی زبانوں کے برعکس اُردو کا رسم الخط جمیل نستعلیق ہے، جس کی ڈیجیٹل نمائندگی اور خودکار پراسیسنگ ہمیشہ سے مشکل رہی ہے۔

ڈیجیٹل ترجمے کے عمل میں دو بڑی اقسام کی مشکلات سامنے آتی ہیں: لسانی چیلنجز اور تکنیکی مسائل۔ لسانی چیلنجز میں الفاظ کے ایک سے زائد معانی، جملے کی ساخت، محاورات اور ثقافتی اظہار کی پیچیدگیاں شامل ہیں۔ مثال کے طور پر اُردو میں "دل" کا لفظ نہ صرف جسمانی عضو کے لیے استعمال ہوتا ہے بلکہ جذبات، احساسات اور تعلقات کے لیے بھی ایک علامتی لفظ ہے۔ ایسے الفاظ کے درست ترجمے کے لیے صرف لغت کافی نہیں بلکہ سیاق وسباق کی گہری فہم ضروری ہے۔ اسی طرح محاورے اور کہاوتیں اکثر لفظی ترجمے سے اپنا اصل مفہوم کھو دیتی ہیں۔(12)

دوسری جانب، تکنیکی سطح پر Natural Language Processing (NLP) کے مسائل جیسے Tokenization ، Morphological Parsing، اور Syntax Analysis اُردو کے لیے ابھی تک مکمل طور پر حل نہیں ہو سکے۔ اُردو کی رسم الخطی مشکلات، مثلاً الفاظ کے درمیان اسپیس کا غیر مستقل استعمال، اعراب کی کمی، اور ہم آواز الفاظ کی کثرت، کمپیوٹر کو درست تجزیہ اور ترجمہ کرنے میں مشکلات پیدا کرتی ہیں۔ Unicode Compatibility اور User Interface جیسے مسائل مزید پیچیدگی میں اضافہ کرتے ہیں کیونکہ بیشتر عالمی ترجمہ پلیٹ فارمز اُردو کے لیے وہ سہولت اور درستی فراہم نہیں کرتے جو دیگر زبانوں مثلاً انگریزی یا فرانسیسی کے لیے ممکن ہے۔

ذیل میں ان لسانی و تکنیکی چیلنجز کا جائزہ لیں گے جو اُردو ڈیجیٹل ترجمہ نگاری کی راہ میں حائل ہیں:

اُردو زبان کا رسم الخط اس کی شناخت کا نہایت اہم جزو ہے۔ اُردو، عربی اور فارسی رسم الخط سے اخذ شدہ ہے اور زیادہ تر خطِ نستعلیق میں لکھی جاتی ہے۔ یہ رسم الخط جمالیات اور حسنِ تحریر میں اپنی مثال آپ ہے لیکن جب ہم اس کی ڈیجیٹل نمائندگی یا مشینی ترجمے کی طرف بڑھتے ہیں تو یہی خوبی ایک بڑی تکنیکی پیچیدگی میں بدل جاتی ہے۔

اُردو میں اسپیس (Space) کا استعمال انگریزی کی طرح سخت قاعدے کا پابند نہیں ہے۔ کئی بار کمپیوٹر الفاظ کو ایک دوسرے سے الگ نہیں کر پاتا کیونکہ صارفین اسپیس کا استعمال یا تو غیر مستقل انداز میں کرتے ہیں یا بالکل نظر انداز کر دیتے ہیں۔ مثلاً "خوش خط" اور "خوشخط" دو مختلف اشکال ہیں لیکن معنوی طور پر ایک ہیں۔ مشینی ترجمہ ان دونوں کو مختلف یونٹس کے طور پر سمجھتا ہے، جس سے ترجمہ کی درستی متاثر ہوتی ہے۔

اُردو زبان میں روزمرہ تحریر میں اعراب (زبر، زیر، پیش) عموماً نہیں لگائے جاتے۔ یہ اعراب معنی میں بڑا فرق پیدا کرتے ہیں۔ مثال کے طور پر "عَلَم" اور "عِلم" کے الگ معانی ہیں۔ لیکن اعراب نہ ہونے کی وجہ سے کمپیوٹر اکثر ان میں فرق نہیں کر پاتا۔ یہی ابہام ترجمے میں غلطیوں کا باعث بنتا ہے۔

ڈیجیٹل دنیا میں اُردو کو زیادہ تر نسخ فونٹ میں پیش کیا جاتا ہے کیونکہ یہ سیدھا اور تکنیکی طور پر کم پیچیدہ ہے۔ لیکن اُردو قارئین اور ادبی روایت نستعلیق سے مانوس ہے۔ یہ فرق نہ صرف جمالیاتی مسئلہ پیدا کرتا ہے بلکہ بعض اوقات ترجمے کے پلیٹ فارمز پر حروف کی ترتیب بھی بگڑ جاتی ہے، جس سے متن کی معنویت اور روانی متاثر ہوتی ہے۔

اُردو رسم الخط کو عالمی ڈیجیٹل نظاموں میں شامل کرنے کے لیے Unicode بنایا گیا ہے۔ اگرچہ Unicode نے اُردو کے لیے کافی سہولت فراہم کی ہے لیکن اس کے باوجود کئی پلیٹ فارمز پر مکمل مطابقت موجود نہیں۔ بعض اوقات حروف ٹوٹے ہوئے یا الٹے سیدھے نظر آتے ہیں، اور کمپیوٹر ان کو درست طریقے سے پراسیس نہیں کر پاتا۔ یہ مسئلہ خاص طور پر ان مشینی ترجمہ پلیٹ فارمز پر سامنے آتا ہے جو اُردو کو کم ترجیح دیتے ہیں۔

یہ تمام رسم الخطی پیچیدگیاں براہِ راست مشینی ترجمے کو متاثر کرتی ہیں۔ جب بنیادی یونٹ یعنی "لفظ" ہی درست طریقے سے شناخت نہ ہو سکے تو کمپیوٹر کے لیے اس کے معنی اور سیاق کو سمجھنا تقریباً ناممکن ہو جاتا ہے۔ اس کے نتیجے میں ترجمہ یا تو غلط نکلتا ہے یا بالکل غیر مربوط ہوتا ہے۔ مثال کے طور پر گوگل ٹرانسلیٹ میں کئی بار اُردو جملوں کو لفظ بہ لفظ توڑ کر الگ الگ ترجمہ کر دیا جاتا ہے، جس سے مفہوم مکمل طور پر بدل جاتا ہے۔

اُردو زبان کی سب سے بڑی خوبصورتی اور سب سے بڑی مشکل اس کی معنوی وسعت ہے۔ اُردو کے الفاظ اکثر کئی کئی معانی رکھتے ہیں، اور ان معانی میں نہ صرف لغوی فرق پایا جاتا ہے بلکہ سیاق وسباق کے لحاظ سے ان کے استعمال کی جہت بھی بدل جاتی ہے۔ یہی خصوصیت جہاں زبان کو ادبی اور شعری رنگ عطا کرتی ہے، وہیں مشینی ترجمہ کے لیے ایک نہایت بڑا چیلنج بن جاتی ہے۔

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

لسانیات میں ایسے الفاظ کو کثیر المعنی الفاظ (Polysemous Words) کہا جاتا ہے جو ایک ہی شکل رکھتے ہوں مگر مختلف معنوں میں استعمال ہوتے ہوں۔ اُردو میں یہ رجحان بہت زیادہ ہے کیونکہ اُردو نے مختلف زبانوں سے الفاظ مستعار لیے اور ان کو اپنے سیاق میں ڈھالا۔ مثلاً عربی کا لفظ "علم" اُردو میں بھی آ گیا، لیکن یہ کبھی "جاننے" کے فعل کے طور پر، کبھی "علم" کے اسم کے طور پر، اور کبھی "پرچم" کے معنی میں استعمال ہوتا ہے۔

انسان جب زبان کو سمجھتا ہے تو وہ الفاظ کو الگ الگ یونٹ کے طور پر نہیں پڑھتا بلکہ پورے جملے اور اس کے تناظر کو دیکھتا ہے۔ ہم جانتے ہیں کہ "کتاب کھولو" میں "کھولو" کا مطلب کسی شے کو کھولنے کی حرکت ہے، لیکن "کھولو کے جنگل" میں یہی لفظ ایک مقام یا علاقے کے نام کے طور پر آتا ہے۔ انسان سیاق و سباق سے فوراً پہچان لیتا ہے، مگر کمپیوٹر کے لیے یہ فیصلہ آسان نہیں ہوتا۔

اُردو میں مترادفات کا کثرت سے پایا جانا بھی ترجمے کو پیچیدہ بنا دیتا ہے۔ مثلاً "غم"، "دکھ"، "کرب"، "صدمہ"، "رنج"، "ملال" سب ایک دوسرے کے قریب المعنی ہیں لیکن ہر لفظ کا استعمال اور جذباتی شدت مختلف ہے۔ ایک انسانی مترجم اس فرق کو سمجھتا ہے اور موقع کے مطابق موزوں لفظ منتخب کرتا ہے، لیکن مشینی ترجمہ اکثر ان سب کو ایک ہی زمرے میں ڈال دیتا ہے اور نتیجتاً ترجمہ اپنی باریکی اور لطافت کھو بیٹھتا ہے۔

اسی طرح انگریزی کے لفظ "Light" کو لیجیے۔ یہ کبھی "روشنی" کے طور پر ترجمہ ہوگا، کبھی "ہلکا" کے طور پر، اور کبھی "آگ جلانا" یا "ہلکا پھلکا" کے معنوں میں۔ مشینی ترجمے کے لیے یہ فیصلہ کرنا نہایت دشوار ہے کہ مخصوص جملے میں "Light" کا صحیح مطلب کیا ہے۔ اُردو میں بھی یہی کیفیت بارہا سامنے آتی ہے۔

مشینی ترجمے میں ابہام کو دور کرنے کے لیے ماڈلز عموماً **Statistical Models** یا **Neural Networks** استعمال کرتے ہیں جو بڑے ڈیٹا سے سیکھتے ہیں۔ لیکن اُردو کے لیے کارپس (Corpus) اور معیاری ڈیٹا کی کمی اس عمل کو مشکل بنا دیتی ہے۔ جہاں انگریزی کے لیے لاکھوں جملے اور مثالیں موجود ہیں، وہاں اُردو کے لیے ایسا مواد محدود ہے۔ نتیجہ یہ ہے کہ مشینی ترجمہ اکثر اُردو کے کثیر المعنی الفاظ کا صحیح مطلب منتخب نہیں کر پاتا۔(13)

مزید یہ کہ اُردو کا نحوی ڈھانچہ (Syntax) انگریزی سے مختلف ہے۔ انگریزی میں جملے کی ترتیب زیادہ تر Subject-Verb-Object (SVO) ہوتی ہے، لیکن اُردو میں اکثر Subject-Object-Verb (SOV) ترتیب استعمال ہوتی ہے۔ اس فرق کی وجہ سے مشینی ترجمہ صرف الفاظ کے معنی ہی نہیں بلکہ جملے کی ترتیب میں بھی غلطیاں کر دیتا ہے۔

جب کثیر المعنی الفاظ ادبی یا شعری متن میں استعمال ہوتے ہیں تو ترجمے کا مسئلہ کئی گنا بڑھ جاتا ہے۔ شاعری میں ایک ہی لفظ کئی پرتوں میں معنی رکھتا ہے، اور قاری اپنی بصیرت اور تجربے کے مطابق ان معنوں کو سمجھتا ہے۔ مثال کے طور پر غالب کا شعر:

دل ہی تو ہے نہ سنگ و خشت، درد سے بھر نہ آئے کیوں

اس میں "دل" نہ صرف جسمانی عضو ہے بلکہ جذبات، احساسات اور انسان کے وجود کی علامت بھی ہے۔ مشینی ترجمہ اس گہرائی کو نہیں سمجھ سکتا اور "دل" کو صرف "Heart" کے طور پر پیش کر دیتا ہے، جس سے شعر کی تہہ دار معنویت ختم ہو جاتی ہے۔

لسانی ماہرین اور کمپیوٹر سائنسدانوں نے اس مسئلے کے حل کے لیے مختلف تکنیکیں استعمال کی ہیں۔ ایک اہم طریقہ **Word Sense Disambiguation (WSD)** ہے جس کا مقصد یہ ہے کہ جملے کے سیاق کو دیکھ کر فیصلہ کیا جائے کہ لفظ کا کون سا مطلب مراد ہے۔ لیکن اُردو کے لیے یہ عمل ابھی تک ابتدائی مراحل میں ہے کیونکہ اس کے لیے بڑے پیمانے پر لیبل شدہ ڈیٹا درکار ہے۔

اسی طرح **Neural Machine Translation (NMT)** میں Contextual Embeddings جیسے BERT یا GPT ماڈلز استعمال کیے جا رہے ہیں، جو سیاق کے لحاظ سے الفاظ کے معنی پہچاننے کی صلاحیت رکھتے ہیں۔ یہ طریقہ اُردو کے لیے بھی امکانات رکھتا ہے، لیکن اس کے لیے اُردو کی زیادہ وسیع اور معیاری ڈیٹا بیس تیار کرنا ضروری ہے۔

اُردو زبان کی دلکشی اور اظہار کی قوت کا سب سے نمایاں پہلو اس کے محاورات، کہاوتوں اور استعاراتی اظہار میں مضمر ہے۔ یہ وہ پہلو ہیں جو اُردو کو صرف ایک ذریعۂ ابلاغ نہیں بلکہ ایک تہذیبی اور ادبی ورثہ بھی بناتے ہیں۔ تاہم یہی پہلو مشینی ترجمہ اور ڈیجیٹل زبان پراسیسنگ کے لیے سب سے زیادہ چیلنجنگ ہیں۔ محاورے اور کہاوتیں ایسی پیچیدگیاں پیدا کرتے ہیں جو صرف لغت سے حل نہیں ہوتیں بلکہ گہری ثقافتی اور ادبی بصیرت کی متقاضی ہیں۔

محاورہ بنیادی طور پر ایسے جملے یا ترکیب کو کہا جاتا ہے جو عام الفاظ سے زیادہ گہرا اور طے شدہ معنوی تاثر رکھتی ہو۔(14) مثال کے طور پر "آسمان سر پر اٹھا لینا" کا مطلب یہ نہیں کہ واقعی آسمان کو سر پر رکھا جا رہا ہے، بلکہ اس سے مراد شور مچانا یا ہنگامہ کھڑا کرنا ہے۔ انسانی قاری فوری طور پر یہ مفہوم سمجھ لیتا ہے کیونکہ وہ زبان اور معاشرتی سیاق سے واقف ہوتا ہے۔ لیکن مشینی ترجمہ عموماً اسے "To lift the sky on the head" کے طور پر ترجمہ کر دیتا ہے، جو غیر فطری اور ناقابلِ فہم ہے۔

محاورے زبان میں لطافت، مزاح اور تہذیبی اشارے پیدا کرتے ہیں۔ اُردو میں ہزاروں محاورے ہیں جو مختلف خطوں اور بولیوں سے آئے ہیں۔ ان سب کو ڈیجیٹل ترجمے میں درست طور پر شامل کرنا ایک انتہائی مشکل کام ہے۔

کہاوتیں یا Proverbs اُردو زبان میں عوامی دانش کی نمائندہ ہیں۔ مثلاً "اونٹ کے منہ میں زیرا" ایک عام کہاوت ہے جس سے مراد ہے کہ کسی بڑے نظام میں معمولی چیز کا کوئی اثر نہیں۔ اگر اس کا لفظی ترجمہ کیا جائے تو یہ بے معنی معلوم ہوگا، لیکن اصل کہاوت کی معنویت اس کے پس منظر اور صدیوں پرانی عوامی حکمت سے جڑی ہوئی ہے۔(15) مشینی ترجمہ ایسی کہاوتوں کو لفظ بہ لفظ ترجمہ کر دیتا ہے، اور یوں قاری اصل مقصد کو سمجھنے سے قاصر رہتا ہے۔ مثلاً انگریزی میں اس کا درست متبادل "A drop in the ocean" یا "Too little to make a difference" ہے، مگر اس مطابقت کو مشین کے لیے پہچاننا تقریباً ناممکن ہے جب تک کہ اس کے پاس وسیع پیمانے پر محاوراتی ڈیٹا اور کراس-لینگول کلچر میپنگ موجود نہ ہو۔

استعارہ زبان کی وہ تکنیک ہے جس کے ذریعے ایک لفظ یا ترکیب کو اس کے اصل معنی کے بجائے کسی اور مفہوم میں استعمال کیا جاتا ہے تاکہ جذباتی یا تخیلی اثر پیدا ہو۔ اُردو شاعری اور نثر میں استعارات کا بہت استعمال ہوتا ہے۔ مثال کے طور پر "وقت کا پہیہ رک نہیں سکتا" میں پہیہ کوئی حقیقی پہیہ نہیں بلکہ ایک استعاراتی علامت ہے۔

اسی طرح "زندگی ایک دریا ہے" میں دریا بطور علامت استعمال ہوا ہے۔ مشینی ترجمہ جب ان جملوں کو پراسیس کرتا ہے تو اکثر لفظی ترجمہ کر دیتا ہے اور استعاراتی معنویت کھو بیٹھتا ہے۔ انگریزی میں بھی استعاراتی زبان پائی جاتی ہے، مگر دونوں زبانوں کے استعاراتی نظام یکساں نہیں، اس لیے براہِ راست ترجمہ اکثر ناکام رہتا ہے۔

اُردو زبان میں طنز اور مزاح اکثر محاورات اور کہاوتوں کے ذریعے ادا کیا جاتا ہے۔ مثلاً "اونٹ رے اونٹ تیری کون سی کل سیدھی"۔ یہ کہاوت طنزیہ ہے اور اس کا مطلب ہے کہ کسی فرد میں کوئی خوبی یا توازن نہیں۔ اگر مشین اس کا لفظی ترجمہ ، "O camel which of your joints is straight?" کر دے تو یہ نہ صرف بامعنی نہیں رہتا بلکہ طنز کا پہلو بھی کھو جاتا ہے۔ یہی وجہ ہے کہ طنزیہ اور مزاحیہ ادب کو مشینی ترجمے میں سب سے زیادہ نقصان پہنچتا ہے۔ انسانی مترجم بھی طنز کے صحیح ترجمے میں مشکل محسوس کرتا ہے، تو ایک خودکار نظام کے لیے یہ اور بھی زیادہ دشوار ہے۔

اُردو چونکہ برصغیر کی ایک بڑی زبان ہے، اس لیے اس کے محاورے اور کہاوتیں مختلف خطوں اور ثقافتی ماحول کے مطابق مختلف ہیں۔ پنجاب، سندھ، خیبر پختونخوا اور بھارت کے مختلف علاقوں میں الگ الگ کہاوتیں اور محاورے رائج ہیں۔ مثال کے طور پر "جیسے کو تیسا" ایک عام محاورہ ہے لیکن بعض خطوں میں یہی مفہوم "اینٹ کا جواب پتھر سے دینا" کے ذریعے بیان کیا جاتا ہے۔

مشینی ترجمے کے لیے یہ تنوع ایک الگ مسئلہ پیدا کرتا ہے، کیونکہ وہ اس علاقائی فرق کو پہچاننے سے قاصر ہے۔ اسی لیے جب کوئی قاری اپنے علاقے کا مخصوص محاورہ استعمال کرتا ہے تو مشینی ترجمہ اس کا مطلب نہیں سمجھ پاتا اور غیر موزوں ترجمہ فراہم کرتا ہے۔

لسانی ماہرین کے مطابق محاورات اور کہاوتوں کے درست ترجمے کے لیے ضروری ہے کہ الگ **Idiomatic Corpus** بنایا جائے جس میں ہزاروں محاورے اور ان کے انگریزی یا دیگر زبانوں میں مساوی متبادل درج ہوں۔ اس کے ساتھ ساتھ **Phrase-Based Translation Models** اور **Neural Machine Translation** میں Idiom Recognition کی صلاحیت پیدا کی جائے تاکہ سسٹم محاورے کو عام الفاظ کے بجائے ایک "فکسڈ یونٹ" کے طور پر پہچان سکے۔

جدید NLP میں Contextualized Embeddings جیسے GPT ، BERT یا LLaMA میں اس مسئلے کو کافی حد تک حل کرنے کی صلاحیت ہے، کیونکہ یہ ماڈلز پورے جملے کے تناظر کو دیکھ کر فیصلہ کرتے ہیں۔ لیکن اُردو کے لیے ابھی تک ایسے بڑے پیمانے پر محاوراتی ڈیٹا موجود نہیں جس سے ان ماڈلز کو موثر طریقے سے تربیت دی جا سکے۔

ترجمہ محض الفاظ کا تبادلہ نہیں بلکہ دو زبانوں اور دو تہذیبوں کے درمیان مکالمہ ہے۔ لسانی ساخت اور قواعد تو کسی حد تک مشین کے ذریعے سمجھائے جاسکتے ہیں، مگر زبان میں پیوستہ وہ ثقافتی معانی جو صدیوں کے تجربے، عقیدے، ادب اور معاشرتی رویوں کا نچوڑ ہوتے ہیں، ان کا ترجمہ سب سے زیادہ پیچیدہ مرحلہ ثابت ہوتا ہے۔ اُردو زبان اپنی ساخت میں محض ایک اظہار کا ذریعہ نہیں بلکہ اس کے پس منظر میں برصغیر کی تہذیبی، مذہبی اور ادبی روایات کا ایک وسیع سرمایہ موجود ہے۔ یہی وجہ ہے کہ جب مشینی ترجمہ ان متون کو دوسری زبانوں میں منتقل کرتا ہے تو اکثر معانی مسخ ہو جاتے ہیں، محاورات اور استعارے کھوکھلے ہو جاتے ہیں، اور قاری اصل متن کے ثقافتی تناظر کو سمجھنے سے قاصر رہتا ہے۔

اُردو میں قرآن، حدیث، اور اسلامی روایات کے حوالے عام ہیں۔ ان کا مشینی ترجمہ اکثر غیر مناسب یا سیاق سے ہٹ کر ہو جاتا ہے، کیونکہ ان کے الفاظ صرف لغوی نہیں بلکہ مقدس اور عقیدتی حیثیت رکھتے ہیں۔

اُردو ادب اور صحافت میں ایسے جملے عام ہیں جو معاشرتی رویوں یا تاریخی تجربات کو ظاہر کرتے ہیں، مثلاً "تقسیم کا المیہ" اس کے ترجمے میں صرف لفظی تبدیلی کافی نہیں، بلکہ اس کے تاریخی اور ثقافتی تناظر کی تفہیم ضروری ہے۔

مشینی ترجمہ زیادہ تر الفاظ کو لفظی سطح پر ڈھالتا ہے، اس لیے محاورہ یا کہاوت اپنے اصل معنوی جوہر سے محروم ہو جاتی ہے۔ اُردو میں ایک ہی لفظ کے کئی معنوی پہلو ہو سکتے ہیں، جو صرف سیاق کے ذریعے واضح ہوتے ہیں۔ لیکن مشین اکثر پورے سیاق کو مدنظر نہیں رکھتی۔ اُردو ادب کی سب سے بڑی خوبی اس کا اسلوب اور شعری جمالیات ہیں۔ مشینی ترجمہ ان جمالیاتی پہلوؤں کو برقرار نہیں رکھ سکتا۔ (16) نتیجہ یہ نکلتا ہے کہ ایک پر اثر شعر یا افسانے کا ترجمہ محض سادہ بیانیہ لگتا ہے۔ مشین کے پاس دنیا کے تمام ثقافتی پس منظر کی مکمل معلومات نہیں ہوتیں۔

ثقافتی متون کے مؤثر ترجمے کے لیے یہ تجاویز پیش کی جاسکتی ہیں: اُردو کے عام محاورے، کہاوتیں اور اصطلاحات کا ایک جامع ڈیٹابیس تیار کیا جائے جسے مشینی ترجمے کے ساتھ ضم کیا جائے۔ ایسے ماڈلز تیار کیے جائیں جو مخصوص ثقافتی حوالوں کو پہچان سکیں اور انہیں متعلقہ زبان کے مناسب متبادلات سے بدل سکیں۔ ثقافتی متون کے ترجمے میں صرف مشین پر انحصار درست نہیں۔ انسانی مترجمین کے ساتھ ہائبرڈ ماڈلز استعمال کیے جائیں تاکہ مشینی کارکردگی اور انسانی فہم کا امتزاج ہو۔ ادبی اور ثقافتی متون پر مبنی مخصوص کارپس تیار کر کے جدید ماڈلز جیسے GPT، BART یا T5 کو Fine-Tune کیا جائے تاکہ وہ محاورات اور ثقافتی جملوں کو بہتر طور پر سمجھ سکیں۔ Crowdsourcing پلیٹ فارمز کے ذریعے عوامی سطح پر مشینی ترجمے کی غلطیوں کو درست کر کے سسٹم کو مزید بہتر بنایا جاسکتا ہے۔

اُردو ڈیجیٹل ترجمہ نگاری میں سب سے بڑا چیلنج یہ ہے کہ زبان کو ایسے انداز میں ڈیجیٹل بنایا جائے کہ کمپیوٹر اس کے جملوں، الفاظ اور معانی کو درست طور پر پہچان سکے۔ اس مقصد کے لیے **Natural Language Processing (NLP)** بنیادی کردار ادا کرتا ہے۔ NLP ایک ایسا علمی و عملی میدان ہے جو زبان اور کمپیوٹر کے درمیان فاصلہ کم کرنے کی کوشش کرتا ہے۔ انگریزی اور یورپی زبانوں کے لیے NLP پر دہائیوں سے تحقیق ہو رہی ہے، لیکن اُردو کے لیے اس شعبے میں تحقیق نسبتاً نئی ہے اور ابھی تک ابتدائی نوعیت کے مسائل حل ہو رہے ہیں۔

اُردو NLP میں سب سے پہلا مسئلہ **Tokenization** کا ہے۔ Tokenization کا مطلب یہ ہے کہ کسی جملے کو ٹکڑوں (tokens) میں تقسیم کیا جائے تاکہ کمپیوٹر ہر لفظ کو الگ الگ سمجھ سکے۔ انگریزی میں الفاظ کے درمیان اسپیس ہمیشہ واضح ہوتی ہے، لیکن اُردو میں اسپیس کے استعمال میں یکسانیت نہیں پائی جاتی۔ اس مسئلے کی وجہ سے اُردو کے مشینی ترجمے میں غلطیاں بڑھ جاتی ہیں۔

دوسرا مسئلہ **Morphological Parsing** کا ہے۔ Morphology کا تعلق الفاظ کے اندرونی ڈھانچے سے ہوتا ہے۔ اُردو میں ایک لفظ کئی طرح کے صیغوں، لاحقوں اور سابقوں کے ذریعے اپنی شکل بدل لیتا ہے۔ کمپیوٹر کے لیے ان سب کو پہچاننا اور ان کے درست معنی اخذ کرنا ایک نہایت پیچیدہ کام ہے۔

Syntax Parsing یا نحوی تجزیہ اُردو کے لیے ایک اور بڑا چیلنج ہے۔ Syntax کا مطلب ہے جملے کی ساخت اور الفاظ کے درمیان تعلق۔ انگریزی میں جملے کی ترتیب عموماً Subject-Verb-Object (فاعل، فعل، مفعول) پر مشتمل ہوتی ہے، لیکن اُردو میں ترتیب بدل سکتی ہے۔ یہی وجہ ہے کہ مشینی ترجمے میں اُردو جملوں کی ساخت اکثر بگڑ جاتی ہے اور ترجمہ درست طور پر سامنے نہیں آتا۔

اُردو میں **Named Entity Recognition (NER)** بھی ایک مشکل مرحلہ ہے۔ NER کا مقصد یہ ہے کہ کسی جملے میں موجود شخصیات، مقامات یا اداروں کے ناموں کو پہچانا جائے۔ انگریزی میں یہ نسبتاً آسان ہے کیونکہ نام اور عام الفاظ کی ساخت میں فرق ہوتا ہے، لیکن اُردو میں "ایوب" ایک شخص کا نام بھی ہو سکتا ہے اور "ایوبی" ایک صفت بھی۔ یہ فرق کمپیوٹر کے لیے آسانی سے سمجھنا ممکن نہیں ہوتا۔

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Unicode Compatibility بھی ایک عملی مسئلہ ہے۔ اُردو کے رسم الخط کے لیے مختلف سافٹ ویئر اور پلیٹ فارمز پر مختلف فونٹس اور کوڈنگ استعمال کی جاتی ہے۔ بعض اوقات ایک ہی لفظ مختلف سسٹمز پر مختلف انداز میں ظاہر ہوتا ہے۔ مثال کے طور پر "ی" اور "ے" کی پہچان یا "ہ" اور "ۃ" کا فرق کمپیوٹر کو الجھا دیتا ہے۔ یہی وجہ ہے کہ ایک ہی ٹیکسٹ مختلف پلیٹ فارمز پر درست طور پر پہچانا نہیں جاپاتا۔

ان مسائل کو حل کرنے کے لیے دنیا بھر میں مختلف تحقیقی کوششیں جاری ہیں۔ اُردو کے لیے Corpus بنانا، Treebanks تیار کرنا، اور Rule-based کے بجائے Machine Learning پر مبنی ماڈلز تشکیل دینا اس سمت میں اہم اقدامات ہیں۔(17) تاہم، ابھی تک انگریزی یا چینی جیسی زبانوں کی طرح جامع حل دستیاب نہیں۔ اُردو کے لیے ایک بڑے، معیاری اور متوازن ڈیٹا بیس کی ضرورت ہے تاکہ مشین لرننگ اور ڈیپ لرننگ ماڈلز کو درست طور پر تربیت دی جاسکے۔

ترجمہ نگاری کے لیے بنیادی ضرورت ایک جامع اور معیاری لغت ہے۔ اُردو کے حوالے سے سب سے بڑی رکاوٹ یہ ہے کہ اگرچہ کاغذی لغات کی ایک طویل روایت موجود ہے لیکن ڈیجیٹل لغات اور ان کے قابلِ استعمال ورژن ابھی تک محدود ہیں۔ بیشتر لغات پی ڈی ایف یا اسکین کی صورت میں دستیاب ہیں جو مشین ریڈ ایبل نہیں۔ اس کے علاوہ بڑے پیمانے پر ایسے لغوی ذخائر (lexical databases) کی کمی ہے جن میں الفاظ کے معنوی تعلقات، مترادفات، متضاد الفاظ، محاوراتی استعمال اور ثقافتی باریکیاں باقاعدہ ڈیٹا بیس کی شکل میں درج ہوں۔ انگریزی میں **WordNet** یا **Oxford Dictionary API** جیسے وسائل موجود ہیں لیکن اُردو کے لیے ان کی کمی محسوس ہوتی ہے۔

ڈیجیٹل ترجمہ نگاری میں کامیابی اسی وقت ممکن ہے جب لغت کے ذخائر وسیع، معیاری اور مشین کے لیے قابلِ فہم ہوں۔ بدقسمتی سے اُردو میں اس سطح کے وسائل ابھی تک پوری طرح تیار نہیں ہوسکے۔ اگرچہ **CRULP (Center for Research in Urdu Language Processing)** اور دیگر تحقیقی ادارے اُردو لغت سازی اور کارپس ڈویلپمنٹ پر کام کر رہے ہیں، لیکن ان کا دائرہ محدود ہے اور عالمی سطح کے مشینی ماڈلز جیسے گوگل ٹرانسلیٹ یا ChatGPT کے لیے یہ ڈیٹا ناکافی ہے۔ یہی وجہ ہے کہ اُردو ترجمے میں اکثر مشینیں درستی اور معنوی گہرائی کھو بیٹھتی ہیں۔

اُردو زبان کے ڈیجیٹل ترجمے کے مسائل صرف لسانی اور معنوی سطح تک محدود نہیں رہتے بلکہ عملی اور تکنیکی سطح پر بھی نہایت پیچیدہ ہیں۔ جب کسی زبان کو کمپیوٹر یا موبائل سافٹ ویئر کے ذریعے دنیا بھر کے صارفین تک پہنچایا جاتا ہے تو اس میں صرف ترجمے کے الگورتھمز ہی اہم نہیں ہوتے، بلکہ اس زبان کی نمائندگی (representation)، دکھائی دینے کی صورت (rendering)، اور یوزر انٹرفیس (User Interface) پر بھی بہت زیادہ انحصار ہوتا ہے۔ اُردو کے تناظر میں یہ چیلنج دو بڑے حصوں میں تقسیم کیے جاسکتے ہیں: ایک جانب یونیکوڈ (Unicode) مطابقت کا مسئلہ ہے، اور دوسری جانب یوزر انٹرفیس (UI) میں اُردو رسم الخط اور ان پٹ کے مسائل۔ ان دونوں پر اگر تفصیل سے غور نہ کیا جائے تو کوئی بھی ترجمہ پلیٹ فارم یا ویب ایپلی کیشن اُردو صارفین کے لیے مؤثر اور قابلِ اعتماد نہیں بن سکتا۔

اُردو زبان مشینی ترجمہ اور ڈیجیٹل لسانیات کے میدان میں ابھرتی ہوئی زبان ہے۔ گوکہ اُردو کے لیے کئی پلیٹ فارمز جیسے گوگل ٹرانسلیٹ، Bing Translate، اور ChatGPT موجود ہیں، مگر یہ تمام ٹولز ابھی اُردو کی نحوی، صرفی اور لغوی پیچیدگیوں کو مکمل طور پر نہیں سنبھال پاتے۔ اس صورتحال میں ضرورت اس امر کی ہے کہ اُردو کے لیے مخصوص ماڈلز تیار کیے جائیں، جن میں اُردو کی ساختی اور ثقافتی خصوصیات کا بھرپور لحاظ رکھا جائے۔ اس حصے میں ہم ان ممکنہ ماڈلز اور ان کے انضمام کی تجاویز کو تفصیل سے بیان کریں گے۔

اُردو کے لیے الگ سے مشینی ترجمہ ماڈلز تیار کرنا ضروری ہے کیونکہ عمومی ماڈلز جیسے گوگل یا Bing کا انحصار زیادہ تر یورپی زبانوں کے کارپس پر ہے۔ جدید دور میں Neural Machine Translation (NMT) سب سے زیادہ مؤثر ہے۔ اُردو کے لیے ایسے نیورل ماڈلز تیار کیے جاسکتے ہیں جو BERT، GPT یا LLaMA جیسے ڈھانچوں پر مبنی ہوں۔ ان ماڈلز کو بڑے اُردو کارپس پر تربیت دینے سے اُردو کے لیے زیادہ درست اور قدرتی ترجمہ ممکن ہوگا۔ ایک واحد ماڈل اُردو کے تمام مسائل کا حل نہیں دے سکتا۔ اس لیے **ہائبرڈ ماڈلز** جن میں Rule-Based، Statistical اور Neural تکنیکوں کو ملا کر استعمال کیا جائے، زیادہ مؤثر ثابت ہو سکتے ہیں۔(18)

کوئی بھی مشینی ماڈل اسی وقت مؤثر اور قابلِ اعتماد نتائج دے سکتا ہے جب اس کی تربیت بڑے پیمانے پر معیاری اور متنوع ڈیٹا پر کی گئی ہو، اس لیے اُردو زبان کے لیے جامع لسانی وسائل کی تیاری ناگزیر ہے۔ ان وسائل میں سب سے اہم اُردو–انگریزی بائی لنگوئل کارپس ہے جو ہزاروں نہیں بلکہ لاکھوں ہم معنی جملوں کے جوڑوں پر مشتمل ہو تاکہ مشینی ترجمہ معنوی سطح پر درستی حاصل کرسکے۔ اسی طرح محاورات اور کہاوتوں کا ایک منظم ڈیٹا بیس ضروری ہے تاکہ نظام لفظی ترجمے کے بجائے مفہومی ترجمہ پیش کرسکے۔ مزید برآں اُردو کے نحوی اور صرفی ڈھانچے کو بہتر طور پر سمجھنے کے لیے معیاری Treebanks اور تشریح شدہ (Annotated) ڈیٹا کی دستیابی اہم

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

ہے، جبکہ مختلف علمی و فکری میدانوں کی ضروریات کو پورا کرنے کے لیے سائنسی، ادبی، صحافتی اور مذہبی متون پر مشتمل ڈومین مخصوص کارپس کی تیاری بھی تحقیق اور اطلاق دونوں کے لیے بنیادی حیثیت رکھتی ہے۔

اُردو ترجمے میں AI اور NLP کا اطلاق:

اُردو زبان کے لیے مصنوعی ذہانت پر مبنی ماڈلز کی ضرورت کو سمجھنے کے لیے سب سے پہلے اس عالمی تناظر کو مدنظر رکھنا ضروری ہے جس میں AI اور زبان باہم مربوط ہو چکے ہیں۔ اکیسویں صدی میں زبان محض اظہارِ خیال کا ذریعہ نہیں رہی بلکہ ڈیجیٹل معیشت، خودکار نظاموں، علمی ذخائر اور انسانی–مشین تعامل (Human–Computer Interaction) کا بنیادی ستون بن چکی ہے۔ ایسی صورتِ حال میں وہ زبانیں جو مصنوعی ذہانت کے نظاموں میں مؤثر نمائندگی حاصل نہیں کر پاتیں، آہستہ آہستہ ڈیجیٹل حاشیے پر چلی جاتی ہیں۔ اُردو، جو کروڑوں افراد کی مادری اور رابطے کی زبان ہے، بدقسمتی سے اب تک AI ماڈلز میں اس درجے کی توجہ حاصل نہیں کر سکی جو انگریزی، چینی یا یورپی زبانوں کو حاصل ہے۔ یہی وجہ ہے کہ اُردو زبان کے لیے مخصوص AI ماڈلز کی ضروریات پر سنجیدہ اور تحقیقی بنیادوں پر غور ناگزیر ہو چکا ہے۔(19)

مصنوعی ذہانت کے ماڈلز کی کارکردگی بڑی حد تک اس ڈیٹا پر منحصر ہوتی ہے جس پر انہیں تربیت دی جاتی ہے۔ اسی لیے کسی بھی زبان کے لیے مؤثر AI نظام کی تشکیل کے لیے معیاری اور وسیع لسانی ڈیٹا کی دستیابی بنیادی شرط سمجھی جاتی ہے۔ اُردو کے تناظر میں سب سے بڑا مسئلہ یہی ہے کہ اگرچہ اُردو میں ادبی، صحافتی اور مذہبی مواد کی ایک طویل روایت موجود ہے، تاہم یہ مواد ڈیجیٹل صورت میں یا تو منتشر ہے یا پھر مشینی استعمال کے قابل نہیں۔ جدید AI ماڈلز، بالخصوص Deep Learning پر مبنی ماڈلز، ایسے ڈیٹا کا تقاضا کرتے ہیں جو صاف شدہ(cleaned)، تشریح شدہ (annotated) اور معیاری فارمیٹس میں دستیاب ہو۔ اُردو زبان کے لیے AI ماڈلز کی ایک بنیادی ضرورت یہ ہے کہ بڑے پیمانے پر ایسے corpora تیار کیے جائیں جو مختلف اسالیب، موضوعات اور لسانی سطحوں کی نمائندگی کرتے ہوں، تاکہ ماڈلز اُردو زبان کی فطری ساخت، معنوی گہرائی اور اسلوبی تنوع کو سیکھ سکیں۔

اُردو رسم الخط کی پیچیدگی بھی AI ماڈلز کے لیے ایک اہم تقاضا بن جاتی ہے۔ اُردو دائیں سے بائیں لکھی جانے والی زبان ہے اور اس کا رسم الخط عربی–فارسی روایت سے جڑا ہوا ہے، جس میں حروف کی اشکال سیاق و سباق کے مطابق بدلتی ہیں۔ اس کے علاوہ اُردو میں اعراب کا اختیاری استعمال، ہم آواز الفاظ، اور ایک ہی لفظ کے متعدد املا، AI ماڈلز کے لیے چیلنج پیدا کرتے ہیں۔ اس لیے اُردو کے لیے AI ماڈلز کی ایک اہم ضرورت یہ ہے کہ وہ Unicode کے ساتھ مکمل مطابقت رکھتے ہوں اور رسم الخط کی باریکیوں کو درست طور پر پروسیس کرنے کی صلاحیت رکھتے ہوں۔ اگر ماڈل رسم الخط کی سطح پر درست کام نہ کر سکے تو اس کے نتیجے میں ترجمہ، خلاصہ نویسی اور سوال و جواب جیسے کام شدید طور پر متاثر ہوتے ہیں۔

لسانی سطح پر اُردو کی صرفی اور نحوی ساخت بھی مخصوص نوعیت کی ہے جو AI ماڈلز کے لیے خصوصی تقاضے پیدا کرتی ہے۔ اُردو ایک نسبتاً آزاد ترتیب (free word order) رکھنے والی زبان ہے، جس میں جملے کے اجزا کی ترتیب سیاق، زور اور اسلوب کے مطابق بدل سکتی ہے۔ مزید یہ کہ اُردو میں لاحقوں اور سابقوں کے ذریعے الفاظ کی تشکیل، مرکب الفاظ کا استعمال، اور فعل کے ساتھ جنس، عدد اور زمانے کی مطابقت، مشینی تجزیے کو پیچیدہ بنا دیتی ہے۔ اُردو کے لیے AI ماڈلز کی ضرورت اس بات کی متقاضی ہے کہ وہ صرف سطحی شماریاتی پیٹرنز پر انحصار کرنے کے بجائے زبان کی گہری نحوی اور صرفی ساخت کو بھی سمجھ سکیں، جس کے لیے لسانی علم اور computational linguistics کا امتزاج ناگزیر ہے۔

معنوی سطح پر اُردو زبان میں محاورات، استعارات، تشبیہات اور ثقافتی حوالہ جات کی بھرمار ہے۔ یہی عناصر اُردو کو ایک جمالیاتی اور تہذیبی زبان بناتے ہیں، مگر AI ماڈلز کے لیے یہی عناصر سب سے زیادہ مشکل ثابت ہوتے ہیں۔ عمومی AI ماڈلز اکثر لفظی ترجمے یا سطحی معنوی مماثلت پر اکتفا کرتے ہیں، جس کے نتیجے میں اُردو کے محاوراتی اور ثقافتی معانی ضائع ہو جاتے ہیں۔ اس لیے اُردو زبان کے لیے AI ماڈلز کی ایک اہم ضرورت یہ ہے کہ انہیں ایسے ڈیٹا پر تربیت دی جائے جو ثقافتی سیاق و سباق کو بھی سموئے ہوئے ہو، تاکہ ماڈلز محض الفاظ نہیں بلکہ مفاہیم کو سمجھ سکیں۔ اس تناظر میں cultural embeddings اور context-aware models کی ضرورت اُردو کے لیے اور بھی بڑھ جاتی ہے۔

اُردو زبان کے لیے AI ماڈلز کی ایک اور بنیادی ضرورت multilingual اور cross-lingual صلاحیتوں سے جڑی ہوئی ہے۔ چونکہ اُردو تاریخی طور پر فارسی، عربی اور ہندی کے ساتھ گہرے روابط رکھتی ہے، اور جدید دور میں انگریزی سے بھی شدید طور پر متاثر ہے، اس لیے اُردو کے AI ماڈلز کو محض monolingual دائرے تک محدود رکھنا علمی طور پر نقصان دہ ہوگا۔ ایسے ماڈلز کی ضرورت ہے جو اُردو کو دیگر زبانوں کے ساتھ تقابلی اور اشتراکی انداز میں سیکھ سکیں،

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

تاکہ ترجمہ، code-switching اور mixed-language texts کو بہتر طور پر سمجھا جا سکے۔ اس مقصد کے لیے multilingual pre-trained models کو اُردو کے مخصوص ڈیٹا کے ساتھ fine-tune کرنا ایک اہم تقاضا بن جاتا ہے۔

اُردو زبان کے لیے AI ماڈلز کی تیاری میں ایک اہم عملی ضرورت تکنیکی وسائل اور انفراسٹرکچر سے بھی جڑی ہوئی ہے۔ بڑے زبان ماڈلز کی تربیت کے لیے اعلیٰ درجے کی کمپیوٹنگ طاقت، GPUs اور cloud-based systems درکار ہوتے ہیں، جو اکثر اُردو تحقیق کے اداروں میں دستیاب نہیں ہوتے۔ اس خلا کو پُر کرنے کے لیے یا تو قومی سطح پر مشترکہ تحقیقی پلیٹ فارمز قائم کرنے کی ضرورت ہے یا پھر ایسے ہلکے (lightweight) اور کم وسائل پر چلنے والے ماڈلز تیار کیے جائیں جو اُردو کے لیے قابلِ عمل ہوں۔ یہ ضرورت خاص طور پر اس وقت اہم ہو جاتی ہے جب اُردو بولنے والے صارفین کی بڑی تعداد ایسے خطوں میں رہتی ہے جہاں وسائل محدود ہیں۔

ایک اور اہم ضرورت اُردو زبان کے لیے AI ماڈلز میں اخلاقی اور سماجی پہلوؤں کے ادراک سے متعلق ہے۔ چونکہ AI ماڈلز اپنے تربیتی ڈیٹا سے تعصبات سیکھ لیتے ہیں، اس لیے اُردو کے لیے ماڈلز تیار کرتے وقت اس بات کا خیال رکھنا ضروری ہے کہ ڈیٹا میں لسانی، مذہبی، صنفی یا علاقائی تعصبات نہ ہوں۔ اُردو ایک متنوع سماجی اور تہذیبی پس منظر رکھنے والی زبان ہے، اور اگر AI ماڈلز اس تنوع کی درست نمائندگی نہ کر سکیں تو وہ سماجی سطح پر غلط فہمیوں اور عدم توازن کو فروغ دے سکتے ہیں۔ لہٰذا اُردو زبان کے لیے AI ماڈلز کی ایک اہم ضرورت responsible AI اور ethical AI کے اصولوں کی پاسداری ہے۔

تعلیمی اور تحقیقی سطح پر بھی اُردو کے لیے AI ماڈلز کی ضرورت ایک منظم حکمتِ عملی کا تقاضا کرتی ہے۔ اُردو زبان کو محض اطلاقی سطح پر نہیں بلکہ نظریاتی سطح پر بھی AI تحقیق کا حصہ بنانا ہوگا۔ اس مقصد کے لیے اُردو میں computational linguistics، NLP اور AI سے متعلق نصابی مواد، تحقیقی مقالے اور تربیتی پروگرامز کی ضرورت ہے، تاکہ ایک نئی نسل ایسے محققین کی تیار ہو جو اُردو اور AI دونوں میں مہارت رکھتے ہوں۔ بغیر انسانی سرمائے کے، کسی بھی زبان کے لیے پائیدار AI ماڈلز تیار کرنا ممکن نہیں۔(20)

اُردو زبان کے لیے مؤثر AI ماڈلز کی تیاری صرف تکنیکی مسئلہ نہیں بلکہ اس میں لسانی تحقیق، ثقافتی تناظر، تعلیمی منصوبہ بندی اور سماجی ضروریات سب شامل ہوتی ہیں۔ اُردو کے لیے مؤثر AI ماڈلز کی تیاری اسی وقت ممکن ہے جب ڈیٹا، لسانی علم، جدید الگورتھمز اور اخلاقی اصولوں کو ایک جامع فریم ورک میں یکجا کیا جائے۔ اگر اُردو زبان کو مستقبل کی ڈیجیٹل دنیا میں ایک فعال اور باوقار مقام دلانا مقصود ہے تو اُردو کے لیے مخصوص اور معیاری AI ماڈلز کی تیاری اب ایک انتخاب نہیں بلکہ ایک ناگزیر ضرورت بن چکی ہے۔

اُردو کے ڈیجیٹل ترجمے میں جہاں عالمی سطح پر تیار ہونے والے خودکار نظاموں سے استفادہ کیا جا رہا ہے، وہیں یہ حقیقت بھی سامنے آ چکی ہے کہ اُردو زبان اپنی ساخت، رسم الخط اور تہذیبی پس منظر کے باعث ایسے مصنوعی ذہانت کے ماڈلز کی متقاضی ہے جو اس کی داخلی خصوصیات کو پیشِ نظر رکھ کر تشکیل دیے جائیں۔ چنانچہ اُردو زبان کے لیے AI ماڈلز کی ضروریات کا تعین دراصل اُردو ڈیجیٹل ترجمہ نگاری کے مستقبل کا تعین ہے۔

اُردو زبان لسانی اعتبار سے ایک ہمہ جہت زبان ہے جس میں عربی، فارسی، سنسکرت اور مقامی بولیوں کے اثرات شامل ہیں۔ یہی امتزاج اس کے حسن کا سبب ہے، مگر مشینی ترجمے کے لیے یہی عنصر سب سے بڑا چیلنج بھی بن جاتا ہے۔ مصنوعی ذہانت پر مبنی کسی بھی ماڈل کے لیے سب سے پہلی اور بنیادی ضرورت یہ ہے کہ وہ اُردو زبان کے صرفی، نحوی اور معنوی نظام کو بطور ایک مربوط اکائی سمجھے، نہ کہ محض الفاظ کے مجموعے کے طور پر۔ موجودہ عالمی ماڈلز اکثر اُردو کو ایک ثانوی یا کم وسائل والی زبان کے طور پر دیکھتے ہیں، جس کے نتیجے میں ترجمہ سطحی اور سیاق سے کٹا ہوا نظر آتا ہے۔ اس صورتِ حال کا تقاضا ہے کہ اُردو کے لیے مخصوص AI ماڈلز تیار کیے جائیں یا کم از کم موجودہ ماڈلز کو اُردو لسانی ڈھانچے کے مطابق تربیت دی جائے۔

اُردو کے لیے AI ماڈلز کی ایک اہم ضرورت معیاری اور وسیع متنی ذخائر کی دستیابی ہے۔ اُردو ڈیجیٹل ترجمہ نگاری کی سب سے بڑی کمزوری یہ ہے کہ اُردو کے مستند، صاف شدہ اور متنوع کارپس کی شدید کمی ہے۔ زیادہ تر دستیاب مواد یا تو اخباری زبان تک محدود ہے یا مذہبی و ادبی متون پر مشتمل ہے، جب کہ جدید سائنسی، فنی اور سماجی علوم کی زبان بہت کم نمائندگی رکھتی ہے۔ مصنوعی ذہانت کے ماڈلز چونکہ مشاہدے اور تکرار سے سیکھتے ہیں، اس لیے اگر ان کے سامنے اُردو کے متنوع اسالیب اور موضوعات نہ ہوں تو وہ معیاری ترجمہ فراہم نہیں کر سکتے۔ چنانچہ اُردو کے لیے AI ماڈلز کی بنیادی شرط یہ ہے کہ انہیں ایسے متنی ذخائر فراہم کیے جائیں جن میں ادب، صحافت، تدریس، قانون، سائنس اور روزمرہ زبان سب شامل ہوں۔

ISSN E: 2709-8273
ISSN P:2709-8265
JALT
JOURNAL OF APPLIED LINGUISTICS AND TESOL

اُردو رسم الخط خود ایک الگ نوعیت کا تقاضا رکھتا ہے۔ دائیں سے بائیں لکھی جانے والی زبان، حروف کی مختلف اشکال، اعراب کی عدم موجودگی اور املا کے متعدد طریقے وہ مسائل ہیں جن سے اُردو کا ہر ڈیجیٹل نظام دوچار ہوتا ہے۔ مصنوعی ذہانت کے ماڈلز کے لیے یہ ضروری ہے کہ وہ اُردو رسم الخط کی ان پیچیدگیوں کو بطور مسئلہ نہیں بلکہ بطور حقیقت قبول کریں۔ اس ضمن میں ایسے ماڈلز درکار ہیں جو اُردو حروف کی شناخت، لفظ کی سرحد بندی اور جملے کی ساخت کو درست طور پر سمجھ سکیں۔ اگر ماڈل لفظ ہی کو درست طور پر پہچاننے میں ناکام رہے تو اس سے آگے معنوی فہم کا سوال ہی پیدا نہیں ہوتا۔

اُردو زبان میں الفاظ کے کئی معانی ہونا ایک عام بات ہے۔ ایک ہی لفظ مختلف سیاق و سباق میں مختلف مفاہیم رکھتا ہے۔ انسانی مترجم سیاق، لہجے اور موقع کی بنیاد پر معنی متعین کرتا ہے، مگر مصنوعی ذہانت کے لیے یہ مرحلہ سب سے زیادہ مشکل ہے۔ اُردو کے لیے AI ماڈلز کی ایک بڑی ضرورت یہ ہے کہ وہ سیاقی فہم پر مبنی ہوں، یعنی لفظ کو تنہا نہیں بلکہ پورے جملے اور متن کے تناظر میں سمجھ سکیں۔ اس کے بغیر اُردو ڈیجیٹل ترجمہ نگاری محض لفظی ترجمے تک محدود رہے گی، جو ترجمہ نگاری کی روح کے منافی ہے۔

ثقافتی اور محاوراتی اظہار اُردو زبان کا امتیازی وصف ہے۔ اُردو میں ایسے محاورے اور تراکیب کثرت سے پائی جاتی ہیں جن کا لفظی ترجمہ کسی دوسری زبان میں ممکن نہیں۔ مصنوعی ذہانت کے ماڈلز اگر اُردو کے لیے مؤثر ثابت ہونا چاہتے ہیں تو ان میں ثقافتی شعور کا عنصر شامل کرنا ناگزیر ہے۔ اس کے لیے اُردو محاورات، کہاوتوں اور استعارات پر مشتمل خصوصی ذخائر کی ضرورت ہے، تاکہ ماڈل محض لفظ نہیں بلکہ مفہوم منتقل کر سکے۔ اُردو ڈیجیٹل ترجمہ نگاری میں معیار کا دار و مدار اسی نکتے پر ہے کہ ترجمہ پڑھنے والے کو اجنبی نہ لگے بلکہ فطری محسوس ہو۔

اُردو کے لیے AI ماڈلز کی ایک اور بنیادی ضرورت لسانی تشریحی وسائل ہیں۔ اُردو زبان کے ایسے ذخائر جن میں الفاظ کی قسم، نحوی کردار اور جملے کی ساخت واضح ہو، نہایت محدود ہیں۔ جب تک مصنوعی ذہانت کو یہ نہ بتایا جائے کہ کون سا لفظ اسم ہے، کون سا فعل اور کون سا حرف، وہ درست ترجمے میں مسلسل غلطیاں کرتا رہے گا۔ اس لیے اُردو کے لیے AI ماڈلز کی تیاری کے ساتھ ساتھ اُردو زبان کے تشریحی کارپس اور لغوی وسائل کی تیاری بھی ناگزیر ہے۔ یہ عمل محض کمپیوٹر سائنس کا نہیں بلکہ لسانیات اور اُردو تحقیق کا مشترکہ منصوبہ ہونا چاہیے۔

اُردو چونکہ ایک زندہ اور متحرک زبان ہے، اس لیے اس میں نئی اصطلاحات اور نئے اسالیب مسلسل شامل ہو رہے ہیں۔ مصنوعی ذہانت کے ماڈلز کے لیے ضروری ہے کہ وہ جامد نہ ہوں بلکہ وقت کے ساتھ خود کو تازہ کر سکیں۔ اُردو ڈیجیٹل ترجمہ نگاری کے تناظر میں ایسے ماڈلز کی ضرورت ہے جو صارف کے تعامل سے سیکھ سکیں اور اپنی کارکردگی بہتر بناتے جائیں۔ یہ خصوصیت اُردو کے لیے خاص طور پر اہم ہے، کیونکہ اُردو کے صارفین مختلف خطوں، لہجوں اور ثقافتی پس منظر سے تعلق رکھتے ہیں۔

اُردو زبان کے لیے AI ماڈلز کی ضروریات محض تکنیکی نہیں بلکہ فکری اور تہذیبی بھی ہیں۔ اُردو کو اگر واقعی ایک عالمی ڈیجیٹل زبان کے طور پر متعارف کرانا ہے تو مصنوعی ذہانت کے نظاموں کو اُردو کے مزاج کے مطابق ڈھالنا ہوگا، نہ کہ اُردو کو مشینی تقاضوں کے تحت محدود کرنا ہوگا۔ یہی نقطۂ نظر اُردو ڈیجیٹل ترجمہ نگاری کو محض ایک سہولت نہیں بلکہ ایک معتبر علمی ذریعہ بنا سکتا ہے۔

اُردو ڈیجیٹل ترجمہ نگاری کی موجودہ صورتِ حال کو اگر مصنوعی ذہانت کے جدید ماڈلز کے تناظر میں دیکھا جائے تو یہ بات واضح ہو جاتی ہے کہ حالیہ برسوں میں زبان سے متعلق ٹیکنالوجی نے ایک نئی کروٹ لی ہے۔ یہ بات سامنے آچکی ہے کہ روایتی قاعدہ پر مبنی اور شماریاتی مشینی ترجمہ کے طریقے اُردو جیسی زبان کے لسانی تنوع، صرفی ساخت، نحوی پیچیدگی اور ثقافتی سیاق کو مکمل طور پر نہیں سمو سکتے۔ یہی وہ مقام ہے جہاں جدید گہرے عصبی نیٹ ورک پر مبنی ماڈلز، خصوصاً ChatGPT، BERT اور LLaMA، اُردو ڈیجیٹل ترجمہ نگاری میں ایک نئے امکان کے طور پر سامنے آتے ہیں۔

BERT کو بنیادی طور پر ایک ایسا لسانی ماڈل سمجھا جاتا ہے جو الفاظ کو صرف انفرادی اکائی کے طور پر نہیں بلکہ جملے اور سیاق و سباق کے اندر معنی کے ساتھ سمجھنے کی صلاحیت رکھتا ہے۔ اُردو زبان میں، جہاں ایک ہی لفظ مختلف سیاق میں مختلف مفاہیم رکھتا ہے، یہ خصوصیت نہایت اہم ہو جاتی ہے۔ مثلاً اُردو میں لفظ "حال" کبھی کیفیت کے معنوں میں آتا ہے اور کبھی زمانے کے مفہوم میں۔ روایتی ترجمہ نظام اکثر اس فرق کو نظر انداز کر دیتے ہیں، جب کہ BERT جیسے ماڈلز سیاقی اشاروں کی بنیاد پر زیادہ درست مفہوم اخذ کرنے کی صلاحیت رکھتے ہیں۔ اُردو ڈیجیٹل ترجمہ میں BERT کا اطلاق خاص طور پر جملوں کی درجہ بندی، معنوی تجزیے اور جزوی ترجمہ کے معاون مراحل میں مفید ثابت ہو سکتا ہے۔

تاہم یہ بات بھی پیشِ نظر رہنی چاہیے کہ BERT بذاتِ خود مکمل مشینی ترجمہ کے لیے تیار نہیں کیا گیا بلکہ یہ زیادہ تر فہمِ زبان، معنوی تعلقات اور متن کے تجزیے کے لیے استعمال ہوتا ہے۔ اُردو ترجمہ نگاری میں اس کا اصل فائدہ تب سامنے آتا ہے جب اسے دیگر ترجمہ ماڈلز کے ساتھ مربوط کیا جائے۔ مثال کے طور پر، اُردو اور انگریزی کے متوازی متون پر تربیت یافتہ کسی ترجمہ نظام میں BERT کو بطور معاون ماڈیول شامل کر کے معنوی صحت اور جملوں کی ساخت کو بہتر بنایا جا سکتا ہے۔

ChatGPT کی صورت میں ہمیں ایک ایسا ماڈل ملتا ہے جو محض لفظی ترجمہ تک محدود نہیں رہتا بلکہ متن کے مجموعی مفہوم، اسلوب اور لہجے کو بھی مدِنظر رکھتا ہے۔ اُردو ڈیجیٹل ترجمہ نگاری میں یہ پہلو نہایت اہم ہے، کیونکہ اُردو محض اطلاعاتی زبان نہیں بلکہ جذبات، تہذیب اور ادبی نزاکتوں کی حامل زبان ہے۔ ChatGPT جیسے ماڈلز کا اطلاق اُردو ترجمہ میں خاص طور پر ادبی متون، صحافتی تحریروں اور عمومی فہم کے مواد میں نسبتاً بہتر نتائج دے سکتا ہے۔ یہ ماڈل اُردو سے دوسری زبانوں اور دوسری زبانوں سے اُردو میں ترجمہ کرتے ہوئے محض لفظ بہ لفظ ترجمہ نہیں کرتا بلکہ جملے کی مجموعی معنوی ساخت کو برقرار رکھنے کی کوشش کرتا ہے۔

اُردو کے حوالے سے ChatGPT کے اطلاق میں چند بنیادی مسائل بھی سامنے آتے ہیں۔ ایک اہم مسئلہ تربیتی مواد کا ہے۔ چونکہ اُردو عالمی سطح پر نسبتاً کم ڈیجیٹل مواد رکھتی ہے، اس لیے ChatGPT جیسے بڑے ماڈلز میں اُردو کا حصہ محدود رہ جاتا ہے۔ نتیجتاً بعض اوقات ترجمہ سطحی، غیر فصیح یا ثقافتی لحاظ سے غیر موزوں ہو جاتا ہے۔ مزید یہ کہ اُردو رسم الخط، مرکب الفاظ اور محاوراتی اظہار میں جو باریکیاں پائی جاتی ہیں، وہ ہر بار درست طور پر منتقل نہیں ہو پاتیں۔ اس کے باوجود، عمومی صارفین کے لیے فوری اور قابلِ فہم ترجمہ فراہم کرنے میں ChatGPT ایک مؤثر ذریعہ بن سکتا ہے، خاص طور پر وہاں جہاں رفتار اور سہولت کو اولیت حاصل ہو۔

LLaMA جیسے ماڈلز کی اہمیت اُردو ڈیجیٹل ترجمہ نگاری میں ایک اور زاویے سے سامنے آتی ہے۔ یہ ماڈلز نسبتاً کھلے اور قابلِ تخصیص ہیں، جس کا مطلب یہ ہے کہ انہیں خاص زبان یا خاص لسانی ضروریات کے مطابق ڈھالا جا سکتا ہے۔ اُردو کے لیے یہ پہلو نہایت اہم ہے، کیونکہ اُردو ترجمہ کے مسائل کا حل عمومی عالمی ماڈلز میں مکمل طور پر دستیاب نہیں۔ LLaMA جیسے ماڈلز کو اُردو متون، لغات، اور متوازی ترجمہ مواد پر دوبارہ تربیت دے کر ایک ایسا نظام تشکیل دیا جا سکتا ہے جو اُردو کی صرفی، نحوی اور اسلوبی خصوصیات کو بہتر طور پر سمجھ سکے۔

اُردو ڈیجیٹل ترجمہ کے تناظر میں LLaMA کا اطلاق خاص طور پر تحقیقی اور ادارہ جاتی سطح پر زیادہ مفید ثابت ہو سکتا ہے۔ جامعات، تحقیقی ادارے اور لسانی مراکز اگر اُردو کے لیے مخصوص ڈیٹا تیار کر کے ان ماڈلز میں شامل کریں تو ایک ایسا ترجمہ نظام وجود میں آ سکتا ہے جو مقامی ضروریات سے ہم آہنگ ہو۔ اس کے ذریعے نہ صرف عمومی ترجمہ بلکہ قانونی، سائنسی اور تکنیکی متون کا نسبتاً معیاری اُردو ترجمہ بھی ممکن بنایا جا سکتا ہے۔(21)

ان تینوں ماڈلز کا تقابلی جائزہ یہ بتاتا ہے کہ اُردو ڈیجیٹل ترجمہ نگاری کے لیے کوئی ایک ماڈل بذاتِ خود مکمل حل فراہم نہیں کرتا۔ BERT اُردو متن کے فہم اور معنوی تجزیے میں معاون ہے، ChatGPT عمومی اور سیاقی ترجمہ میں سہولت فراہم کرتا ہے، جب کہ LLaMA جیسے ماڈلز اُردو کے لیے مخصوص اور حسبِ ضرورت نظام تیار کرنے کا امکان دیتے ہیں۔ اُردو ترجمہ نگاری کے مستقبل میں زیادہ مؤثر حکمتِ عملی یہی ہو سکتی ہے کہ ان ماڈلز کو ایک دوسرے کے ساتھ مربوط کر کے استعمال کیا جائے۔ مثال کے طور پر، کسی مجوزہ اُردو ترجمہ ویب ایپلی کیشن میں BERT کو متن کی پیشگی درجہ بندی اور معنوی جانچ کے لیے، ChatGPT کو عمومی ترجمہ اور صارف کے ساتھ تعامل کے لیے، اور LLaMA کو مخصوص شعبہ جاتی ترجمہ کے لیے استعمال کیا جا سکتا ہے۔

اس تناظر میں یہ بھی ضروری ہے کہ اُردو ترجمہ نگاری میں ان ماڈلز کے اطلاق کو محض تکنیکی کامیابی کے طور پر نہ دیکھا جائے بلکہ اسے ایک تہذیبی اور علمی ذمہ داری کے طور پر بھی سمجھا جائے۔ اگر اُردو کے لیے مصنوعی ذہانت کے یہ ماڈلز درست سمت میں ترقی کرتے ہیں تو نہ صرف اُردو زبان عالمی ڈیجیٹل منظرنامے میں مضبوط مقام حاصل کر سکتی ہے بلکہ علمی و ادبی ترجمے کے ذریعے اُردو قارئین کو جدید علم تک براہِ راست رسائی بھی ممکن ہو سکتی ہے۔

اُردو NLP کے لیے مضبوط، معیاری اور منظم لسانی وسائل کے بغیر نہ تو مؤثر مشینی ترجمہ ممکن ہے اور نہ ہی مصنوعی ذہانت کے جدید ماڈلز سے مطلوبہ نتائج حاصل کیے جا سکتے ہیں۔ لسانی وسائل سے مراد وہ منظم متنی ذخائر ہیں جو زبان کے مختلف پہلوؤں کی نمائندگی کرتے ہیں۔ اُردو کے تناظر میں یہ وسائل خاص طور پر اس لیے اہم ہو جاتے ہیں کہ اُردو کو عالمی سطح پر ایک کم وسائل رکھنے والی زبان سمجھا جاتا ہے۔ اس کمی کا براہِ راست اثر اُردو ڈیجیٹل ترجمہ نگاری پر پڑتا ہے، کیونکہ مصنوعی ذہانت کے نظام زبان کو انسان کی طرح فطری طور پر نہیں سیکھتے بلکہ انہیں مثالوں، نمونوں اور تشریح شدہ مواد کی ضرورت ہوتی ہے۔ اگر یہ مواد ناکافی یا غیر معیاری ہو تو ترجمہ بھی غیر مستند اور غیر فصیح ہو گا۔

Corpus یعنی متنی ذخیرہ اُردو NLP کی بنیاد کی حیثیت رکھتا ہے۔ یہ ذخیرہ محض متن کا مجموعہ نہیں بلکہ زبان کے حقیقی استعمال کی عکاسی کرتا ہے۔ اُردو کے لیے معیاری Corpus کی اہمیت اس لیے دوچند ہو جاتی ہے کہ اُردو زبان مختلف اسالیب، لہجوں اور موضوعات میں استعمال ہوتی ہے۔ ادبی زبان، صحافتی زبان، تدریسی

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

اسلوب اور روزمرہ گفتگو—یہ سب اُردو کے مختلف چہرے ہیں۔ اگر اُردو Corpus ان تمام جہات کا احاطہ نہ کرے تو مصنوعی ذہانت کے ماڈلز اُردو کو ایک محدود اور یک رخی زبان کے طور پر سیکھیں گے، جس کا نتیجہ ناقص ترجمے کی صورت میں نکلے گا۔

اُردو ڈیجیٹل ترجمہ نگاری کے لیے متوازی Corpus کی اہمیت بھی غیر معمولی ہے۔ متوازی Corpus وہ ذخیرہ ہوتا ہے جس میں ایک ہی متن دو یا زیادہ زبانوں میں دستیاب ہو۔ مشینی ترجمہ کے نظام انہی متوازی متون سے سیکھتے ہیں کہ کس زبان کا کون سا جملہ دوسری زبان میں کس طرح منتقل ہوتا ہے۔ اُردو کے معاملے میں مسئلہ یہ ہے کہ معیاری اُردو۔ انگریزی یا اُردو۔ دیگر زبانوں کے متوازی ذخائر نہایت محدود ہیں۔ اس کمی کے باعث اُردو ترجمہ کے خودکار نظام اکثر انگریزی مرکزیت کا شکار ہو جاتے ہیں اور اُردو کے فطری اسلوب کو برقرار نہیں رکھ پاتے۔

Corpus کے ساتھ ساتھ Treebanks اُردو NLP کے لیے ایک اور بنیادی ضرورت ہیں۔ Treebank دراصل ایسا تشریح شدہ ذخیرہ ہوتا ہے جس میں جملوں کی نحوی ساخت، الفاظ کے باہمی تعلق اور گرامری کردار کو واضح کیا جاتا ہے۔ اُردو زبان چونکہ نحوی اعتبار سے لچکدار ہے اور اس میں جملے کی ترتیب بدلنے سے معنی ہمیشہ متاثر نہیں ہوتے، اس لیے اُردو Treebanks کی اہمیت اور بھی بڑھ جاتی ہے۔ اگر مصنوعی ذہانت کے ماڈلز کو یہ نہ بتایا جائے کہ جملے میں فاعل، مفعول اور فعل کا آپس میں کیا تعلق ہے تو وہ درست ترجمہ کرنے سے قاصر رہیں گے۔

اُردو Treebanks کی کمی کا اثر براہِ راست اُردو مشینی ترجمہ میں دیکھا جا سکتا ہے۔ اکثر خودکار نظام اُردو جملے کی ساخت کو درست طور پر نہیں سمجھ پاتے، جس کے نتیجے میں ترجمہ غیر فطری یا معنوی اعتبار سے غلط ہو جاتا ہے۔ اس کے برعکس، اگر اُردو کے لیے معیاری Treebanks دستیاب ہوں تو مصنوعی ذہانت کے ماڈلز جملے کی گہرائی تک جا کر اس کے مفہوم کو بہتر انداز میں سمجھ سکتے ہیں۔ یہ خصوصیت خاص طور پر علمی اور فنی متون کے ترجمے میں نہایت اہم ہے، جہاں ایک چھوٹی سی نحوی غلطی پورے مفہوم کو بدل سکتی ہے۔

اُردو NLP کے لیے لسانی وسائل کی اہمیت صرف مشینی ترجمہ تک محدود نہیں بلکہ یہ پورے ڈیجیٹل لسانی نظام کی بنیاد ہیں۔ متن کی درجہ بندی، خلاصہ سازی، سوال و جواب کے نظام اور خودکار تدوین جیسے تمام جدید اطلاقات انہی وسائل پر انحصار کرتے ہیں۔ اگر اُردو کے لیے یہ وسائل مضبوط ہوں تو اُردو زبان ڈیجیٹل دنیا میں محض صارف کی حیثیت سے نہیں بلکہ تخلیقی اور علمی زبان کے طور پر سامنے آ سکتی ہے۔ یہی نکتہ اُردو ڈیجیٹل ترجمہ نگاری کو ایک وسیع تر علمی منصوبے سے جوڑتا ہے۔

یہاں یہ بات بھی قابلِ غور ہے کہ اُردو کے لسانی وسائل کی تیاری محض تکنیکی ماہرین کا کام نہیں۔ اس میں اُردو کے محققین، لسانیات کے ماہرین اور ترجمہ نگاروں کا کردار بھی بنیادی ہے۔ Corpus اور Treebanks کی تیاری میں اگر لسانی باریکیوں کو نظر انداز کیا جائے تو تیار ہونے والا مواد مصنوعی ذہانت کے لیے نقصان دہ بھی ثابت ہو سکتا ہے۔ اس لیے اُردو NLP کے لیے لسانی وسائل کی تیاری کو ایک بین الشعبہ جاتی منصوبے کے طور پر دیکھنا چاہیے، جہاں زبان اور ٹیکنالوجی ایک دوسرے کی تکمیل کریں۔(22)

اُردو ڈیجیٹل ترجمہ نگاری کے مستقبل کے حوالے سے یہ کہا جا سکتا ہے کہ لسانی وسائل میں سرمایہ کاری دراصل زبان کے مستقبل میں سرمایہ کاری ہے۔ اگر اُردو کے لیے معیاری Corpus اور Treebanks تیار کر لیے جائیں تو نہ صرف موجودہ ترجمہ نظام بہتر ہو سکتے ہیں بلکہ آئندہ آنے والی مصنوعی ذہانت کی نسلیں بھی اُردو کو زیادہ فطری اور مؤثر انداز میں سمجھ سکیں گی۔

اُردو NLP کے لیے لسانی وسائل محض ایک تکنیکی ضرورت نہیں بلکہ ایک تہذیبی ذمہ داری بھی ہیں۔ اُردو زبان کا ڈیجیٹل مستقبل اسی وقت محفوظ اور روشن ہو سکتا ہے جب اس کے لسانی سرمائے کو منظم، محفوظ اور قابلِ استعمال بنایا جائے۔ یہی شعور اُردو ڈیجیٹل ترجمہ نگاری کو ایک مضبوط علمی بنیاد فراہم کرتا ہے اور اسے عالمی ڈیجیٹل منظرنامے میں باوقار مقام دلانے کی راہ ہموار کرتا ہے۔

مجوزہ اُردو ڈیجیٹل ترجمہ ویب ایپلی کیشن فریم ورک:

اُردو ڈیجیٹل ترجمہ نگاری کے میدان میں ایک جامع، مربوط اور عملی تحقیقی پلیٹ فارم کی تشکیل اب ایک موجودہ اور ناگزیر ضرورت کے طور پر سامنے آئی ہے۔ زیرِ تجویز اُردو ترجمہ ویب ایپلی کیشن کا مقصد صرف ترجمے کو لغوی یا سطحی تبدیلی تک محدود رکھنا نہیں بلکہ اسے معنوی صحت، اسلوبی مطابقت اور تہذیبی مناسبت کے ساتھ ایک مکمل فکری عمل کے طور پر نافذ کرنا ہے، تاکہ اُردو ترجمہ محض لفظوں کا تبادلہ نہ رہے بلکہ مفہوم، سیاق و سباق اور اسلوب کی یکساں منتقلی کا ذریعہ بن سکے۔ اس ایپلی کیشن میں اُردو زبان کو مرکزیت حاصل ہے، اور یہ نہ صرف اُردو۔ انگریزی اور انگریزی۔ اُردو ترجمے کی صلاحیت فراہم کرتی ہے بلکہ مستقبل میں دیگر زبانوں

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

کے ساتھ اُردو کے ترجمے کی توسیع کی گنجائش بھی رکھتی ہے۔ اس نظام کی داخلی ساخت اُردو زبان کے نحوی، صرفی اور اسلوبی تقاضوں کے مطابق ترتیب دی گئی ہے تاکہ ترجمہ نہ صرف درست بلکہ بامعنی، فصیح اور سیاق کے لحاظ سے موزوں ہو۔ نظریاتی سطح پر یہ ویب اپیلی کیشن ایک ایسے ذہین مترجم کے تصور کو عملی شکل دیتی ہے جو اُردو کے قواعد، محاورات اور اسلوبی روایت سے بخوبی واقف ہو اور ادبی، علمی اور صحافتی متون کے مابین فرق کو سمجھنے کی صلاحیت رکھتا ہو۔

اس اپیلی کیشن کی لسانی بنیاد اُردو کارپس پر استوار ہے، جو متون کے ذخیرے کے بجائے زبان کے زندہ، متنوع اور سیاقی استعمال کی نمائندگی کرتا ہے۔ اس کارپس میں ادبی نثر اور شاعری، صحافتی و اخباری زبان، درسی و علمی متون اور روزمرہ مکالماتی زبان کے منتخب نمونے شامل ہیں، تاکہ نظام یہ ادراک حاصل کر سکے کہ اُردو مختلف مواقع اور متون میں کس طرح استعمال ہوتی ہے۔ اس کے ساتھ متوازی کارپس بھی شامل کیا گیا ہے تاکہ اُردو اور دیگر زبانوں کے درمیان ترجمے کے عملی نمونے حاصل کیے جا سکیں اور یہ مواد مصنوعی ذہانت کے ماڈلز کی تربیت اور جانچ کے لیے استعمال ہو۔ ویب اپیلی کیشن کا مرکزی فنی جزو مصنوعی ذہانت پر مبنی API ہے، جو ترجمے کے بنیادی عمل کو انجام دیتی ہے اور اسے اُردو کارپس، لغوی وسائل اور اسلوبی اصولوں کے ساتھ ہم آہنگ رکھا گیا ہے۔ مصنوعی ذہانت کا کردار محض جملوں کی تبدیلی تک محدود نہیں بلکہ متن کے مفہوم کو سمجھنا، ممکنہ تراجم کی تشکیل اور اُردو کے فطری اسلوب کے مطابق مناسب انتخاب کرنا ہے، جس سے AI کی تکنیکی صلاحیت اور اُردو لسانیات کی فکری بصیرت کا امتزاج ممکن ہوتا ہے۔

ترجمے کا عمل اس اپیلی کیشن میں مرحلہ وار اور منظم ہوتا ہے۔ ابتدائی مرحلے میں صارف کا فراہم کردہ متن اُردو رسم الخط، املا اور جملہ بندی کے اعتبار سے جانچا جاتا ہے تاکہ بنیادی لسانی شناخت حاصل ہو۔ اس کے بعد متن کو اُردو کارپس اور لغوی نظام کے ساتھ مربوط کیا جاتا ہے، جس سے ممکنہ معانی، مترادفات اور اسلوبی پہلو سامنے آتے ہیں۔ اگلے مرحلے میں مصنوعی ذہانت کی API ترجمے کی تجاویز پیش کرتی ہے، تاہم یہ تجاویز اُردو کے لسانی اور اسلوبی دائرے میں محدود رہتی ہیں۔ آخری مرحلے میں ترجمے کو فصاحت، روانی اور معنوی صحت کے معیار پر پرکھا جاتا ہے اور صارف کو ایک بامعنی، فطری اور معیاری اُردو ترجمہ فراہم کیا جاتا ہے۔ اس نظام میں ترجمہ بامحاورہ، رواں اور سیاق و سباق سے ہم آہنگ ہوتا ہے اور اس کے لیے اُردو محاورات، اسلوبی نمونے اور مختلف النوع متون کے لیے الگ ترجیحی اصول شامل کیے گئے ہیں، جس سے ادبی متن کا ترجمہ ادبی ذوق کے مطابق اور علمی متن کا ترجمہ علمی سنجیدگی کے ساتھ ممکن ہو پاتا ہے۔ صارف انٹرفیس سادہ، واضح اور مکمل اُردو زبان پر مبنی ہے، جس میں دائیں سے بائیں تحریر، مناسب خط اور اُردو اصطلاحات شامل ہیں، اور صارف کو متبادل تراجم دیکھنے، رائے یا اصلاح درج کرنے اور ترجمے کی نوعیت منتخب کرنے کی سہولت دی گئی ہے، جس سے ترجمہ ایک تعاملاتی اور اشتراکی عمل کے طور پر انجام پاتا ہے۔ تحقیقی اعتبار سے یہ ویب اپیلی کیشن محض عملی ٹول نہیں بلکہ ایک تجرباتی پلیٹ فارم کے طور پر صارفین کے استعمال، اصلاحات اور فیڈبیک کی بنیاد پر ترجمے کے معیار اور اُردو NLP پر مزید تحقیق کے دروازے کھولتی ہے۔

فنکشنل ماڈیول کے تناظر میں، اپیلی کیشن کو ایسے مربوط مگر خود مختار حصوں میں تقسیم کیا گیا ہے جو ہر ایک مخصوص لسانی یا تکنیکی ذمہ داری سر انجام دیتے ہیں۔ صارف تعامل اور ان پٹ ماڈیول صارف سے متن وصول کرتا ہے اور اسے اُردو یونیکوڈ کے مطابق معیاری شکل میں نظام کے اندر منتقل کرتا ہے، جس سے ترجمے کا عمل صاف اور قابلِ اعتماد بنیاد پر شروع ہوتا ہے۔ لسانی پیشگی تجزیہ ماڈیول اُردو متن کی صرفی، نحوی اور ساختی خصوصیات کی شناخت کرتا ہے اور جملوں، الفاظ اور گرامری اکائیوں کی نشاندہی کرتا ہے تاکہ ترجمہ بامعنی اور فصیح ہو۔ کارپس اور لغوی ربط ماڈیول متن کو اُردو کارپس اور لغوی وسائل سے مربوط کرتا ہے، جس سے الفاظ، تراکیب، محاورات اور سیاقی استعمال کی معلومات ترجمے میں شامل ہوتی ہیں۔ مصنوعی ذہانت ترجمہ ماڈیول مختلف ممکنہ تراجم تجویز کرتا ہے، مگر یہ تجاویز اُردو کارپس اور لغوی اصولوں کے دائرے میں رہتی ہیں۔ معیار، اسلوب اور اصلاح ماڈیول ترجمے کو فصاحت، روانی اور معنوی صحت کے اعتبار سے جانچتا اور بہتری کے لیے متبادل تجاویز فراہم کرتا ہے۔ آؤٹ پٹ اور صارف فیڈبیک ماڈیول صارف کو ترجمہ فراہم کرتا ہے اور اسے متبادل تراجم دیکھنے، رائے دینے اور بہتری کی تجاویز شامل کرنے کی سہولت فراہم کرتا ہے، جبکہ تحقیقی ڈیٹا اور بہتری ماڈیول صارفین کے استعمال اور فیڈبیک کو محفوظ کر کے نظام کی مسلسل بہتری اور اُردو NLP کے لیے تحقیقی ڈیٹا فراہم کرتا ہے۔

استعمال ہونے والے ٹولز میں کارپس مینجمنٹ کے لیے Sketch Engine اور AntConc شامل ہیں جو اُردو متون کے استعمال، سیاق اور اسلوب کی شناخت میں معاونت فراہم کرتے ہیں۔ اُردو NLP کے لیے Urduhack اور Stanza ماڈیول استعمال کیے گئے ہیں، جو جملوں کی تقسیم، Tokenization اور نحوی شناخت کو ممکن بناتے ہیں۔ مصنوعی ذہانت کے مرکزی ماڈیول کے لیے (ChatGPT) OpenAI API اور Hugging Face Transformers کے کثیر لسانی ماڈلز جیسے mBERT اور XLM-R شامل ہیں تاکہ اُردو اور دیگر زبانوں کے درمیان معنوی تعلق قائم ہو اور تجرباتی تربیت ممکن ہو۔ لغوی اور اسلوبی وسائل میں Urdu WordNet اور تحقیقی بنیاد پر تیار شدہ اُردو لغت شامل ہیں تاکہ ترجمہ لغوی، معنوی اور اسلوبی اعتبار سے معیاری ہو۔ Django بیک اینڈ فریم ورک، React.js فرنٹ اینڈ، PostgreSQL ڈیٹابیس اور Pandas و NumPy تجزیاتی ٹولز اس نظام کو عملی، منظم اور تحقیقی اعتبار سے قابلِ اعتماد بناتے ہیں۔ یہ تمام ٹولز

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

اور ماڈیولز مل کر ایک جامع، مربوط اور تحقیقی پلیٹ فارم تشکیل دیتے ہیں جو اُردو زبان کو عالمی ڈیجیٹل منظر نامے میں باوقار، مؤثر اور تحقیقی اعتبار سے مضبوط مقام دلانے کے قابل بناتا ہے۔

References:

1. Hutchins, W. J. *Machine Translation: A Brief History*. Amsterdam: John Benjamins, 1995.
2. Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention Is All You Need." In *Advances in Neural Information Processing Systems 30*, 5998–6008. Red Hook, NY: Curran Associates, 2017.
3. Koehn, Philipp. *Statistical Machine Translation*. Cambridge: Cambridge University Press, 2010.
4. Toral, Antonio, Martijn Wieling, and Andy Way. "Post-Editing Effort of a Novel with Statistical and Neural Machine Translation." *Frontiers in Digital Humanities* 5 (2018): 9.
5. Dione, C. M. B., A. Lo, E. M. Nguer, and S. Ba. "Low-Resource Neural Machine Translation: Benchmarking State-of-the-Art Transformer for Wolof ↔ French." In *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022)*, 3210–3217. Marseille: European Language Resources Association, 2022.
6. Shterionov, Dimitar, Riccardo Superbo, Pat Nagle, Laura Casanellas, Tony O'Dowd, and Andy Way. "Human versus Automatic Quality Evaluation of NMT and PBSMT." *Machine Translation* 32, no. 3 (2018): 217–235.
7. Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural Machine Translation by Jointly Learning to Align and Translate." In *Proceedings of the International Conference on Learning Representations*. 2015.
8. Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to Sequence Learning with Neural Networks." In *Advances in Neural Information Processing Systems 27*, 3104–3112. Red Hook, NY: Curran Associates, 2014.
9. Wu, Yonghui, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, et al. "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation." arXiv:1609.08144, 2016.
10. Liu, Yinhan, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. "Multilingual Denoising Pre-training for Neural Machine Translation." *Transactions of the Association for Computational Linguistics* 8 (2020): 726–742.
11. Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. "BLEU: A Method for Automatic Evaluation of Machine Translation." In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 311–318. Philadelphia: Association for Computational Linguistics, 2002.
12. Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. "The Mathematics of Statistical Machine Translation: Parameter Estimation." *Computational Linguistics* 19, no. 2 (1993): 263–311.
13. Ali, Daud, Wahab Khan, and Dunren Che. "Urdu Language Processing: A Survey." *Artificial Intelligence Review* 47, no. 3 (2017): 279–311.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL

14. Toral, Antonio, and Andy Way. "Is Machine Translation Ready for Literary Translation? A Case Study on English–Portuguese Translation." *Digital Scholarship in the Humanities* 33, no. 1 (2018): 47–62.
15. Callison-Burch, Chris, Philipp Koehn, Christof Monz, and Omar F. Zaidan. "Findings of the 2011 Workshop on Statistical Machine Translation." In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, 22–64. Edinburgh: Association for Computational Linguistics, 2011.
16. Abbas, Qaiser. "Building a Hierarchical Annotated Corpus of Urdu: The URDU.KON-TB Treebank." In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*, *Lecture Notes in Computer Science* 7181, 66–79. Berlin: Springer, 2012.
17. Hussain, Sarmad. "Resources for Urdu Language Processing." In *Proceedings of the 6th Workshop on Asian Language Resources*, 99–107. Hyderabad: Association for Computational Linguistics, 2008.
18. Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186. Minneapolis: Association for Computational Linguistics, 2019.
19. Johnson, Melvin, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, et al. "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation." *Transactions of the Association for Computational Linguistics* 5 (2017): 339–351.
20. Neubig, Graham, and Junjie Hu. "Rapid Adaptation of Neural Machine Translation to New Languages." arXiv:1808.04189, 2018.
21. Conneau, Alexis, and Guillaume Lample. "Cross-lingual Language Model Pretraining." In *Advances in Neural Information Processing Systems 32*, 7059–7069. Red Hook, NY: Curran Associates, 2019.
22. Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. "Language Models Are Unsupervised Multitask Learners." Technical Report. OpenAI, 2019.