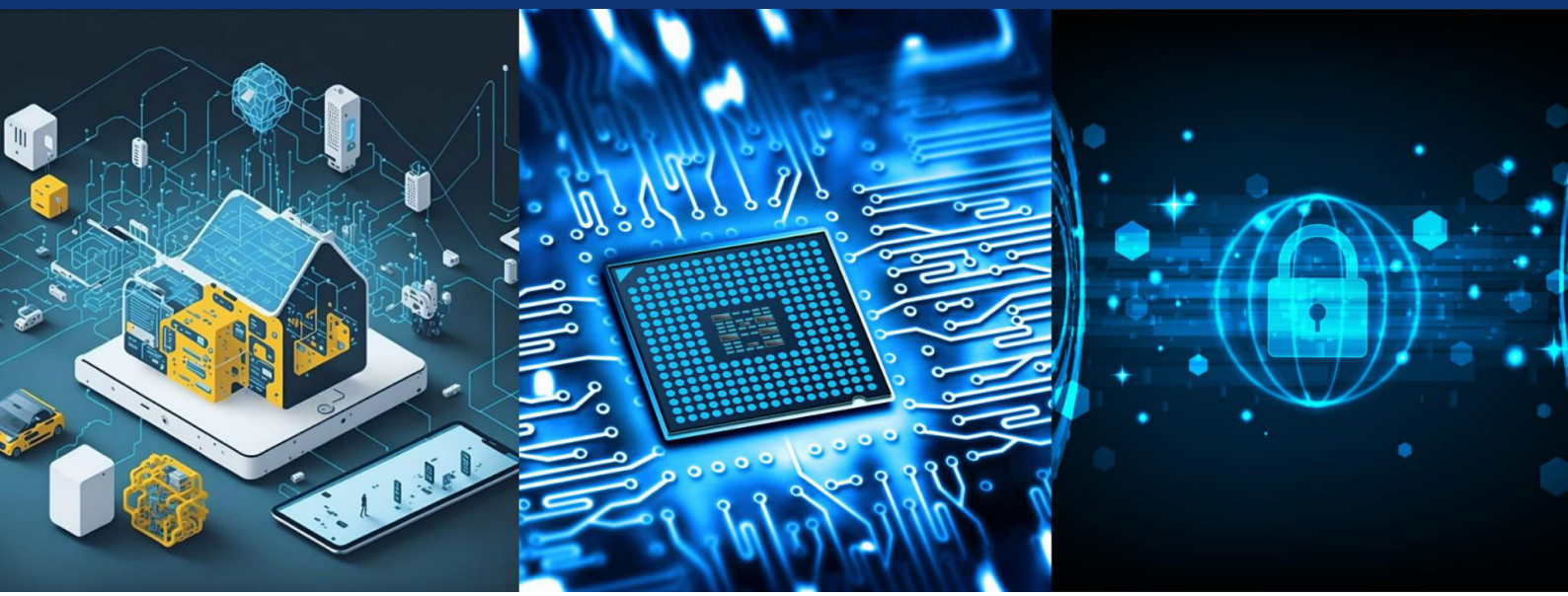
IJIRCCE

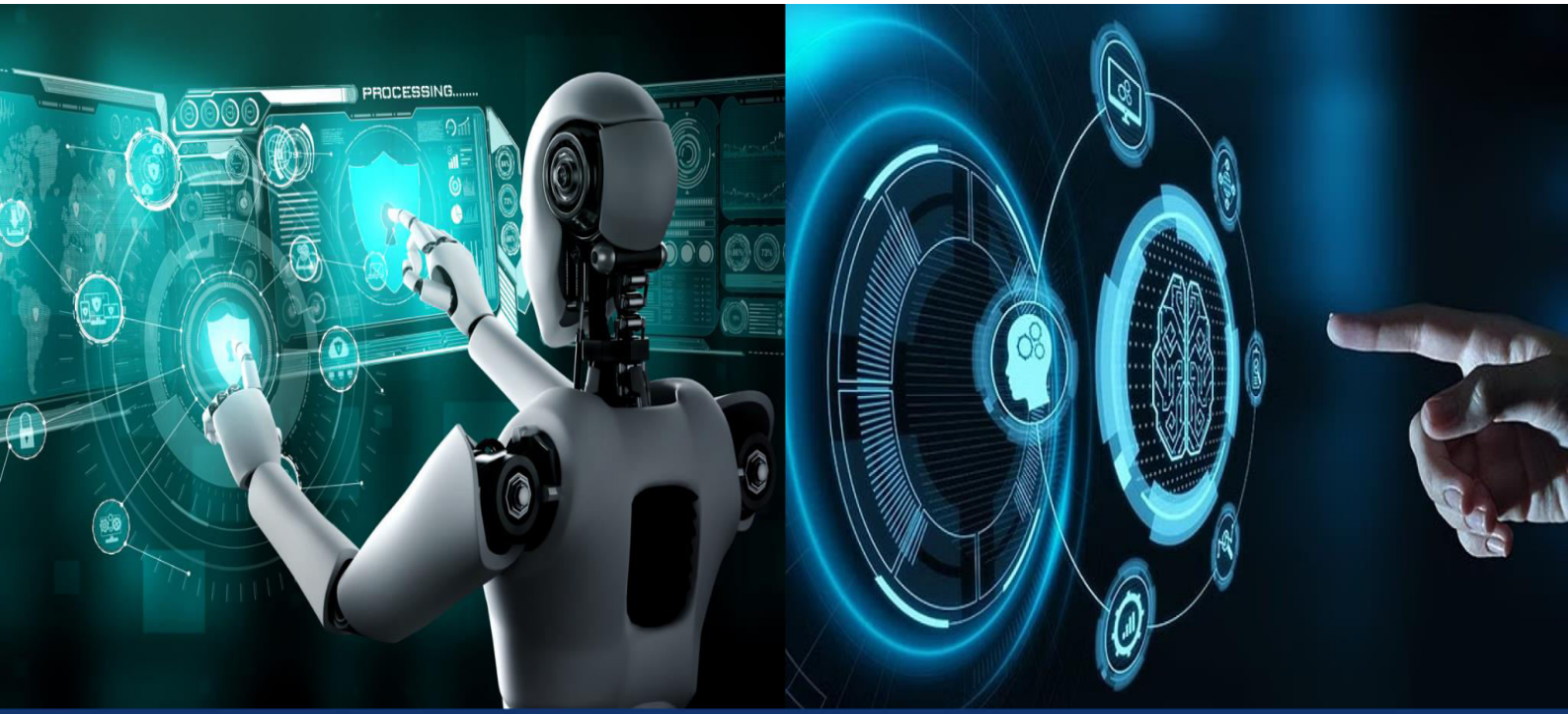
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.625

Volume 13, Issue 1, January 2025

© 2025 IJIRCCE | Volume 13, Issue 1, January 2025| DOI: 10.15680/IJIRCCE.2025.1301149

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Comparative Study of Microsoft Translator vs. Google Translate: Applications in Multilingual CV Parsing

Mrs. Varsha Pandagre[1], Mrs. Deepali Hajare[1], Rhugved Kadam[2], Satej Dhakane[2], Shreyas Kallurkar[2], Gaurav Kulkarni[2]

Professor, Dr. D. Y. Patil International University, Akurdi, Pune, India [1]

Student , Department of Artificial Intelligence and Data Science, D. Y. Patil Institute of Engineering, Management and Research, Akurdi, Pune, India [2]

ABSTRACT: Technology has brought us all closer in this age. One of the several challenges that keep us away and not able to interact with each other is the complexity of various languages. Along with this comes the automated retrieval of information for these several languages, and due to the complexity and difficulty involved in them, semantic analysis is a must. A review study is presented here where a comparison is made between Google Translate and Microsoft Translator with an emphasis on how these translation technologies can extract relevant data from multilingual resumes. Both of these translation technologies have proven to be indispensable in the job market as they allow recruiters to break down linguistic barriers in accessing talent from around the world. Moreover, it enables potential employees to get jobs with good companies without any language-related and other issues that may occur due to language proficiency. The study compares the translation models, speed, possibilities of integration with document parsing systems, and general usability for hiring processes of these tools to analyse their technical aspects. This paper seeks to assess some of these factors and ultimately establish which one is best utilized to simplify the screening process for multilingual resumes to maximize international hiring efficiency.

KEYWORDS: Multilingual CV parsing, machine translation, Google translate versus Microsoft translator, NLP and Neural machine translation, Automated Resume screening, cross-linguistic hiring, AI in Recruitment, translation accuracy, HR Technology and AI, text analysis in recruitment, Transformer models in NLP, Semantic analysis, Language processing in hiring, and AI-based resume parsing.

I. INTRODUCTION

NLP enables machines to understand, interpret, and even generate human language. Multilingual NLP takes this further with more complex systems that involve translation of languages, entity recognition, sentiment analysis, text categorization, and information extraction, due to structural and semantic differences between languages, supported by neural networks and transformer models. The growing global work environment requires businesses to process and evaluate multilingual resumes; otherwise, they would end up missing a pool of talent sources. As resumes written in other languages present another barrier to extracting useful information, the role of translating software such as Google Translate and Microsoft Translator has, therefore, increased in recent times. These solutions automatically translate high volumes of multilingual papers for HR professionals and recruitment platforms, thereby making material across languages easier to interpret. In addition, it makes possible getting a position in a reputed firm without language barriers or challenges arising from the person's language ability. Today, the job market is increasingly becoming global. Consequently, processing multilingual resumes has emerged as an essential need of the organizations searching for diversified talent pools. The barriers in language will severely restrict the traditional CV parsing tools to do justice to resumes presented in a non-native language. This is the time when tools such as Google Translate and Microsoft Translator are at their best for filling up this void. High volume recruiters as well as HR systems require the translation solution that automatically makes multilingual content interpretation in a manner that treats each candidate on par without getting biased by language. However, the quality and robustness of such translation systems in applications such as CV parsing remain high evaluation spaces. The primary contribution of this paper is comparative analysis with

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

the help of application use of Google Translate and Microsoft Translator for multilingual CV parsing. This research, to enhance connectivity as well as cross-linguistic easy communication, compares technical features like translation models, operation speed, potential for integration into document parsing systems, and general usability in hiring procedures. The article will give an in-depth overview of the technical aspects of Google Translate and Microsoft Translator, specially focused on the concepts behind them and exactly how these systems work. Part of this also includes giving a brief synopsis of the background history of these tools as well. Checking out and contrasting all the technical aspects is prime in trying to figure out which tool is the best to use with multilingual CV parsing and analysis

II. LITERATURE SURVEY

Brief History of Google Translate-
This happened eighteen years ago, accompanied by the rather conservative method of publishing Google Translate. Language structure consists of words and phrases and semantics and grammar rules each varying in its patterns. More importantly, variations in writing scripts like Latin, Cyrillic, or Devanagari influence character sets translation models work upon when applying them in the context of translational accuracy or sentiment analysis.

Originally, Google Translate relied on a very basic word-to-word translation method, whereby it typically went wrong most of the time. But Google has improved by identifying patterns in languages from the massive volumes of online data that have resulted in better accuracy in its translation. Currently, Google Translate uses neural machine translation that has taken forward translation to an extent with the use of models such as encoder- decoder frameworks. This new approach encapsulates elements of an embedding layer, which would present data in a numerical format and of an encoder, which can process raw data into meaningful representations for the computer; the context vector preserves text semantics, and a decoder, almost always improved with an attention mechanism, that produces faithful translations.
Now it's on neural machine translation, which is NMT. Its development ideas can be traced with more ancient methods like a word and sentence embedding technique as well as neural language models. Among the cutting edge NMT designs that would never have been possible to begin with without these foundation ideas is the encoder decoder framework. NMT uses more complex encoder decoder framework. [3]

Architecture Components of NMT: [3]

a) Embedding Layer: text data represented numerically or vector
b) Encoder: converts raw data into a latent space or fixed-size representation, encapsulating its key characteristics for applications like classification or translation.
c) Context Vector: a rich representation that encapsulates the semantic meaning of a string of words, enabling models to understand and preserve the context of a text or conversation while doing tasks like text generation or translation.
d) Decoder - in charge of translating the context vector into the appropriate translation. Additionally, it may incorporate an attention mechanism

Brief history of Microsoft Translator
The first version of Microsoft Translator was built between 1999 and 2000 by using the semantic structures referred as the as logical forms or LF. It was built for Microsoft Word grammar check. Microsoft could translate their Knowledge Base into Spanish, German, French, and Japanese based on this LFs-based approach.

Microsoft Translator had to transition to machine learning techniques that allowed for a data-driven approach, whereas translation algorithms learned from large parallel texts rather than any pre-defined linguistic rules. More recently, Microsoft developed its "treelet" system, where these logical forms were arranged in dependency trees so that translation could be efficiently performed and more languages supported with the launch of the user-facing translation service then known as Windows Live translator in 2007, free online text and website translations became available. During translations undertaken through the functionality of website Bilingual Viewer, a text translation of a person is undertaken on the translation webpage provided at Bing. In 2011, service was taken and offered on a scale on several other Microsoft Translator's products through a cloud application programming interface for consumer, as well as for its commercial purpose. It offered an enhanced version in the May edition of the following year in 2018. This version supported and utilized neural machine translation as a widely accepted translation medium. Except translation, the

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

revised edition encompasses transliteration, along with bilingual dictionary and word searching features that let you explore the further meanings and translate these words while discovering example uses in phrases.

III. OVERVIEW OF BOTH TECHNOLOGIES

A.Microsoft Translator: Microsoft Translator is an API-based translation service hosted in the cloud, supporting over 50 languages and designed for seamless enterprise integration. It's compatible with various developer and consumer applications, including Visual Studio, Bing, Microsoft Office, SharePoint, Skype Translator, and the Microsoft Translator apps for multiple platforms.
B.Google Translate: Google Translate, launched in 2006, is a free online translation service capable of translating words, phrases, web pages, and documents. Over time, Google has extended its functionality to recognize text in images as well. Originally based on statistical machine translation (SMT), Google Translate now integrates with many apps, such as Google Sheets, Gmail, and Dropbox, and various business platforms like HubSpot CRM and Instagram for Business, enhancing its versatility across personal and professional domains.

IV. TRANSLATION TECHNOLOGIES USED BY MICROSOFT TRANSLATOR

A.Neural machine translation : NMT utilizes artificial neural networks to autonomously translate text between languages. The NMT model's encoder-decoder framework first interprets the input text into numeric vectors representing meaning, then uses these vectors to generate a coherent translation in the target language. By focusing on essential context through attention mechanisms, NMT models trained on large multilingual datasets deliver more context-sensitive and fluent translations than traditional methods.

B.Bitext word alignment : Bitext word alignment maps corresponding words or phrases in bilingual texts, helping improve translation accuracy by using statistical neural models to align terms between languages effectively.

C.Statistical machine translation : SMT relies on parsing the syntactic structures of both source and target languages to guide translation, analyzing the grammar of words to ensure linguistic coherence and fluency.

D.Phrase-based Statistical Machine Translation (SMT) : This approach segments sentences into phrases rather than individual words and translates these phrases based on probabilities learned from large bilingual datasets, ensuring better alignment and smoother sentence flow between languages.

E.Language modeling : Language modeling is an NLP technique that anticipates word sequences based on extensive datasets, enabling Microsoft Translator to generate meaningful and coherent text for various applications.

V. METHODLOGY

1) Language Pair : There were many language pairs. These included English, Spanish, French, Russian, Hindi, Tamil, and Marathi. The languages were chosen to check the performance of the translator in different linguistic structures. In addition, texts from various domains, such as technical, legal, academic, and casual dialogue, were also considered to test the translator's ability to deal with specific terminology and colloquial language.

2) Evaluation : Translations were compared with pre- translated texts for an effective assessment of grammatical correctness. Objective measures of translation quality were provided through automated tools like BLEU and TER.

3) Contextual Comprehension Test: The translators were exposed to complex texts like philosophical literature and religious scriptures so that the tools' inability to handle fine meaning and complex language may be probed.

4) Integration : Google Translate and Microsoft Translator APIs were integrated with frameworks like Spring Boot/Java and NextJs to assess ease of integration and overall affordability.

© 2025 IJIRCCE | Volume 13, Issue 1, January 2025| DOI: 10.15680/IJIRCCE.2025.1301149

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. RESULT AND DISCUSSION

Feature	Microsoft Translator	Google Translate
Translation Accuracy	Excellent at translating industry- specific c vocabulary and technical jargon, but may struggle with subtle terms in formal settings.	performs admirably in the majority of translations but could have trouble with technical jargon.
Languages Supported	Support for 138 languages and dialects as of October 2024.	Support for 243 languages and dialects, including minor indigenous languages and well-known international languages, in October 2024.
Contextual Understanding	Superior and robust context retention in longer sentences and paragraphs, effectively handling discourse-level dependencies.	Coherence loss may result from context drift in longer or syntactically complicated translations
Translation Speed	The maximum latency for standard models is 15 seconds, and the maximum la tency for custom models is 1 20 seconds. For text with 100 characters, the response time is between 150 and 300 milliseconds.	Real-time translation typically takes less than a second for brief texts. For longer texts or documents, translations usually are completed within a few seconds.
Fluency	Translations tend to be more natural and closer to native phrasing.	Translations may occasionally sound unnatural or slightly off.
Integration Skills	Integrates well with Microsoft Office products like Word, Excel, Visual Studio, VS Code, Bing, SharePoint, Lync, Yammer, Skype Translator, for linkedin translations and Microsoft Translator apps for various devices.	Works seamlessly with third-party apps and Google services (Sheets, Docs, Gmail, Google Drive), as well as platforms like Pipedrive CRM, Dropbox, Shopify, Slack, HubSpot CRM, Facebook Pages, Microsoft 365 Email (Outlook), and Instagram.
Text Formatting Preservation	Generally, preserves formatting such as bold and italics during translation.	May lose formatting in complex documents and with multiple formatting’s
Pronunciation Features	Provides clear audio pronunciation	Offers pronunciation, but clarity may

VII. CONCLUSION

Multilingual NLP growth has opened up opportunities for hiring globally because more companies are seeking diverse talent for innovation and competitiveness. In this regard, translation tools, such as Microsoft Translator and Google Translate, play a very important role in the multilingual parsing of CVs because they can provide the extraction of candidate information in an accurate manner without language barriers.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

These translation tools will enable recruiters to standardize content contained in the CV and look into candidate's qualifications/skills, thus unbarred by the restrictions that are placed on one based on his language proficiency level. Integration with other Microsoft applications, like Word or Excel, supports translation by maintaining format for both straightforward information as well as those specific terms found within businesses. Google Translate, supporting multiple languages, also facilitates parsing CVs by allowing HR platforms to automatically translate resumes into a common language, making candidate matching and extraction of skills much easier. Google Translate may not always be accurate with certain terms, but its popularity and constant updating make it a very helpful tool in processing multilingual resumes.

REFERENCES

1] https://www.youtube.com/watch?v=_GdSC1Z1Kzs &t=8s
[2] https://research.google/blog/a-neural-network-for-m achine-translation-at-production-scale/
[3] https://arxiv.org/abs/1912.02047
[4] https://en.wikipedia.org/wiki/Microsoft_Translator# cite_note-4
[5]https://www.microsoft.com/en-us/research/projec t/machine-translation-2/?from=https%3A%2F%2Fresearch.microsoft.com%2Fen-us%2Fprojects%2Fm t%2F&type=exact
[6]https://web.archive.org/web/20190902184414/https: //www.microsoft.com/en-us/translator/business/tran slator-api/
[7] https://microsoft.com/en-us/translator/business/trans lator-api/
[8]https://learn.microsoft.com/en-us/azure/ai-servic es/translator/service-limits#:~:text=The%20Translator%20has%20a%20maximum,150%20milliseconds %20to%20300%20milliseconds.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details