Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 10, 2025

ADVANCES IN TRANSFER LEARNING FOR NATURAL LANGUAGE PROCESSING

Sohrab khan[1], Bijjar Aziz[2], Maliha Hafeez[3], Ehsan ullah khadim[4], Rizwan Ullah[5], Khalid Ali[6], Abdul Moiz[7], Ghulam Yasin[*8]

[1]*Department of Computer science, University of Peshawar*
[2]*Departmen of computer science, lasbela University of Agricultural water and marine science Uthal*
[3]*Govt. Sadiq college Women University*
[4]*Balochistan Agriculture College Quetta.*
[5]*Balochistan University of information Technology, Engineering and Management Sciences.*
[6]*Department of Computer Science, University college of Dera Murad Jamali, Lasbela University of Agriculture, Water and Marine Science Uthal, Balochistan, Pakistan*
[7]*Department of computer science, The Islamia University of BahwalPur*
[*8]*Department of Computer Science, University of Makran Panjgur*

[1]sohrab.khann6@outlook.com,[2]bijjarazizbaloch111@gmail.com,[3]malihahafeez728@gmail.com, [4]ehsankhan158@gmail.com,[5]rizwanmusakhel@gmail.com,[6]khalid.ali@ucdmj.luawms.edu.pk, [7]itx-moiz@outlook.com, [*8]yasinarman38@gmail.com

DOI: https://doi.org/10.5281/zenodo.17430902

Keywords
Transfer Learning; Natural Language Processing (NLP); Pre-trained Language Models; BERT; GPT; T5; Transformer Architecture; Deep Learning; Fine-Tuning; Few-Shot Learning; Zero-Shot Learning; Multilingual NLP; Domain-Specific Models; Text Classification; Sentiment Analysis; Question Answering; Summarization; Conversational AI; Model Interpretability; Data Bias; Low-Resource Languages; Multimodal Integration; Efficient NLP; Green AI; Sustainable Artificial Intelligence.

Article History
Received: 02 September 2025
Accepted: 12 October 2025
Published: 24 October 2025

Copyright @Author
Corresponding Author: *
Ghulam Yasin

Abstract
Transfer learning has emerged as a transformative approach in Natural Language Processing (NLP), revolutionizing how language models acquire, transfer, and apply knowledge across diverse linguistic tasks. By leveraging large-scale pre-trained models such as BERT, GPT, and T5, transfer learning has significantly reduced the dependence on extensive labeled datasets, improving efficiency and generalization across applications including text classification, sentiment analysis, translation, summarization, and conversational AI. This review highlights the evolution of transfer learning from early word embeddings to transformer-based architectures, emphasizing their adaptability in multilingual and domain-specific contexts. Despite remarkable success, challenges persist regarding computational cost, interpretability, data bias, and limited support for low-resource languages. The paper further explores future directions such as few-shot and zero-shot learning, efficient fine-tuning, multimodal integration, and sustainable "Green AI" approaches to enhance scalability and environmental efficiency. Overall, transfer learning stands at the forefront of NLP innovation, offering a robust pathway toward more generalizable, interpretable, and energy-efficient intelligent language systems.

INTRODUCTION

Multi-positives are of special interest to NLP because of transfer learning and other reasons. Transfer learning has made a significant impact on research in NLP, enabling models pre-trained on large datasets to be used for downstream tasks with relatively few labeled examples (Devlin et al., 2019). It improves the performance significantly for all NLP tasks by incorporating the broad knowledge already encoded in pre-trained models like BERT and GPT-3 which have demonstrated state-of-the-art results in several applications (Sharma and Verma 2024).

Transfer learning makes it possible to transform a variety of NLP tasks into one framework with better results on multiple benchmarks including summarization and text classification. It is highly robust: we demonstrate that it can outperform models devoid of parameters with an order of magnitude fewer comparisons (Roberts et al., 2019; Xie et al., 2020). Despite this challenge, there has been enhancing interests in transfer learning where solutions have been proposed to solve overfitting and saving of knowledge when fine-tuning deep models (Jiang et al., 2020).

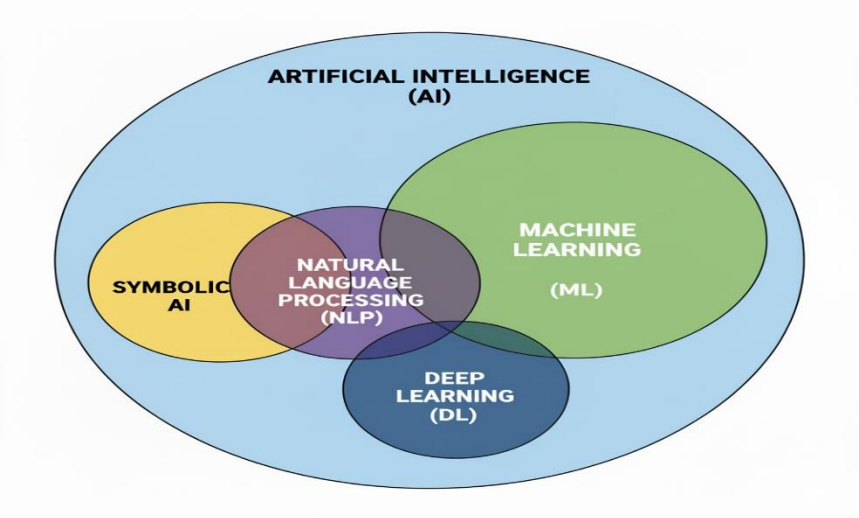


Figure 1. Relationship Between Artificial Intelligence, Machine Learning, Deep Learning, and NLP

Further, transfer learning provides a bridge that connects general purposes with downstream tasks in specific domains (Dai et al., 2021; Young et al., 2018), Medical records automation for patient screening of psychiatric disorders, which is especially useful when having smaller and imbalanced data where general models might not perform well.

Basics of Transfer Learning in NLP

The basic idea of transfer learning in NLP is to use the knowledge obtained from one task or domain to improve performance on another task or domain. The conventional supervised learning framework involves the training of models independently for a single task in isolation using a single dataset. Transfer learning is a finer-grained version of this process, where data other from the source domain or task with which to train models that can generalize well in new domains/tasks (Ruder et al., 2019). The history of transfer learning in NLP also can be dated back from the rise in word embeddings to transformer models. Word embeddings, such as Word2Vec and GloVe were a major breakthrough in NLP, offering a means of continuous vector space representations for words on the basis of context usage patterns in large corpora. These embeddings are capable of retaining the semantical relationships, based on which they have been served as basic features for a number of NLP systems (El Mahdaouy et al., 2016; Young et al.,2018). Notwithstanding, the word

embeddings also had some limitations including inability to capture relationships at sentence or document level. This motivated learning representation of sentences and larger text chunks. Unfortunately, obtaining high performance with sentence embeddings relied on training data that was expensive to obtain, which limited the scope for languages and domains with little available labeled examples (Giorgi et al., 2020).

Table.2 Applications in NLP

Application	Description of Transfer Learning Enhancement	Examples/Benefits	Citation
Text Classification	Pre-trained models BERT, ULMFiT) fine-tuned on large corpora reduce error rates and utilize semantic resources.	Significant performance improvements; Outperforms parameter-free models with fewer comparisons.	(Howard & Ruder, 2018; Roberts et al., 2019)
Sentiment Analysis	Supervised fine-tuning (BERT) captures contextual nuances across languages and domains.	Generalizes to diverse expressions; Surpasses traditional ML with less feature engineering.	(Prottasha et al., 2022; Liu et al., 2019)
Named Entity Recognition (NER)	Fine-tuned models (BERT) handle nested entities and low-resource scenarios with limited labeled data.	High performance in multilingual and biomedical domains.	(Agrawal et al., 2022; Francis et al., 2019; Sun et al., 2018)
Question Answering (QA)	Pre-trained models (BERT) fine-tuned for handling noisy inputs and complex questions in low-data settings.	Accurate answers; Generalizes across diverse tasks.	(Saeed et al., 2023; Namazifar et al., 2021)
Translation	Cross-lingual embeddings and encoders improve quality for low-resource languages without shared vocabularies.	Enhanced generalization in neural machine translation (NMT).	(Kim et al., 2019)
Summarization	Fine-tuned LMs (GPT-3) with reinforcement learning generate concise, informative summaries.	Better generalization on unseen data.	(Wang & Wu, 2023; Keneshloo et al., 2019)
Conversational AI	Pre-trained models (GPT-3) fine-tuned for multi-turn dialogues and contextual responses.	Human-like interactions; Robust in open-domain conversations.	(Talreja Wassan & Ghuriani, 2024; Kulkarni et al., 2019)

The entry of transformer models [BERT (Bidirectional Encoder Representations from Transformers)] redefined that landscape. These models are based on attention mechanisms so that they can measure the importance of sub-parts within words while inferring context and, consequently achieve better semantic representations at different levels, such as words, sentences and documents. Transformers have revolutionised NLP

tasks, and are the backbone of 21st century NLP (O. Kjell et al., 2023). This shift from word embeddings to transformers can be thought of as a mutation driven by the necessity for more contextually informed models that understand subtle language structure and meaning. Through a more intelligent encoding of textual data, transformer-based models greatly extend the scope and utility of transfer learning by producing better adapted solutions that perform well on various NLP tasks (Conneau et al., 2017; O. N. E. Kjell et al., 2021).

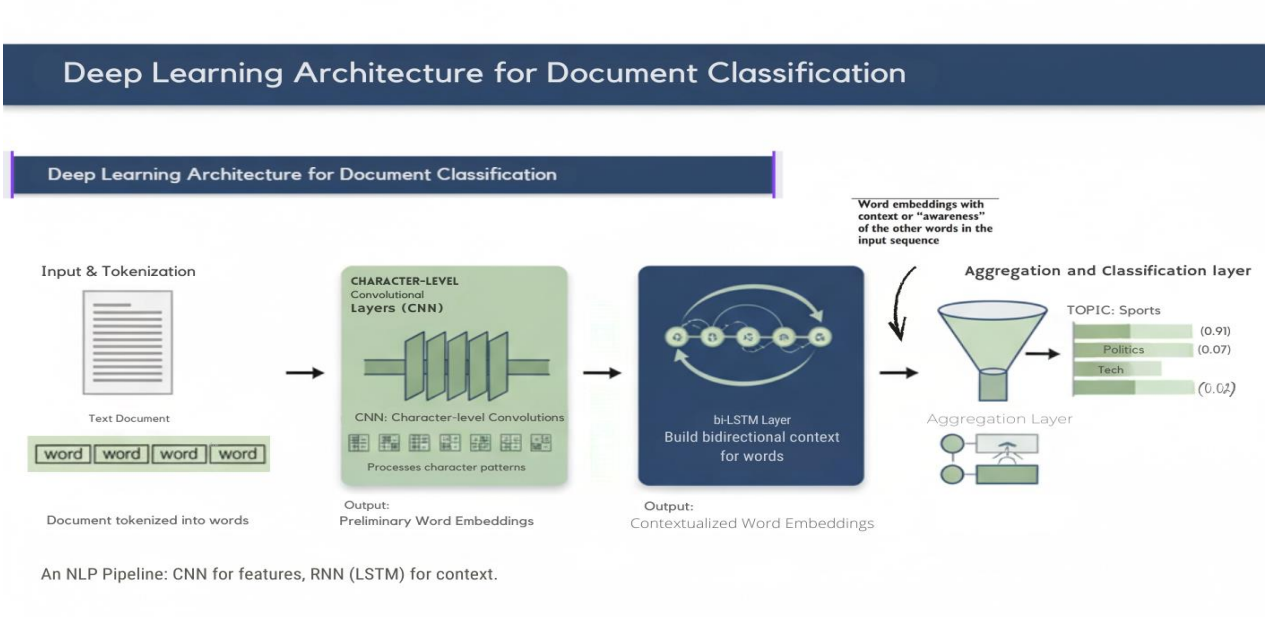


Figure 2. Deep Learning Pipeline for Document Classification in NLP

BERT (Bidirectional Encoder Representations from Transformers) BERT by Google has revolutionized NLP tasks bringing a context understanding to the intense level of understanding. It utilizes a transformer-based neural network architecture and applies methods like masked language modeling to better comprehend the context of words in a text than previous models. By fine-tuning BERT to specific datasets, state-of-the-art results have been achieved on a variety of NLP tasks; this has greatly influenced research and industry practice in the development of language models (Gardazi et al., 2025; Gupta, 2024; Souza et al., 2020). Furthermore, the impact of BERT is also seen for languages beyond English, including morphologically rich languages such as Turkish (Özçift et al., 2021).

GPT (Generative Pre-trained Transformer) models, from OpenAI, are a line of models specializing in generation of coherent and contextually relevant text. GPT models have outperformed state-of-the-art results in text generation tasks throughout benchmarks, including impressive capacity on applications such as chatbots, translation or text summarization. For example, GPT-3 generated

human-like text and its next version GPT-4 performed more proficiently in producing such text using bigger model sizes and better performance for multilingual (Baktash et al., 2023; Thakkar & Jagdishbhai, 2023). These developments have expanded possibilities for automatic text generation and interactive systems.

T5 (Text-to-Text Transfer Transformer): Another important model that shook up the field of NLP, T5 approaches all NLP tasks by casting them as text-to-text tasks. This abstraction makes it possible to use the same model, with different training data, for a wide range of NLP tasks including translation, summarization and classification. The unified style of the T5 approach has been successful in various NLP tasks and follows best practice for newcomers like code-centric software engineering tasks (Buchlak et al., 2021; Mastropaolo et al., 2021).

Multilingual and domain-specific models have greatly advanced NLP by improving model accuracy and efficiency over different tasks and fields. The popularity of large models such as GPT-4 have demonstrated the progress that has been made in designing natural language translators and question answering systems (Baktash et al., 2023), with better multimodality and context awareness displayed.

Multilingual models allow for processing and comprehension of a multitude of languages, which is important for tasks in low-resource languages or domains, such as legal informatics. Handling multiple languages is important since language specific idiosyncrasies can be overcome leading to improvements in tasks such as law consolidation and legislative document interpretation (Krasadakis et al., 2024).

Domain-specific models, including those pretrained on biomedical NLP tasks (bionlp) show that starting from domain-specific corpora yields significant performance improvement compared to the original model trained on general-domain corpora. The technique achieves state-of-the-art performance on multiple biomedical tasks using large-scale unlabeled text in the domain for pretraining, and this work sets a new benchmark in the realm of task-specific applications (Gu et al., 2021).

The use of these models in NLP tasks has also greatly been fueled by advances in transfer learning, where pretrained model (BERT or GPT-3) are fine-tuned on any type visual or linguistic task. Transfer learning has been shown successful (particularly when labeled data is limited) at adjusting models to new domains without needing the extensive training necessary to build a model from scratch (Sharma & Verma, 2024).

Applications in NLP

The developments in transfer learning have had a big influence on several NLP applications, such as for example text classification, sentiment analysis and the named entity recognition (NER). Transfer learning, notably with models such as BERT and ULMFiT has had a significant effect on text classification. Transfer learning is a technique where a language model is first pre-trained on a large corpus and then fine-tuned for particular tasks. This process enables models to utilize semantic resources derived out of large-scale text corpus, which have been shown to bring great improvements in performance across multiple natural language processing (NLP) tasks (Howard & Ruder, 2018; Roberts et al., 2019). ULMFiT, for example is an approach that fine-tune a pre-trained language model and has observed significant decreases in error rates across several text classification tasks (Howard & Ruder, 2018).

Additionally, transfer learning for sentiment analysis has expanded towards understanding contextual nuances of the sentiment across diverse text. Approaches including supervised fine-tuning such as BERT have been employed to address sentiment analysis and obtain the state-of-the-art results. Such models are able to generalise well even faced with changes in how sentiment is expressed across language and domain thanks to their strong transfer learning ability (Prottasha et al., 2022). Deep learning models have recently surpassed traditional machine learning approaches which traditionally require heavy feature engineering efforts by using transfer lear- ning (Liu et al., 2019).

Transfer learning has been crucial in the success of models trained for NER, which can achieve high performance even with small amounts of labelled data. There have been own fine-tuned models like BERT for complex natural language processing task e.g. Nested NER, where the entity boundaries overlap with each other. Transfer learning allows NER models to leverage the wealth of recent large-scale pretrained language model (e.g., (Agrawal et al., 2022; Francis et al., 2019)), and thus to specialize policy with relatively few labeled examples. Transfer learning techniques worked well in multilingual and low-resource scenarios, where labeled data is limited (P. Sun et al., 2018).

Transfer learning has greatly impacted many tasks in natural language processing (NLP) by improving the performance and the efficiency of models on question answering, translation, summarization, and conversational AI. Question Answering (QA) systems have been enhanced with transfer learning that utilizes pretrained models such as BERT, which are finetuned on particular QA tasks. These models are designed to distill and generalize information at a massive scale, which makes them more capable of understanding and answering complex questions accurately. They face issues of handling noisy inputs and unseen or out-of-vocabulary words which are

typical to the QA tasks (Saeed et al., 2023). Transfer learning permits QA models to work well, even in low-data settings, by sharing information across diverse QA tasks (Namazifar et al., 2021).

Transfer learning has been proved very effective in machine translation which is, especially for low-resource languages. For this, methods such as cross-lingual word embeddings and language-agnostic encoders enable successful transfer without shared vocabularies; they achieve significant gains in translation quality (Kim et al., 2019). This potential would be highly beneficial to the generalization abilities of NMT models across a variety of languages and the ability for improvement by fine-tuning on target language.

In text summarization, transfer learning uses deep learning methods to generate more accurate and coherent summaries. Pre-trained language models (LMs) such as GPT fine-tuned for summarization tasks have shown gains in generating summaries which are both concise and informative from challenging source texts (G. Wang & Wu, 2023). Moreover, the use of reinforcement learning in our transfer learning framework has improved the generalization of models on unseen data, some important criteria for a good summarization (Keneshloo et al., 2019). Transfer learning has played an important role in the development of dialog systems. This ensures that models such as GPT-3 have been pre-trained on large corpora and can be fine-tuned for particular conversational styles, which means they are well suited to varied and open-domain conversations. Such models could be capable of human-like responses, and could maintain a context of the user across interactions (Talreja Wassan and Ghuriani, 2024). The use of transfer learning has made such systems robust enough to participate interactively and contextually in multi-turn dialogues (Kulkarni et al., 2019). Transfer learning has redefined what NLP models can do by consuming enormous pre-trained models, leading to achieving state-of-the-art results with less data and solving a wide range of difficult language problems across diverse domains (Roberts et al., 2019; Sharma and Verma, 2024).

Challenges and Limitations

Transfer learning in NLP faces many challenges and limitations such as cost computational, bias, low resource languages, lack of interpretability. Transfer learning approaches, such as LLMs like BERT and GPT are computationally expensive. Unfortunately, training and deploying these models are computationally expensive and memory-consuming. This can be problematic also because of the high computational cost, which may not be easily accessible for some organizations. It is also resource-consuming to evaluate and adapt these models and usually need specialized hardwares, such as GPUs (Jia et al., 2023; Pritam, 2024).

It is widely observed that NLP models can suffer from bias where they may pick up on (or get biased by) the biases present in their training data and sometimes even amplify these biases. Applied to dissimilar populations or circumstances, this may produce biased conclusions and further increase existing societal disparities. For example, models extrapolating equities with regards to racial or gender considerations might undermine (imo significantly) fairness as well ethical actions. It is important to note these biases, and attempt to identify, mitigate and correct them when building such models (Patil et al., 2024; Teferra et al., 2024). One of the primary drawbacks in current transfer learning techniques, is the fact that they are mostly developed for high-resource languages such as English. These two languages are the ones with larger corpora over which models are trained, while low-resource languages (which are those underrepresented) have fewer data and resources. This makes it difficult to find appropriate training datasets and models with state-of-the-art evaluation metrics for these languages. Certain approaches to adapting models for low-resource languages include: the use of smaller architectures or domain-specific adaptations and creative cross-lingual techniques (Kholodna et al., 2024; Munaf et al., 2024). An important difficulty for transfer learning models is the lack of interpretability. Most advanced NLP models, including those that operate on deep learning frameworks, are black boxes and it is hard to see why they make the decisions they do. This lack of transparency can obstruct the diagnostic detection of error, troubleshooting and monitoring to improve performance, and even the accountability in mission critical applications. The complexity versus interpretability trade-off must also be balanced, for

more trust and usability (Chamarthi et al., 2025; Zeng et al., 2018).

Future Directions

Developments in transfer learning are powering progress in many aspects of few-shot and zero-shot learning, good (and bad) for efficient fine-tuning, and multilingual NLP. Transfer learning from existing pre-trained models on large-scale data to downstream tasks with relatively small amounts of labeled data is crucial in few-shot and zero-shot learning settings as well as such for NLP.

Transfer learning brings great benefits to few-shot and zero-shot learning where models can transfer the knowledge from very limited data. For example, few-shot learning approaches like fine-tuning (transferring pre-trained weights of models), meta-learning can help models in adapting fast to new tasks by re-using knowledge from related prior tasks (Bhati et al., 2024; A. Gupta et al., 2020). This is especially beneficial for the domains, such as bioinformatics and image classification with limited of available labeled data (X. Sun et al., 2020; Xu et al., 2025). On the contrary, zero-shot learning employs semantic representations and mapping strategies strengthened by transfer learning to categorize a test sample without being trained on any example, which is conducive to extend models' application scope for new classes not occurred during training (; X. Sun et al., 2020; W. Wang et al., 2019).

Contemporary advances in transfer learning focus on efficient fine-tuning, where only a fraction of the model's parameters are updated during adaptation to a new task. This kind of approach is very important in NLP tasks because it saves a lot of computational burden while maintaining relatively little performance loss. Fine-tuning methods such as task specific scaling and shifting function may allow models to retain previous knowledge while adjusting for new contexts (Sharma & Verma, 2024).

Weighted combination of multimodal soubasicrces, for example combining text and image or audio with text resulted in improved performance of the NLP models. Transfer learning can help this fusion by transferring knowledge among modalities, refining feature representations and for being able to learn more effective contexts in (multi\- and cross\-modal) tasks that involve various types of data'), ([Noaith N. Wang et al., 2024])). This is evident with the proposed multitask and hybrid models where multiple data streams have been exploited to improve accuracy and generalization in multimodal setup (Sharma & Verma, 2024).

Transfer learning is crucial in allowing few-shot/zero-shot learning, efficient fine-tuning and multimodal NLP by transferring the knowledge of large pre-trained models to domain-specific tasks with small amount of data. These achievements open up some promising research directions, in particular towards better generalization to unseen tasks, optimizing fine-tuning process and generalizing across various modalities in NLP applications (Sharma & Verma, 2024; X. Sun et al., 2020). It is also more important, although challenging, to develop sustainable and energy efficient algorithms for transfer learning in NLP. Here are a few ways in which such approaches are likely to shape the field. While the recent NLP models and data continue to scale up dramatically, not only necessitating more resource (CPU/GPU/TPU etc.) for execution, but also requiring efficient algorithms in order to achieve reasonable performance with a limited number of resources (Treviso et al., 2023). Efficient NLP synthesis encourages us to research into methodologies which consume less data, time, space and energy for the accessibility and affordability for NLP.

The application of sustainable AI, including model optimization techniques and effi- cient algorithm design, is an area for further research. (2024) TAKEGREEN driving for a sustainable increase in AI and artificial life performance. This includes refining the algorithms themselves and also optimizing the hardware these models run on, things like energy-efficient tensor processing units (TPUs). Automated Machine Learning has achieved great successes in NLP, but is associated with a sever energy cost. Approaches to minimize the carbon impact of AutoML, including optimizing databases for hyperparameter algorithms are vital. Model carbon emissions There is evidence that a mainstreaming of green AI practices can yield significant reductions in emissions for only tiny costs to accuracy (Castellanos-Nieves & García-Forte, 2024).

How we allocate cloud resources is important for managing the computational demand of NLP

workloads. Sophisticated techniques that exploit both model and data parallelism along with dynamic scaling can lead to an even better use of resources. This facilitiates fast computation, scalability, and low inference latency for large-scale NLP applications (Gauri Dhopavkar, 2024). The concepts of knowledge distillation and model compression are examined for enabling the models to run efficiently in resource-constrained contexts maintaining high accuracy. Such techniques can enable transfer learning: smaller models learn from larger, pre-trained models without taxing resources (Hahn & Choi, 2019). Sustainable and energy efficient practices will have a major impact on the development of transfer learning in NLP. Such initiatives will not only facilitate wider reach and impact of NLP technologies, but also support integration with larger scale environmental and resource sustainability mandates.

Table. 3 Emerging Future Directions in Transfer Learning for Natural Language Processing

Future Direction	Description	Potential Benefits	Citation
Few-Shot/Zero-Shot Learning	Transfer from pre-trained models to new tasks with minimal/no labeled data; Uses meta-learning and semantic mappings.	Rapid adaptation in domains like bioinformatics; Extends to unseen classes.	(Bhati et al., 2024; Gupta et al., 2020; Sun et al., 2020; Xu et al., 2025; Wang et al., 2019)
Efficient Fine-Tuning	Updates only a subset of parameters (e.g., scaling/shifting); Reduces computational burden.	Maintains performance with less resource use; Retains prior knowledge.	(Sharma & Verma, 2024)
Multimodal NLP	Integrates text with images/audio; Transfers knowledge across modalities for better representations.	Improves accuracy in cross-modal tasks; Enhances context understanding.	(Wang et al., 2024; Sharma & Verma, 2024)
Green/Energy-Efficient Solutions	Focuses on resource-efficient algorithms, model compression, and sustainable AI frameworks (e.g., knowledge distillation).	Reduces energy/carbon footprint; Optimizes for cloud and hardware; Promotes accessibility.	(Treviso et al., 2023; Castellanos-Nieves & García-Forte, 2024; Gauri Dhopavkar, 2024; Hahn & Choi, 2019; Tabbakh et al., 2024)

Conclusion

The rise of transfer learning has had a profound impact on the field of natural language processing (NLP) in that it has resulted in substantial improvements across various NLP tasks. Transfer learning enables models to take advantage of knowledge learnt in one domain for another, so that much less task-specific labeled training data must be used for effective learning. This has proven to be especially impactful in NLP, where obtaining large labeled datasets is often difficult and costly. The transfer learning is an influencer for NLP in many aspects. Firstly, it has achieved state-of-art performances on multiple NLP benchmarks, such as text classification, sentiment analysis and machine translation. Recent models such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have applied transfer learning by pre-training on large general corpora of text, setting state-of-the-art results with very little task-specific data.

Transfer learning has also democratized the opportunity to access and use highly sophisticated NLP models, allowing researchers and developers with limited resources to create robust apps. While

pre-trained models' generalizability across tasks minimizes the reliance on computational resources and huge amounts of data, it democratizes state-of-the-art NLP. As we look into the future, prospects for transfer learning in NLP appear to be bright. An increasing improvement on training times and data needs can also be expected, as models go on evolving. Future work may also consider limitations of the model interpretabilit and the ability to integrate with domain-specific sub-words without performance drop. Moreover, the capability to extend transfer learning structures towards multilingual and multi-model data will open the way for generalizing NLP technologies world-widely.

Conclusion

Transfer learning has fundamentally reshaped modern NLP by enabling models to transfer linguistic and contextual knowledge across tasks, thereby minimizing data requirements and enhancing performance. The advent of transformer-based architectures like BERT, GPT, and T5 has set new benchmarks across nearly all NLP domains, from sentiment analysis to question answering. These models have democratized access to high-level NLP capabilities, allowing efficient deployment even in resource-constrained environments. Nonetheless, critical challenges remain particularly in addressing bias, improving interpretability, and expanding support for underrepresented languages. Future research must prioritize energy-efficient model optimization, low-resource language adaptation, and cross-modal integration to promote equitable and sustainable AI development. Ultimately, transfer learning not only advances NLP toward more adaptive and intelligent systems but also shapes the foundation for the next generation of responsible, scalable, and human-centered artificial intelligence.

REFERENCES

Agrawal, A., Choudhary, G., Vardhan, M., Tripathi, S., Sihag, V., & Dragoni, N. (2022). BERT-Based Transfer-Learning Approach for Nested Named-Entity Recognition Using Joint Labeling. *Applied Sciences*, *12*(3), 976. https://doi.org/10.3390/app12030976

Baktash, J. A., Dawodi, M., & Femtech, A. (2023). GPT-4: A Review on Advancements and Opportunities in Natural Language Processing. *Journal of Electrical Electronics Engineering*, *2*(4). https://doi.org/10.33140/jeee.02.04.19

Bhati, N., Duggar, R., & Alzahrani, A. (2024). *Exploring Few-Shot Learning Approaches for Bioinformatics Advancements* (pp. 303–316). Igi Global. https://doi.org/10.4018/979-8-3693-1822-5.ch016

Buchlak, Q. D., Farrokhi, F., Bennett, C., & Esmaili, N. (2021). Natural Language Processing Applications in the Clinical Neurosciences: A Machine Learning Augmented Systematic Review. *Acta Neurochirurgica. Supplement*, *134*, 277–289. https://doi.org/10.1007/978-3-030-85292-4_32

Castellanos-Nieves, D., & García-Forte, L. (2024). Strategies of Automated Machine Learning for Energy Sustainability in Green Artificial Intelligence. *Applied Sciences*, *14*(14), 6196. https://doi.org/10.3390/app14146196

Chamarthi, B., Polu, O. R., Anumula, S. K., Ushmani, A., Kasralikar, P., & Aleem Syed, A. (2025). Natural Language Processing (NLP)- and Machine Learning (ML)-Enabled Operating Room Optimization: A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Systematic Review Anchored in Project Planning Theory. *Cureus*, *17*(4). https://doi.org/10.7759/cureus.82796

Conneau, A., Schwenk, H., Kiela, D., Barrault, L., & Bordes, A. (2017). *Supervised Learning of Universal Sentence Representations from Natural Language Inference Data*. Cornell University. https://doi.org/10.48550/arxiv.1705.02364

Dai, H.-J., Lee, Y.-Q., Su, C.-H., Wang, C.-K., Zhang, Y.-C., Wu, C.-S., & Kuo, C.-J. (2021). Deep Learning-Based Natural Language Processing for Screening Psychiatric Patients. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.533949

El Mahdaouy, A., El Alaoui, S. O., & Gaussier, E. (2016). *Arabic Text Classification Based on Word and Document Embeddings* (Vol. 533, pp. 32–41). Springer. https://doi.org/10.1007/978-3-319-48308-5_4

Francis, S., Van Landeghem, J., & Moens, M.-F. (2019). Transfer Learning for Named Entity

Recognition in Financial and Biomedical Documents. *Information*, *10*(8), 248. https://doi.org/10.3390/info10080248

Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Alsahfi, T., & Alshemaimri, B. (2025). BERT applications in natural language processing: a review. *Artificial Intelligence Review*, *58*(6). https://doi.org/10.1007/s10462-025-11162-5

Gauri Dhopavkar, E. A. (2024). Optimizing Resource Allocation in Cloud for Large-Scale Deep Learning Models in Natural Language Processing. *Journal of Electrical Systems*, *19*(3), 62–77. https://doi.org/10.52783/jes.652

Giorgi, J., Nitski, O., Bader, G., & Wang, B. (2020). *DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations*. Cornell University. https://doi.org/10.48550/arxiv.2006.03659

Gu, Y., Liu, X., Naumann, T., Tinn, R., Poon, H., Gao, J., Usuyama, N., Lucas, M., & Cheng, H. (2021). Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. ACM *Transactions on Computing for Healthcare*, *3*(1), 1–23. https://doi.org/10.1145/3458754

Gupta, A., Thadani, K., & O'Hare, N. (2020). *Effective Few-Shot Classification with Transfer Learning*. 1061–1066. https://doi.org/10.18653/v1/2020.coling-main.92

Gupta, R. (2024). Bidirectional encoders to state-of-the-art: a review of BERT and its transformative impact on natural language processing. *Информатика. Экономика. Управление - Informatics. Economics. Management*, *3*(1), 0311–0320. https://doi.org/10.47813/2782-5280-2024-3-1-0311-0320

Hahn, S., & Choi, H. (2019). *Self-Knowledge Distillation in Natural Language Processing*. 423–430. https://doi.org/10.26615/978-954-452-056-4_050

Howard, J., & Ruder, S. (2018). *Universal Language Model Fine-tuning for Text Classification*. Cornell University. https://doi.org/10.48550/arxiv.1801.06146

Jia, J., Liang, W., & Liang, Y. (2023). *A Review of Hybrid and Ensemble in Deep Learning for Natural Language Processing*. https://doi.org/10.48550/arxiv.2312.05589

Jiang, H., Liu, X., Gao, J., Chen, W., He, P., & Zhao, T. (2020, January 1). *SMART: Robust and Efficient Fine-Tuning for Pre-trained Natural Language Models through Principled Regularized Optimization*. https://doi.org/10.18653/v1/2020.acl-main.197

Keneshloo, Y., Reddy, C. K., & Ramakrishnan, N. (2019). *Deep Transfer Reinforcement Learning for Text Summarization* (pp. 675–683). Society for Industrial Applied Mathematics. https://doi.org/10.1137/1.9781611975673.76

Kholodna, N., Julka, S., Khodadadi, M., Gumus, M., & Granitzer, M. (2024). *LLMs in the Loop: Leveraging Large Language Model Annotations for Active Learning in Low-Resource Languages*. https://doi.org/10.48550/arxiv.2404.02261

Kim, Y., Gao, Y., & Ney, H. (2019, January 1). *Effective Cross-lingual Transfer of Neural Machine Translation Models without Shared Vocabularies*. https://doi.org/10.18653/v1/p19-1120

Kjell, O., Giorgi, S., & Schwartz, H. A. (2023). The text-package: An R-package for analyzing and visualizing human language using natural language processing and transformers. *Psychological Methods*, *28*(6), 1478–1498. https://doi.org/10.1037/met0000542

Kjell, O. N. E., Giorgi, S., & Schwartz, H. A. (2021). *The text-package: An R-package for Analyzing and Visualizing Human Language Using Natural Language Processing and Deep Learning*. Center for Open Science. https://doi.org/10.31234/osf.io/293kt

Krasadakis, P., Sakkopoulos, E., & Verykios, V. S. (2024). A Survey on Challenges and Advances in Natural Language Processing with a Focus on Legal Informatics and Low-Resource Languages. *Electronics*, *13*(3), 648. https://doi.org/10.3390/electronics13030648

Kulkarni, P., Gadgil, K., Mahabaleshwarkar, A., Kulkarni, M., & Sirsikar, N. (2019). *Conversational AI: An Overview of Methodologies, Applications & Future Scope*. 1–7.

https://doi.org/10.1109/iccubea47591.2019.9129347

Liu, R., Shi, Y., Ji, C., & Jia, M. (2019). A Survey of Sentiment Analysis Based on Transfer Learning. *IEEE Access*, *7*, 85401–85412. https://doi.org/10.1109/access.2019.2925059

Mastropaolo, A., Scalabrino, S., Poshyvanyk, D., Cooper, N., Oliveto, R., Nader Palacio, D., & Bavota, G. (2021). *Studying the Usage of Text-To-Text Transfer Transformer to Support Code-Related Tasks*. 336–347. https://doi.org/10.1109/icse43902.2021.00041

Munaf, M., Afzal, H., Iltaf, N., & Mahmood, K. (2024). Low Resource Summarization using Pre-trained Language Models. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *23*(10), 1–19. https://doi.org/10.1145/3675780

Namazifar, M., Hakkani-Tur, D., Papangelis, A., & Tur, G. (2021). *Language Model is all You Need: Natural Language Understanding as Question Answering*. 7803–7807. https://doi.org/10.1109/icassp39728.2021.9413810

Özçift, A., Akarsu, K., Yumuk, F., & Söylemez, C. (2021). Advancing natural language processing (NLP) applications of morphologically rich languages with bidirectional encoder representations from transformers (BERT): an empirical case study for Turkish. *Automatika*, *ahead-of-print*(ahead-of-print), 226–238. https://doi.org/10.1080/00051144.2021.1922150

Patil, D., Rane, N. L., Rane, J., & Desai, P. (2024). *Machine learning and deep learning: Methods, techniques, applications, challenges, and future research opportunities*. Deep Science. https://doi.org/10.70593/978-81-981367-4-9_2

Pritam, K. (2024). Advancements and Methodologies in Natural Language Processing and Machine Learning: A Comprehensive Review. *International Journal for Research in Applied Science and Engineering Technology*, *12*(4), 1495–1500. https://doi.org/10.22214/ijraset.2024.63359

Prottasha, N. J., Sami, A. A., Kowsher, M., Murad, S. A., Bairagi, A. K., Masud, M., & Baz, M. (2022). Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning. *Sensors (Basel, Switzerland)*, *22*(11), 4157. https://doi.org/10.3390/s22114157

Roberts, A., Raffel, C., Lee, K., Matena, M., Shazeer, N., Liu, P., Narang, S., Li, W., & Zhou, Y. (2019). *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. Cornell University. https://doi.org/10.48550/arxiv.1910.10683

Ruder, S., Swayamdipta, S., Wolf, T., & Peters, M. E. (2019, January 1). *Transfer Learning in Natural Language Processing*. https://doi.org/10.18653/v1/n19-5004

Saeed, N., Jhanjhi, N., & Ashraf, H. (2023). *DEEP LEARNING BASED QUESTION ANSWERING SYSTEM (SURVEY)*. Mdpi Ag. https://doi.org/10.20944/preprints202312.1739.v1

Sharma, N., & Verma, B. (2024). *Recent Advances in Transfer Learning for Natural Language Processing (NLP)* (pp. 228–254). Bentham Science. https://doi.org/10.2174/9789815238488124020014

Souza, F., Lotufo, R., & Nogueira, R. (2020). *BERTimbau: Pretrained BERT Models for Brazilian Portuguese* (pp. 403–417). Springer. https://doi.org/10.1007/978-3-030-61377-8_28

Sun, P., Yang, X., Zhao, X., & Wang, Z. (2018). *An Overview of Named Entity Recognition*. *1*, 273–278. https://doi.org/10.1109/ialp.2018.8629225

Sun, X., Sun, H., & Gu, J. (2020). Research progress of zero-shot learning. *Applied Intelligence*, *51*(6), 3600–3614. https://doi.org/10.1007/s10489-020-02075-7

Tabbakh, A., Al Amin, L., Islam, M., Mahmud, G. M. I., Chowdhury, I. K., & Mukta, M. S. H. (2024). Towards sustainable AI: a comprehensive framework for Green AI. *Discover Sustainability*, *5*(1). https://doi.org/10.1007/s43621-024-00641-4

Talreja Wassan, J., & Ghuriani, V. (2024). *Perspective Chapter: Recent Trends in Deep Learning for*

Conversational AI. Intechopen. https://doi.org/10.5772/intechopen.113250

Teferra, B. G., Rueda, A., Pang, H., Valenzano, R., Samavi, R., Krishnan, S., & Bhat, V. (2024). Screening for Depression Using Natural Language Processing: Literature Review. *Interactive Journal of Medical Research*, *13*(10258), e55067. https://doi.org/10.2196/55067

Thakkar, K. Y., & Jagdishbhai, N. (2023). Exploring the capabilities and limitations of GPT and Chat GPT in natural language processing. *Journal of Management Research and Analysis*, *10*(1), 18–20. https://doi.org/10.18231/j.jmra.2023.004

Treviso, M., Heafield, K., Forde, J. Z., Slonim, N., Strubell, E., Dodge, J., Ciosici, M. R., Raffel, C., Ji, T., Martins, P. H., Martins, A. F. T., Lee, J.-U., Gurevych, I., Hassid, M., Derczynski, L., Schwartz, R., Balasubramanian, N., Aken, B. V., Simpson, E., … Milder, P. (2023). Efficient Methods for Natural Language Processing: A Survey. *Transactions of the Association for Computational Linguistics*, *11*, 826–860. https://doi.org/10.1162/tacl_a_00577

Wang, G., & Wu, W. (2023). Surveying the landscape of text summarization with deep learning: A comprehensive review. *Discrete Mathematics, Algorithms and Applications*, *16*(03). https://doi.org/10.1142/s1793830923300047

Wang, N., Jin, W., Gao, J., Xu, C., & Bi, H. (2024). A Survey on Deep Learning for Few-Shot PolSAR Image Classification. *Remote Sensing*, *16*(24), 4632. https://doi.org/10.3390/rs16244632

Wang, W., Zheng, V. W., Miao, C., & Yu, H. (2019). A Survey of Zero-Shot Learning. *ACM Transactions on Intelligent Systems and Technology*, *10*(2), 1–37. https://doi.org/10.1145/3293318

Xu, H., Liu, L., Sun, S., Zhi, S., & Patel, V. (2025). Deep Learning for Cross-Domain Few-Shot Visual Recognition: A Survey. *ACM Computing Surveys*, *57*(8), 1–37. https://doi.org/10.1145/3718362

Zeng, Z., Luo, Y., Deng, Y., Li, X., & Naumann, T. (2018). Natural Language Processing for EHR-Based Computational Phenotyping. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *16*(1), 139–153. https://doi.org/10.1109/tcbb.2018.2849968

IEER
Institute for Excellence in Education & Research