

Paper Type: Original Article

Challenges and Perspectives in the Mathematical Foundations of Computational Linguistics

Ismail A Mageed* 

PhD, IEEE, IAENG & UK President of the ISFSEA Society, UK; im0486215@gmail.com.

Citation:

Received: 07 February 2025

Revised: 27 March 2025

Accepted: 02 May 2025

Mageed, I. A. (2025). Challenges and perspectives in the mathematical foundations of computational linguistics. *Computational Engineering and Technology Innovations*, 3(2), 45-60.

Abstract

Mathematics forms the theoretical foundation of Computational Linguistics (CL); a multidisciplinary field concerned with the computational modelling of human language. This paper reviews the central role of mathematical frameworks in the evolution of CL, from early formal language theory inspired by Chomsky to modern statistical and machine learning based approaches in Natural Language Processing (NLP). Core mathematical areas such as linear algebra, probability theory, and calculus underpin key techniques including vector space models, word embeddings, language modelling, and the training of deep neural networks through gradient-based optimization. Despite substantial progress, significant challenges remain. Accurately modelling semantics and pragmatics particularly context-dependent meaning, ambiguity, and real-world knowledge continues to be difficult. In addition, the interpretability of complex models such as large language models poses theoretical and practical concerns, alongside issues related to data sparsity in low-resource languages and ethical problems such as bias in AI-driven language systems. Looking forward, the field is expected to benefit from deeper integration of advanced mathematical concepts and emerging architectures, including transformers and graph neural networks. Future research is increasingly shifting toward holistic, goal-oriented models of language use. Continued innovation at the intersection of mathematics, linguistics, and computer science will be essential for addressing the remaining complexities of human language.

Keywords: Computational linguistics, Natural language processing, Mathematical foundations, Probability theory, Linear algebra, Deep learning.

1 | Introduction

Human language represents one of the most sophisticated cognitive abilities of human beings. It is characterized by complex grammatical structures, rich semantics, pragmatic nuances, ambiguity, and an extraordinary level of creativity and variability [1]. These properties make natural language particularly

 Corresponding Author: im0486215@gmail.com

 <https://doi.org/10.48314/ceti.vi.66>



Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

challenging for computational systems, which traditionally rely on formal, well-defined, and unambiguous representations. Computational Linguistics (CL), together with its applied counterpart Natural Language Processing (NLP), emerged to address this challenge by developing theories, models, and algorithms that enable machines to analyze, understand, and generate human language in a principled way [2–5]. From its earliest beginnings, CL has been deeply intertwined with mathematics [6]. Early symbolic approaches were inspired by formal language theory, automata, and logic, most notably in the generative grammar framework proposed by Noam Chomsky [7–10]. These approaches sought to describe linguistic competence using precise formal rules and mathematical abstractions. While rule-based systems provided valuable theoretical insights, they faced limitations in scalability and robustness when confronted with the variability and noise of real-world language use [10–14].

The field experienced a major paradigm shift with the rise of statistical and probabilistic methods in the late twentieth century [11]. Advances in probability theory, statistics, and information theory enabled researchers to model language as a stochastic process, learning patterns directly from large corpora of text and speech data [12]. Techniques such as n-gram language models, Hidden Markov Models (HMMs), and probabilistic context-free grammars demonstrated that mathematically grounded, data-driven approaches could outperform purely symbolic systems in many practical tasks [15].

In recent decades, the rapid growth of computing power and data availability has further accelerated the mathematization of CL [16–19]. Linear algebra and multivariate calculus now play a central role in modern NLP through vector space models, distributed word representations, and deep neural networks. Word embeddings, sentence representations, and large language models rely on high-dimensional vector spaces and optimization techniques based on gradient descent. These models have achieved remarkable success across a wide range of applications, including machine translation, information retrieval, sentiment analysis, and question answering [20].

Despite these advances, fundamental challenges remain. Capturing meaning beyond surface-level statistical regularities such as context-dependent semantics, pragmatics, world knowledge, and reasoning continues to be an open problem [19–22].

Moreover, the increasing complexity of modern models raises concerns about interpretability, theoretical understanding, and ethical implications, including bias and fairness in language technologies [23]. These challenges highlight the need for deeper and more principled mathematical frameworks that can bridge empirical performance with linguistic and cognitive interpretability [24].

This paper explores the role of mathematics as the unifying foundation of CL [25]. By reviewing key mathematical frameworks that have shaped the field, identified current theoretical and practical challenges, and outlined emerging perspectives, the study aims to show how continued innovation at the intersection of mathematics, linguistics, and computer science is essential for advancing our understanding and computational treatment of human language.

1.1 | Mathematical Foundations of Computational Linguistics

The development of CL has undergone a profound transformation, evolving from early rule based and symbolic approaches to statistically grounded and, more recently, deep learning-based models [23–26]. This evolution reflects both the intrinsic ambiguity of natural language and the growing recognition that purely deterministic rules are insufficient to capture its variability. Mathematical foundations have therefore become central to the design, analysis, and evaluation of modern computational language systems [24].

Statistical techniques play a crucial role in addressing linguistic ambiguity and uncertainty [20–27]. By modelling language as a probabilistic process, CL systems can estimate the likelihood of word sequences and infer latent linguistic structures from observed data.

Foundational models such as n-gram language models approximate the probability of a word given its preceding context, providing a simple yet effective framework for tasks such as speech recognition and text prediction. More sophisticated probabilistic models, including HMMs, have been widely applied to sequence labelling problems such as part-of-speech tagging and speech processing, where hidden linguistic states generate observable word or phoneme sequences [26].

Bayesian inference further enriches these approaches by offering a principled framework for incorporating prior linguistic knowledge and updating beliefs based on empirical evidence. Through Bayesian methods, computational systems can balance prior assumptions with data-driven learning, enabling more robust generalization, especially in settings with limited or noisy data. These probabilistic foundations have contributed significantly to the success of early large-scale applications such as automatic speech recognition and machine translation [28–30].

In parallel, information theory has provided powerful quantitative tools for analyzing linguistic phenomena. Concepts such as entropy, cross-entropy, and mutual information allow researchers to measure uncertainty, information content, and dependency relationships within language data.

These measures have guided effective data representation and feature selection, supporting tasks such as topic modelling, text classification, and lexical association analysis [31]. By framing language processing in terms of information flow and uncertainty reduction, information-theoretic approaches have strengthened the mathematical rigor of CL research.

The rise of machine learning and deep learning has further expanded the mathematical landscape of CL. Linear algebra underpins vector space models and distributed representations, while optimization theory and multivariate calculus enable the training of deep neural networks through gradient-based learning. As illustrated in *Fig. 1*, deep learning algorithms now play a central role across multiple domains, including computer vision, voice analytics, and NLP [28–33].

Transformer based architectures, in particular, have demonstrated the ability to capture long-range dependencies and contextual meaning at unprecedented scales. Text mining and NLP, while closely related, emphasize complementary aspects of language processing. Text mining focuses on extracting patterns, trends, and actionable insights from large collections of unstructured text, whereas NLP encompasses a broader interdisciplinary field integrating Artificial Intelligence (AI), machine learning, CL, and computer science to model and process human language more generally.

As shown conceptually in *Figs. 2-9* [19], these areas overlap significantly, sharing mathematical tools and methodological foundations.

Overall, the mathematical foundations of CL provide the unifying framework that enables theoretical insight, empirical effectiveness, and practical scalability. Continued advances in probability theory, information theory, linear algebra, and optimization remain essential for developing more accurate, interpretable, and ethically responsible language technologies.

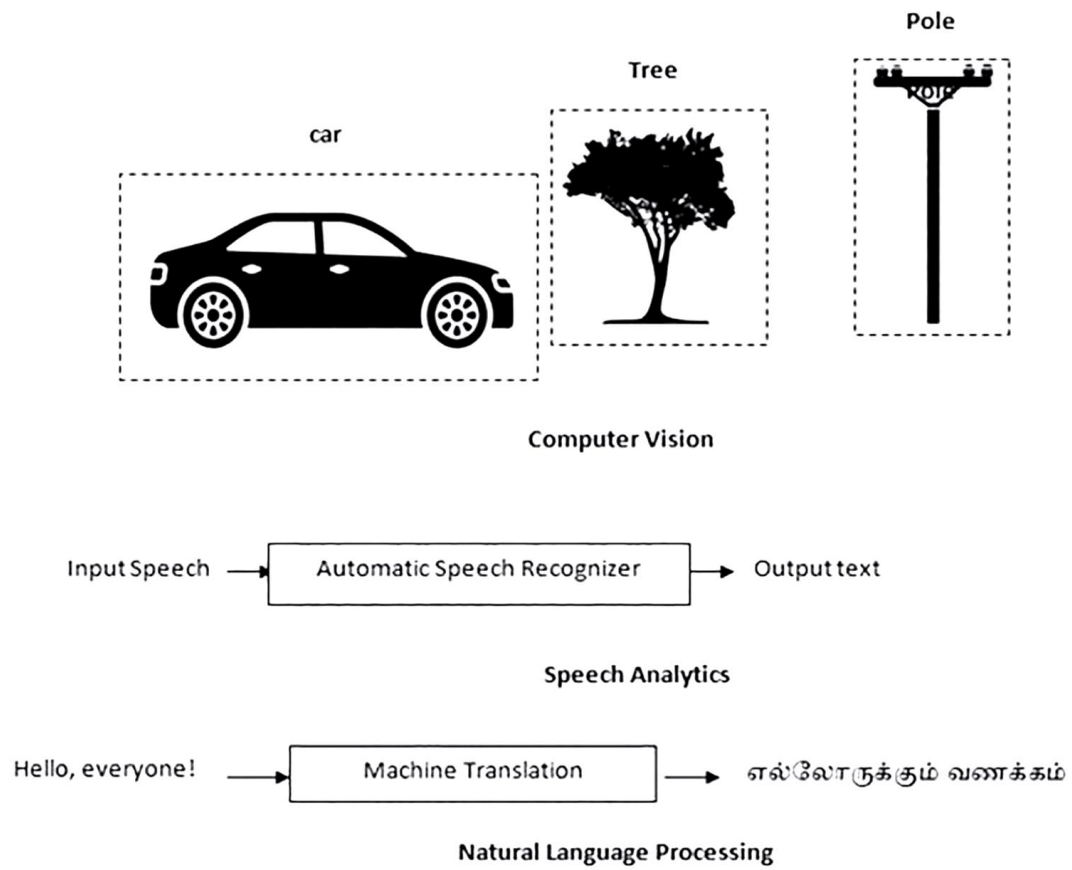


Fig. 1. DL applications.

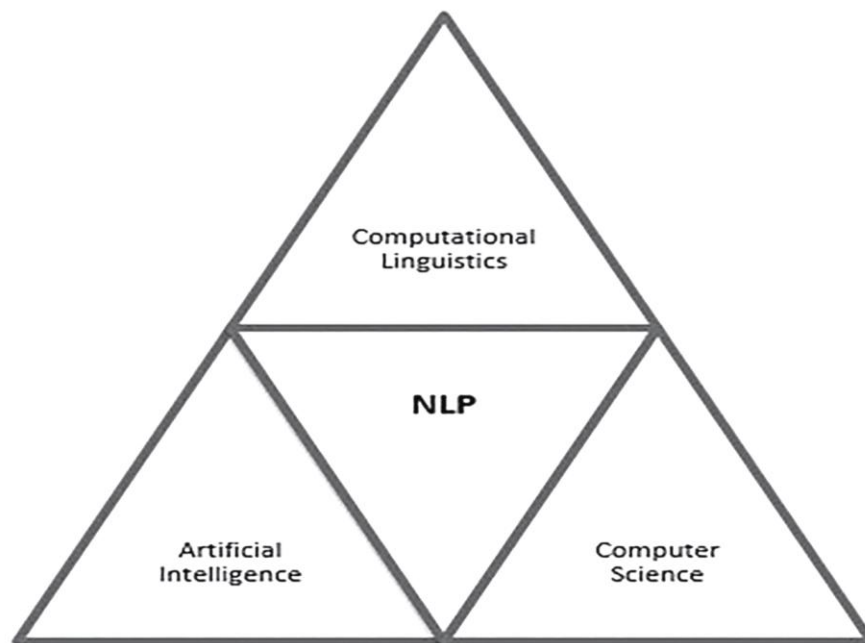


Fig. 2. NLP in a nutshell.

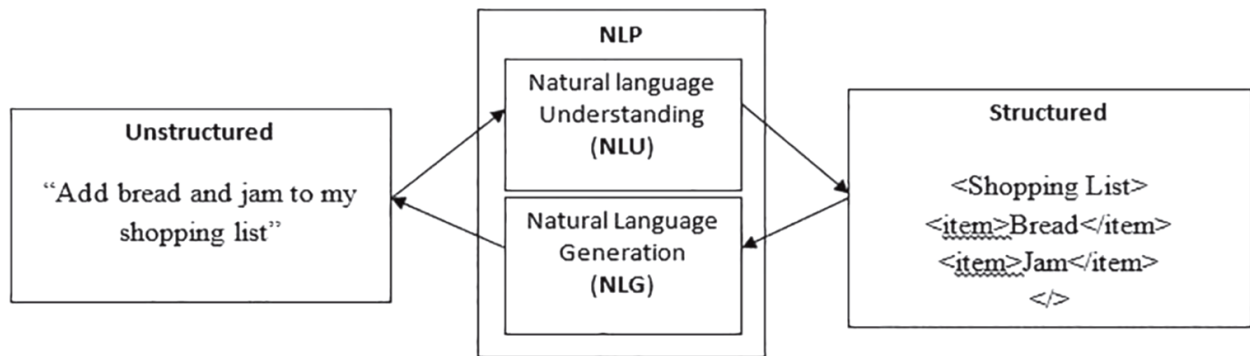


Fig. 3. Variants of NLP.

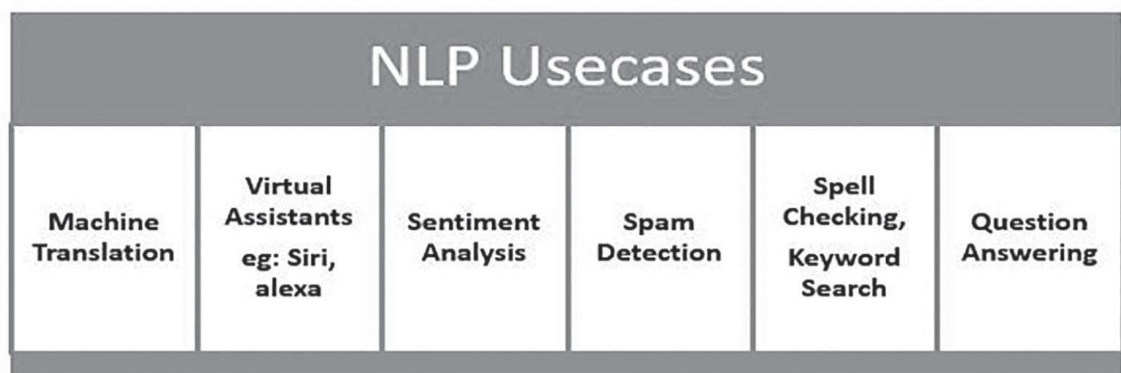


Fig. 4. NLP use cases.

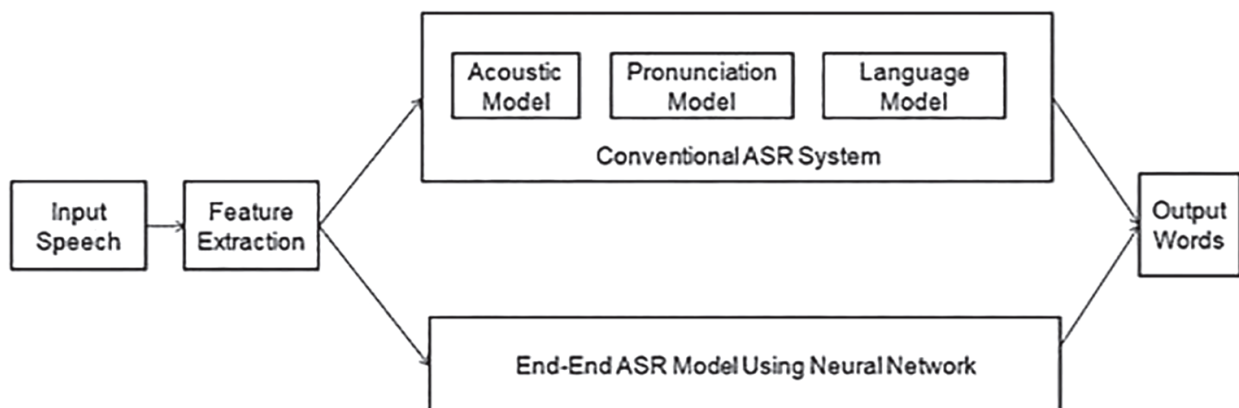


Fig. 5. Automatic speech recognition.

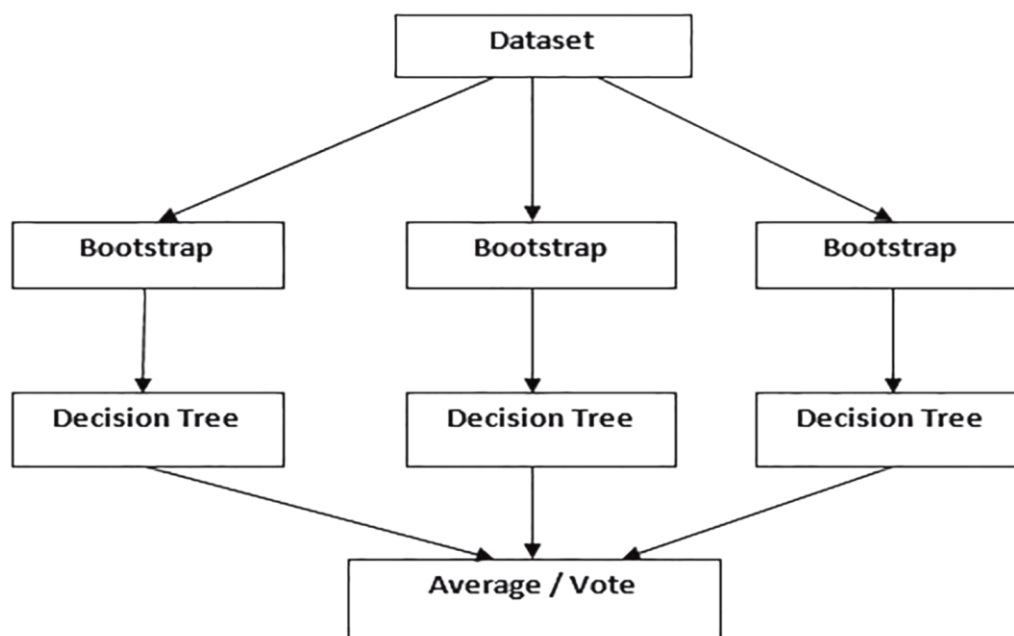


Fig. 6. Random forest algorithm.

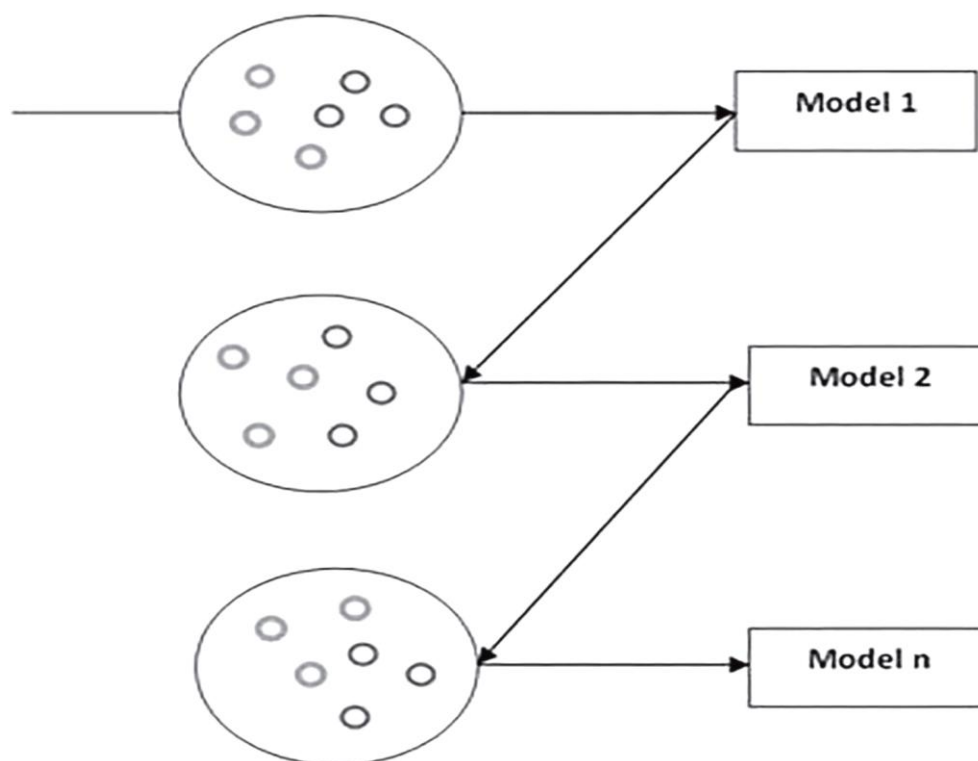


Fig. 7. Boosting model.

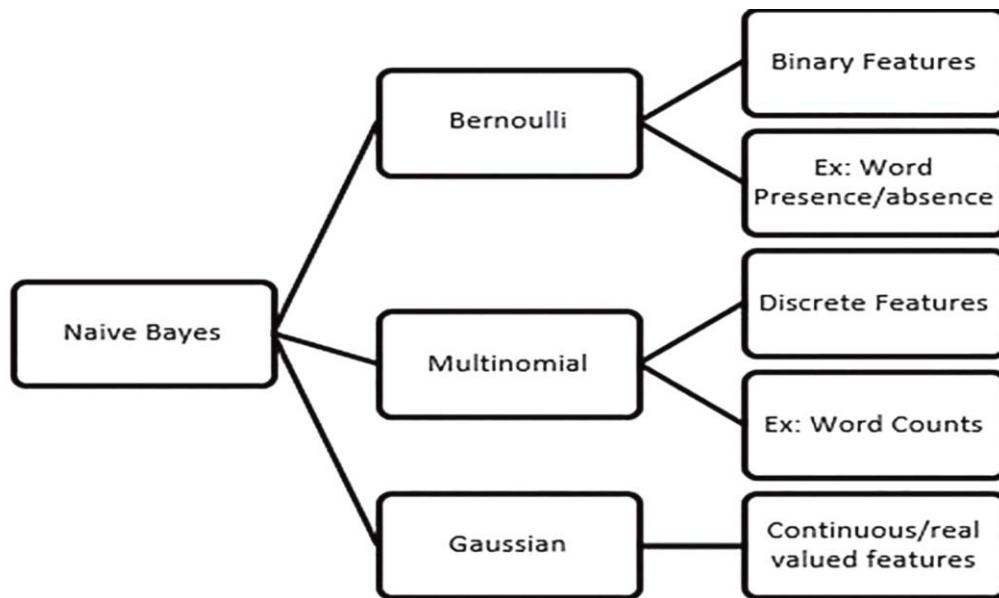


Fig. 8. Variants of naive bayes.

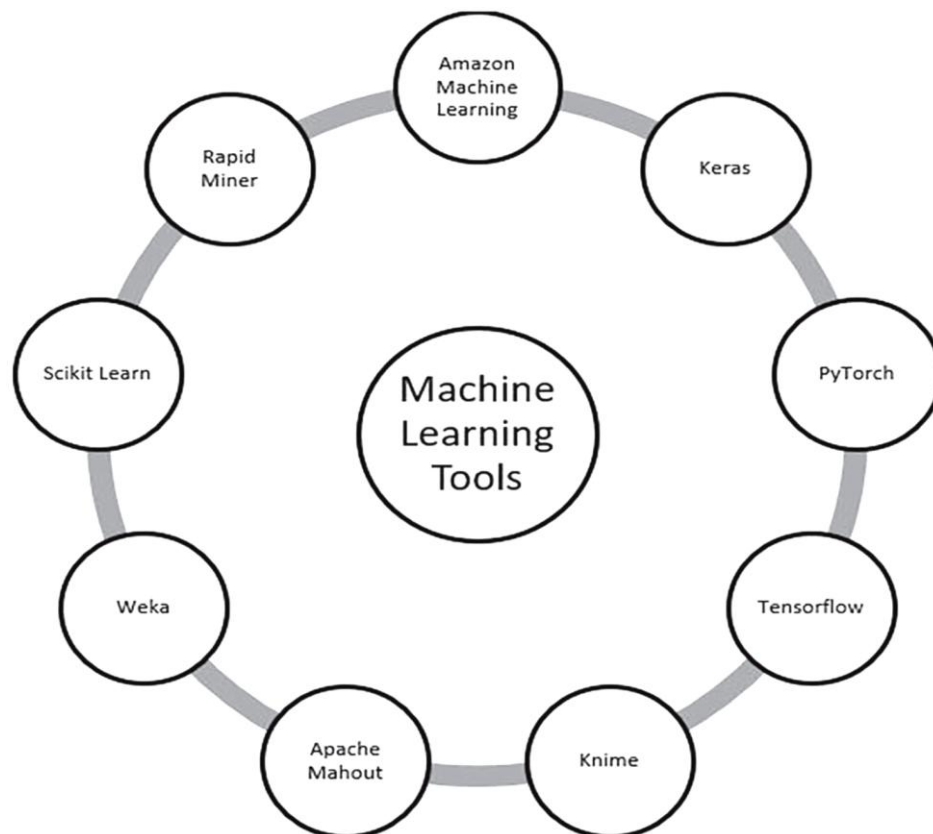


Fig. 9. ML tools.

1.2| Linear Algebra and Vector Spaces

Distributional semantics has fundamentally transformed how computational systems represent and interpret word meaning. Linear algebra provides the mathematical structure required to model words, phrases, sentences, and even entire documents as high-dimensional vectors in continuous vector spaces. Within this framework, linguistic meaning is encoded geometrically, enabling quantitative comparison and manipulation of language units [28], [30–32].

Modern neural networks operate directly on these vector representations as both inputs and internal states. Semantic similarity is reflected through vector proximity, while algebraic operations on vectors can capture meaningful linguistic relationships, as illustrated by the well-known analogy $\text{king} - \text{man} + \text{woman} \approx \text{queen}$. Such properties demonstrate the expressive power of vector space models in representing abstract semantic relationships.

Prominent examples of distributional word representations include Word2Vec, which learns embeddings from local contextual information using shallow neural architectures [34] which integrates global corpus statistics through matrix factorization techniques to capture broader semantic structure [35]. Both approaches rely heavily on linear algebraic operations, including matrix multiplication and optimization in high-dimensional spaces.

Deep learning models, particularly transformer-based architectures, further extend the role of linear algebra in CL. Transformers depend extensively on matrix multiplications and tensor operations during both model construction and training [20].

As shown in *Fig. 10*, the Transformer Model (TM) is composed of an encoder decoder architecture. The encoder consists of N stacked layers, each containing self-attention and feed-forward components, while the decoder comprises M layers that include masked self-attention, encoder decoder attention, and feed-forward components. These structures enable efficient modelling of long-range dependencies and contextual relationships in language.

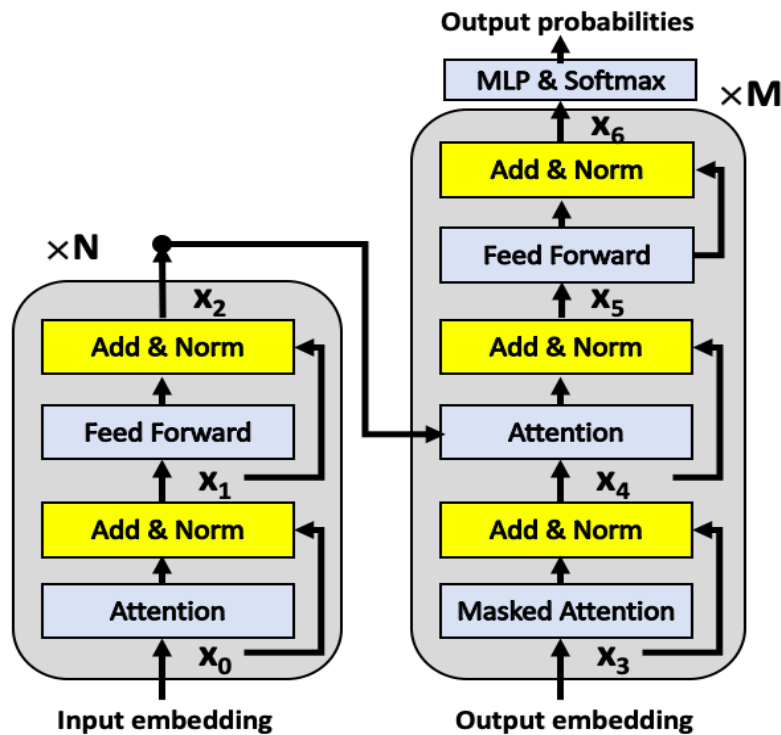


Fig. 10. TM in display.

A Venn diagram with three overlapping circles. The left circle is dark gray and labeled 'learns from'. The top-right circle is dark gray and labeled 'operations over variable structured'. The bottom-right circle is light gray and labeled 'records'. The intersection of 'learns from' and 'operations over variable structured' is a medium gray area containing a green circle labeled 'Transformer'. The intersection of 'operations over variable structured' and 'records' is a medium gray area containing a green circle labeled 'CYC'. The intersection of 'learns from' and 'records' is a medium gray area containing a green circle labeled 'NS-CL'. The intersection of all three circles is a light gray area containing a green circle labeled 'Mitsuku symbolic chatbot'. The area of 'operations over variable structured' not overlapping with the others contains a green circle labeled 'ELIZA'. The area of 'learns from' not overlapping with the others contains two green circles: 'multilayer perceptrons trained by backprop' and 'CNNs trained by backprop'.

Venn diagram shows systems for spatial, physical, psychological, temporal, and causal reasoning. Neural networks lack explicit mechanisms and representation for these types, see *Fig. 12* [36].

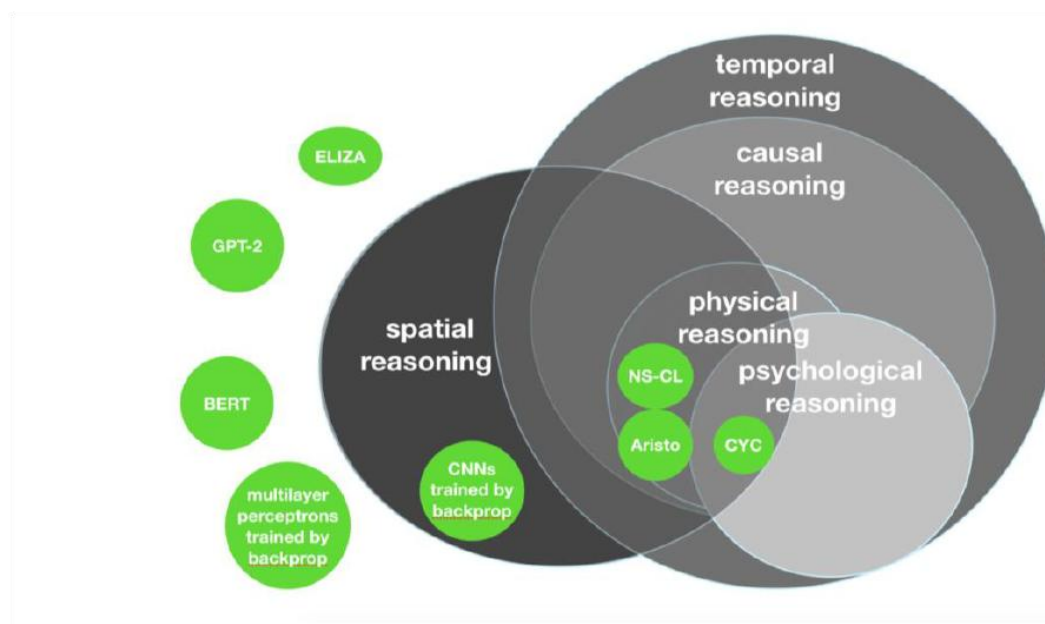


Fig. 12. Macrostructure with extensive knowledge in various fields.

1.3| Calculus and Optimization

DL models and contemporary neural networks are trained mostly with differential calculus from Calculus. By reducing loss functions, which measure the difference between model predictions and real data, methods like gradient descent and its variants are used to optimize model parameters. The main training method for multi-layer neural networks, backpropagation, is basically an application of the chain rule of differentiation [4], [11–19].

2| Challenges in Computational Linguistics

Despite the significant progress achieved in CL, the field continues to face substantial challenges that require further mathematical innovation. One of the primary obstacles is data scarcity and sparsity, as many languages, domains, and linguistic phenomena remain low-resource, thereby limiting the effectiveness of conventional statistical approaches and motivating the development of methods capable of learning from limited data, such as few-shot learning and transfer learning [13].

Moreover, the inherent ambiguity of natural language at lexical, syntactic, and semantic levels complicates meaning disambiguation and the extraction of true semantic understanding beyond surface-level statistical correlations; distinguishing between multiple meanings of a single word and capturing real-world knowledge and causality remain major challenges [3].

Although contemporary models can generate fluent and well-structured text, they often lack a deep understanding of actual events and causal relationships. Another critical issue concerns the interpretability and explainability of deep learning models, particularly large transformer-based architectures, which largely operate as black boxes; ensuring transparency, reliability, and bias awareness is essential for trustworthy real-world deployment [37].

Furthermore, multilingualism and cultural diversity introduce additional complexity, as linguistic expressions, idioms, and cultural nuances are frequently lost in direct translation, necessitating more advanced modeling strategies. Finally, ethical concerns and bias remain central challenges, since NLP models trained on large-scale datasets may inherit and amplify existing societal prejudices, underscoring the need for mathematically rigorous frameworks for bias detection, mitigation, and the promotion of fairness, accountability, and transparency in AI systems [8], [14].

3| Results and Future Directions

Innovative mathematical methodologies and cross-disciplinary collaboration are expected to continue shaping the future evolution of CL. Advances in deep learning architectures, particularly transformer-based models and attention mechanisms, have stimulated ongoing research into more efficient attention formulations, enhanced recurrent neural network designs, and neuro-symbolic hybrid frameworks that integrate statistical learning with symbolic reasoning [20].

In parallel, the growing demand for explainable and trustworthy AI has intensified mathematical research in model interpretability, including the development of formal verification techniques for AI systems [9], robust uncertainty quantification, and causal inference models, with the overarching goal of producing systems capable of providing human-understandable explanations that extend beyond mere statistical correlations. Furthermore, the integration of language with other modalities such as vision, audio, and sensor data represents a critical step toward more human-like intelligence, requiring novel mathematical tools for multimodal representation learning and data fusion.

In this context, embodied AI where language models interact with physical environments necessitates deeper integration with control theory and robotics mathematics. A prominent example is the CLIP framework, which jointly learns image and text encoders to align visual and linguistic representations; unlike conventional image models that rely on a single image feature extractor and linear classifier, CLIP enables zero-shot learning

by embedding class names or textual descriptions through the learned text encoder, effectively forming a linear classifier at inference time, as illustrated in *Fig. 13*.

Empirical evaluations demonstrate that zero-shot CLIP achieves performance comparable to fully supervised baselines. Across 16 benchmark datasets, including ImageNet, a zero-shot CLIP classifier outperforms a fully supervised linear classifier built upon ResNet-50 features, as illustrated in *Fig. 14*.

Moreover, zero-shot CLIP attains accuracy levels comparable to 4-shot learning and approaches 16-shot performance, with BiT-M and SimCLRv2 emerging as the strongest competing models; comparative results across 20 datasets with up to 16 examples per class are presented in *Fig. 15*. Robustness analyses further reveal that zero-shot CLIP significantly improves performance under natural distribution shifts, reducing the accuracy gap between ImageNet and shifted distributions by up to 75%, as shown in *Fig. 16*.

Despite these strengths, CLIP exhibits notable limitations: while competitive with ResNet-50, it remains below State-of-the-Art (SOTA) models, whose performance gains typically require computational costs up to three orders of magnitude higher, rendering such approaches impractical. Consequently, improving CLIP's computational efficiency remains a critical research direction. Additional concerns include reliance on validation sets in semi-supervised settings, sensitivity to evaluation dataset selection, and the lack of standardized benchmarks for fair comparison.

Furthermore, the flexibility of natural-language-based classifiers constrains performance on more complex tasks, and CLIP does not effectively optimize few-shot learning, often exhibiting degraded accuracy in low-shot regimes. Looking further ahead, emerging paradigms such as quantum and neuromorphic computing although largely exploratory offer potential long-term avenues for CL; neuromorphic architectures inspired by biological cognition may enable more efficient language processing, while quantum algorithms for linear algebra could theoretically accelerate certain NLP computations beyond the capabilities of classical von Neumann architectures [38].

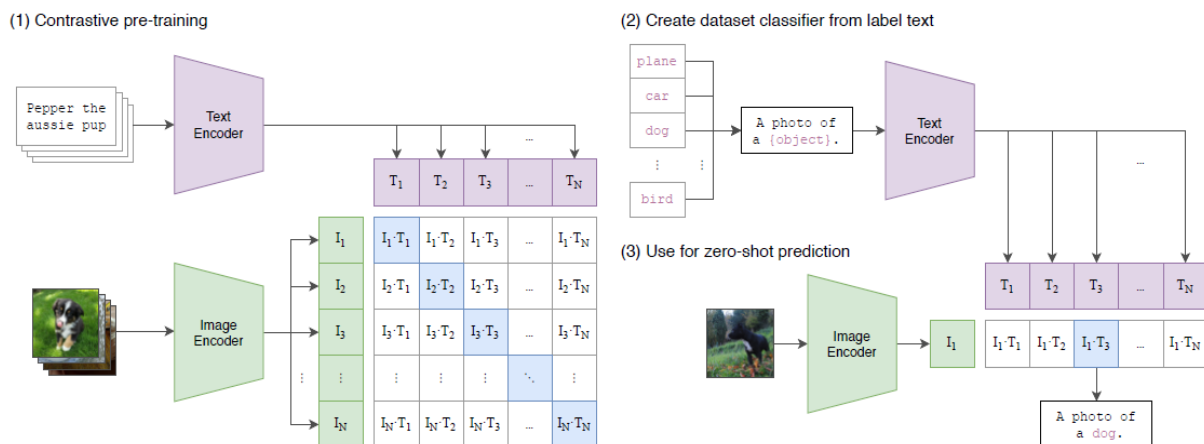


Fig. 13. Visualisation of research dynamics.

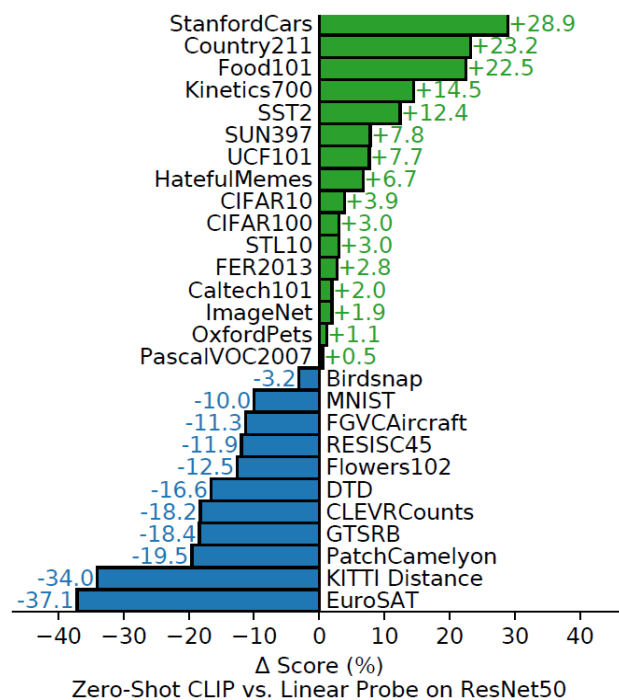


Fig. 14. Zero-shot CLIP outperforms a supervised classifier on 16 of 27 datasets, including ImageNet.

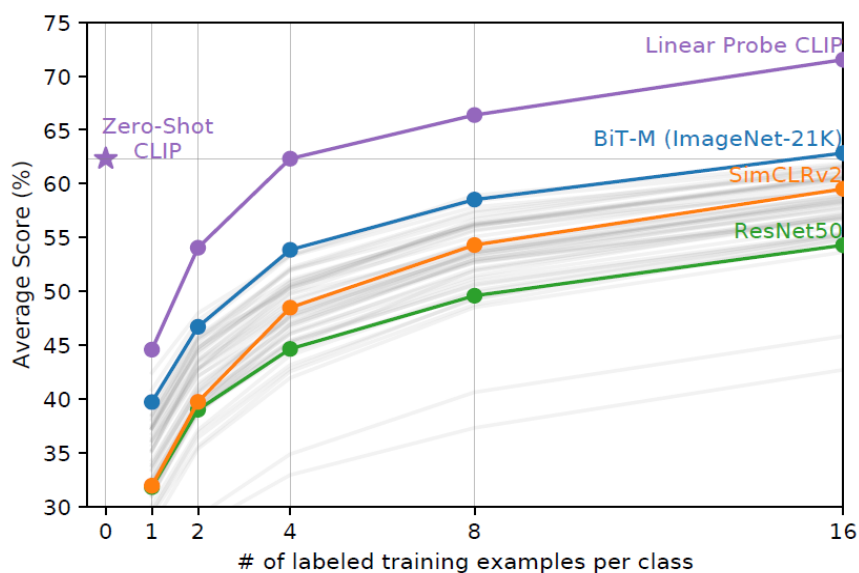


Fig. 15. Few-shot linear probes outperformed by zero-shot CLIP.

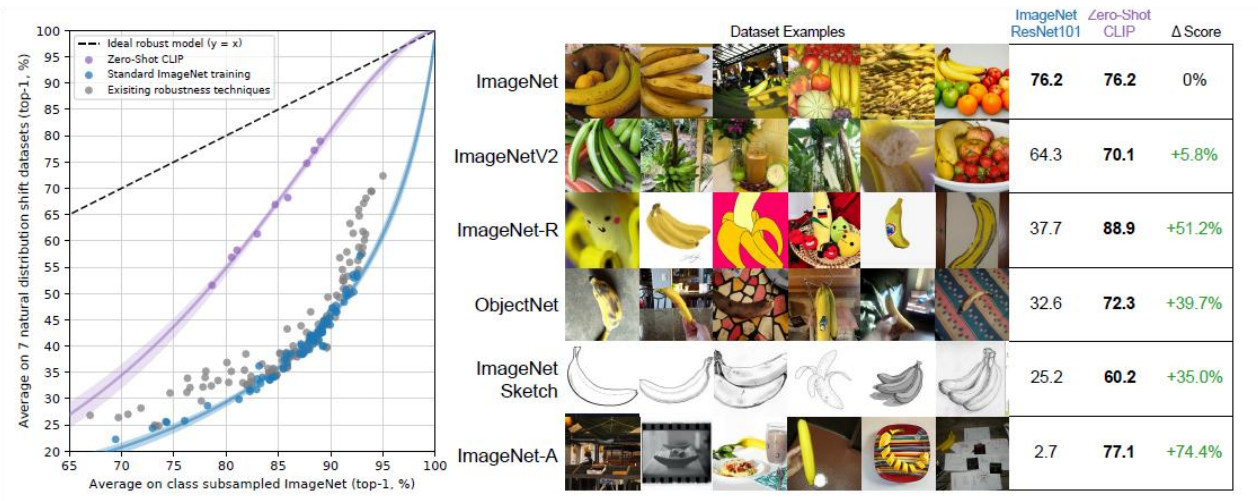


Fig. 16. Zero-shot CLIP is more robust to distribution shift than ImageNet models.

4 | Conclusion

The development of CL provides compelling evidence of the central role mathematics plays in modelling and understanding human language. From early probabilistic frameworks that supported foundational statistical models to the sophisticated linear algebra, optimization techniques, and calculus that underpin contemporary deep learning systems, mathematics has consistently supplied the essential tools for analyzing, predicting, and generating linguistic phenomena. These mathematical foundations have enabled significant advances in areas such as language modelling, machine translation, speech recognition, and semantic representation.

Despite these achievements, CL continues to face substantial challenges. Issues related to data sparsity, uncertainty, interpretability of complex models, and ethical concerns—including bias, fairness, and transparency remain open research problems. Addressing these challenges requires not only larger datasets and more powerful architectures but also deeper theoretical insight and mathematically grounded explanations of language behavior. As the field continues to evolve, the pursuit of robust, context-aware, and human-like language understanding will increasingly depend on advanced mathematical frameworks and interdisciplinary collaboration. This ongoing integration of mathematics, linguistics, and computer science reaffirms the enduring and indispensable relationship between mathematics and CL, positioning it as a key driver of future innovation in intelligent language technologies.

Authors' Contributions

All research activities and manuscript development were conducted by the Ismail A Mageed, who also approved the final manuscript.

Data Availability

All data supporting the reported findings in this research paper are provided within the manuscript.

Funding

No funding was received to conduct this study.

Conflict of Interest

The author declares that in this paper, there is no conflict of interests.

Consent for Publication

The author has provided their consent for the publication of this manuscript.

Ethics Approval and Consent to Participate

No studies involving human participants or animals were performed in this research.

References

- [1] Abusitta, A., Li, M. Q., & Fung, B. C. M. (2024). Survey on explainable AI: Techniques, challenges and open issues. *Expert Systems with Applications*, 255, 124710. <https://doi.org/10.1016/j.eswa.2024.124710>
- [2] Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., ..., & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion*, 99, 101805. <https://doi.org/10.1016/j.inffus.2023.101805>
- [3] Bender, E. M., & Koller, A. (2020). Climbing towards nlu: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185–5198). Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.463>
- [4] Bishop, C. M. (2020). Neural networks: A pattern recognition perspective. In *Handbook of Neural Computation*. CRC Press. <https://doi.org/10.1201/9780429142772-51>
- [5] Black, E. W., Garside, R., & Leech, G. N. (1993). *Statistically-driven Computer Grammars of English: The IBM/Lancaster Approach*. Rodopi. <https://books.google.com/books?id=Hkzr-LYVz2wC&printsec=frontcover>
- [6] Capuano, N., Fenza, G., Loia, V., & Stanzione, C. (2022). Explainable artificial intelligence in cybersecurity: A survey. *IEEE Access*, 10, 93575–93600. <https://doi.org/10.1109/ACCESS.2022.3204171>
- [7] chander, B., John, C., Warriar, L., & Gopalakrishnan, K. (2025). Toward trustworthy artificial intelligence (TAI) in the context of explainability and robustness. *ACM Computing Surveys.*, 57(6). <https://doi.org/10.1145/3675392>
- [8] Dev, S., Li, T., Phillips, J. M., & Srikumar, V. (2020). On measuring and mitigating biased inferences of word embeddings. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05), 7659–7666. <https://doi.org/10.1609/aaai.v34i05.6267>
- [9] Ding, P. (2024). A first course in causal inference. *Chapman and HALL/CRC*. <https://doi.org/10.1201/9781003484080>
- [10] Gerlings, J., Shollo, A., & Constantiou, I. (2020). *Reviewing the need for explainable artificial intelligence (XAI)*. <https://doi.org/10.48550/arXiv.2012.01007>
- [11] Gunning, D., Vorm, E., Wang, Y., & Turek, M. (2021). *DARPA's explainable AI (XAI) program: A retrospective*. <http://dx.doi.org/10.22541/au.163699841.19031727/v1>
- [12] Hamida, S. U., Chowdhury, M. J., Chakraborty, N. R., Biswas, K., & Sami, S. K. (2024). Exploring the landscape of explainable artificial intelligence (XAI): A systematic review of techniques and applications. *Big Data and Cognitive Computing*, 8(11), 149. <https://doi.org/10.3390/bdcc8110149>
- [13] Huang, Z., Xu, P., Liang, D., Mishra, A., & Xiang, B. (2020). *TRANS-BLSTM: Transformer with bidirectional LSTM for language understanding*. <https://doi.org/10.48550/arXiv.2003.07000>

- [14] Huff, D., & Geis, I. (2010). *How to lie with statistics*. W. W. Norton.
<https://books.google.com/books?id=5oSU5PepogEC>
- [15] Islam, M. U., Mozaharul Mottalib, M., Hassan, M., Alam, Z. I., Zobaed, S. M., & Fazle Rabby, M. (2022). The past, present, and prospective future of XAI: A comprehensive review. *Explainable Artificial Intelligence for Cyber Security: Next Generation Artificial Intelligence* (pp. 1–29). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-96630-0_1
- [16] Kouvatsos, D. D., Mageed, I. A., Anisimov, V., & Limnios, N. (2021). Non-extensive maximum entropy formalisms and inductive inferences of stable M/G/1 queue with heavy tails. *Advanced Trends in Queueing Theory*, 2. <https://doi.org/10.1002/9781119755234.ch5>
- [17] Kouvatsos, D. D., & Mageed, I. A. (2021). Formalismes de maximum d'entropie non extensive et inférence inductive d'une file d'attente M/G/1 stable à queues lourdes. *Théorie des files d'attente 2: Théorie et pratique* (pp. 182–198). ISTE Editions. <https://www.istegroup.com/fr/produit/theorie-des-files-dattente-2>
- [18] Kumar, L. A., & Renuka, D. K. (2023). *Deep learning approach for natural language processing, speech, and computer vision: Techniques and use cases*. CRC Press.
<https://books.google.com/books?id=HmqwEAAAQBAJ>
- [19] Liu, X., Duh, K., Liu, L., & Gao, J. (2020). *Very deep transformers for neural machine translation*. <https://doi.org/10.48550/arXiv.2008.07772>
- [20] Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Ser, J. Del, ..., & Stumpf, S. (2024). Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*, 106, 102301. <https://doi.org/10.1016/j.inffus.2024.102301>
- [21] Mohamed, I. A. M., & Kouvatsos, D. D. (2011). *Extended properties of the class of rényi generalized entropies in the discrete time domain*. International Conference on Computer Networks and Information Technology (pp. 1–7). IEEE. <https://doi.org/10.1109/ICCNIT.2011.6020894>
- [22] Mageed, I. A., & Kouvatsos, D. D. (2019). Information geometric structure of stable M/G/1 queue manifold and its matrix exponential. *Proceedings of the 35th Uk Performance Engineering Workshop*. University of Leeds. <https://doi.org/10.5220/0010206801530160>
- [23] Mageed, I. A., & Kouvatsos, D. D. (2021). The impact of information geometry on the analysis of the stable m/g/1 queue manifold. *Proceedings of the 10th International Conference on Operations Research and Enterprise Systems (icores 2021)* (pp. 153–160). SciTePress. <https://doi.org/10.5220/0010206801530160>
- [24] Mageed, I. A., & Zhang, Q. (2022). *An introductory survey of entropy applications to information theory, queueing theory, engineering, computer science, and statistical mechanics*. 2022 27th International Conference on Automation and Computing (ICAC) (pp. 1–6). IEEE. <https://doi.org/10.1109/ICAC55051.2022.9911077>
- [25] Mageed, I. A., Zhang, Q., Kouvatsos, D. D., & Shah, N. (2022). M/G/1 queue with balking shannonian maximum entropy closed form expression with some potential queueing applications to energy. *2022 global energy conference (GEC)* (pp. 105–110). IEEE. <https://doi.org/10.1109/GEC55014.2022.9987144>
- [26] Mageed, I. A., Yilmaz, M., Zhang, Q., Celikel, R., & Sidhu, M. S. (2022). A review of potential applications of tensor decompositions to electromagnetics and energy works. *2022 Global Energy Conference (GEC)* (pp. 385–390). IEEE. <https://doi.org/10.1109/GEC55014.2022.9987191>
- [27] Mageed, I. A., Zhang, Q., Akinci, T. C., Yilmaz, M., & Sidhu, M. S. (2022). Towards abel prize: the generalized brownian motion manifold's fisher information matrix with info-geometric applications to energy works. *2022 Global Energy Conference (GEC)* (pp. 379–384). IEEE. <https://doi.org/10.1109/GEC55014.2022.9987239>
- [28] Mageed, I. A., & Zhang, Q. (2022). *Inductive inferences of z-entropy formalism (ZEF) stable M/G/1 queue with heavy tails*. 2022 27th International Conference on Automation and Computing (ICAC) (pp. 1–6). IEEE. <https://doi.org/10.1109/ICAC55051.2022.9911090>
- [29] Mageed, I. A., & Bhat, A. H. (2022). Generalized Z-Entropy (Gze) and Fractal Dimensions. *Applied Mathematics & Information Sciences*, 16(5), 829–834. <https://doi.org/10.18576/amis/160517>

- [30] Mageed, I. A., & Zhang, Q. (2022). *An information theoretic unified global theory for a stable $m/g/1$ queue with potential maximum entropy applications to energy works*. 2022 global energy conference (GEC) (pp. 300–305). IEEE. <https://doi.org/10.1109/GEC55014.2022.9986719>
- [31] Mageed, I. A., & Mohamed, M. (2023). Chromatin can speak fractals: A review. *FOLS Conference*. University of Bradford. <https://www.researchgate.net/profile/Ismail-A-Mageed-2/publication/374583965>
- [32] Ullah, N., Khan, J. A., De Falco, I., & Sannino, G. (2024). Explainable Artificial Intelligence: Importance, Use Domains, Stages, Output Shapes, and Challenges. *CM Computing Surveys*, 57(4). <https://doi.org/10.1145/3705724>
- [33] Mageed, I. A., & Becheroul, A. (2023). The threshold theorems of generalized z-entropy (GZE) fractal dimension (fd) combined with influential applications of fd to biotechnology engineering. *39th Annual uk Performance Engineering Workshop* (p. 26). University of Birmingham. <https://www.researchgate.net/profile/Ismail-A-Mageed-2/publication/376596201>
- [34] Worth, P. J. (2023). Word embeddings and semantic spaces in natural language processing. *International Journal of Intelligence Science*, 13(1), 1–21. <https://doi.org/10.4236/ijis.2023.131001>
- [35] Wu, J., Belinkov, Y., Sajjad, H., Durrani, N., Dalvi, F., & Glass, J. (2020). Similarity analysis of contextual word representation models. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 4638–4655). Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.422>
- [36] Marcus, G. (2020). *The next decade in AI: Four steps towards robust artificial intelligence*. <https://doi.org/10.48550/arXiv.2002.06177>
- [37] Molnar, C. (2020). *Interpretable Machine Learning*. Leanpub. <https://books.google.com/books?id=jBm3DwAAQBAJ>.
- [38] Martín-Guerrero, J. D., & Lamata, L. (2022). Quantum machine learning: A tutorial. *Neurocomputing*, 470, 457–461. <https://doi.org/10.1016/j.neucom.2021.02.102>