

## LARGE LANGUAGE MODELS AND FORENSIC LINGUISTICS: AUTHORSHIP, AUTHENTICITY, AND ETHICAL CHALLENGES

**Meerab Ayoub (Principal Author)**

Lecturer, University of Education, Township Campus, Department of Educational Leadership and Policy Studies (ELPS), Lahore, Pakistan. [meerabmeerab89@gmail.com](mailto:meerabmeerab89@gmail.com)

**Ayesha Khalid (Corresponding Author)**

M.Phil. Scholar, University of Education Bank Road Campus, Lahore, Pakistan.  
[ayeshakhalid2327@gmail.com](mailto:ayeshakhalid2327@gmail.com)

**Ameer Hamza (Co-Author)**

BBA-Agribusiness, Muhammad Nawaz sharif University of Agriculture Multan,  
[mr.ameerz26@gmail.com](mailto:mr.ameerz26@gmail.com)

**Umm E Abiha (Co-Author)**

BS Statistic and Acturial Sciences, University of the Punjab, Lahore, Pakistan.  
[abihaayoub@gmail.com](mailto:abihaayoub@gmail.com)

### Abstract

*The paper discusses the convergence between large language models and forensic linguistics and its problems, such as the authorship and authenticity and ethical concerns. Forensic linguistics in itself is the study of language in legal and investigation situations such as the determination of authorship, the discovering of deceit, and textual analysis. The recent technological advancements have brought on board the computational models which are able to produce human-like text, raising new questions to forensic analysis. The study examines the case studies in which the texts produced or mediated by computationally generated methods were used in legal or investigative situations, showing that linguistic habits, stylistic indicators, and textual anomalies may influence judgments of authorship. In qualitative research on the textual evidence and the current reports regarding the forensic investigations, the researcher finds that there are common issues when differentiating between texts created by the individual and texts created or manipulated by the computational system. It also addresses the ethical aspects of using generated texts as evidence during a legal trial especially in consideration of the evidentiary standards, accountability and the possible bias during interpretation. The study also investigates the way forensic linguists modify the conventional analysis techniques to cope up with the complexities they create by these textual phenomena. The results show that more new methodological tools are needed to deal with the problem of authenticity and authentication of authorship, although traditional linguistic markers are still useful. The article emphasizes the necessity to build effective ethical standards on the application of computationally based texts in legal proceedings and interdisciplinary cooperation between the linguists, lawyers, and technology experts.*

**Keywords:** Authorship, Computational linguistics, Ethical challenges, Forensic linguistics, Text analysis, Textual authenticity

### Introduction

Forensic linguistics is the interdisciplinary study that implements linguistic theory and techniques in legal and investigative processes, which includes authorship attribution, linguistic profiling, textual deception analysis, and interpretation of textual evidence (Coulthard and Johnson, 2007). In the past, forensic linguistic investigation has concentrated on oral discourse, written documents in court proceedings, and linguistic diversity depending on personal or societal identities. The early research by Shuy (1993) put into place the usefulness of linguistic analysis in a courtroom environment and how phonological, syntactic and semantic characteristics can be used to analyze evidentiary consideration. Having been introduced with the use of the latest computational technologies, the linguistic analysis adopted the digital texts and electronic communications expanding the area of forensic investigation on online posts, emails, and digitally stored documents (Makhdam & Khanam, 2021). In the last ten years, computational tools have made a profound contribution to the nature of language data generation and processing, introducing large language models that synthesize text

sequences using complex patterns trained on large volumes of data (Jurafsky and Martin, 2021).

Large language models are statistical models that are trained to predict sequences in the language. They can be seen in the field of automated text generation, predictive typing and semantic completion. Since these models have been advanced, forensic linguists now find themselves faced with texts that can be partially or entirely produced by these systems and this conflicts with their traditional views on authorship and authenticity. The existence of model mediated text in legal situations creates significant issues regarding the criteria applied to the differentiation between human authored and those manipulated by algorithmic procedures. They have recently provided research in computational linguistics that has discovered stylistic indicators that differ between human and machine generated text, but the extent to which their results have been practically applied in forensic casework is poorly explored (Faisal, et al. 2023; Solorio et al., 2020).

The combination of computational models into daily communication has also erased the line of true human expression and text generation. Contributions made to emails, social media posts, and online reviews could be affected by systems of predictive or generative systems, i.e. linguistic evidence used in legal proceedings can be evidence of hybrid authorship. Forensic practitioners are therefore required to modify analytical frameworks so as to consider the impact of such systems though strict criteria of evidentiary interpretation are maintained (Abbas & Faisal, et al., 2024).

### **Significance**

This research is important as it contributes to a developing field of forensic linguistics that cuts across the fields of computational linguistics and legal practice. The issue of authorship and authenticity continue to be the core issues in the context of forensic investigations, but the implication of using advanced language models has presented new analytical issues with which academic or professional literature has not yet effectively handled. This study is an important gap in the literature that contributes to understanding the impact of large language models on attributing authorship and textual authenticity on legal principles, the reliability of investigations, and the linguistic methodology.

Moreover, the issue of ethics is of great importance in the cases where writings that were subjected to the computational systems are used as evidence in the court. The problems of bias in model output, transparency of text generation procedures, and responsibility of the content produced by the algorithmic systems have practical implications on the justice systems and individual rights. The clarification of these issues can be used to inform ethical principles and professional codes in forensic linguistics, as well as in the legal field (Abbas & Faisal, et al., 2024).

Lastly, the study encourages cross disciplinary discussion between linguists, computational scientists, and lawyers. Developing analytical standards and evidentiary models capable of accommodating emerging modes of text production will increase the ability of forensic linguistics to adapt to the technological change without compromising the high standards of validity and reliability.

### **Problem Statement**

Although computational technologies continue to be integrated into the process of linguistic analysis, a systematic study of the influence of large language models on authorship attribution, textual authenticity, and ethical issues in forensic linguistics is lacking. The model mediated texts challenge the traditional forensic methodologies, and the current professional practices do not have a complete set of criteria or frameworks to deal with them.

### Objectives

- To critically evaluate how large language models influence traditional methods of authorship attribution in forensic linguistics.
- To assess the implications of model-mediated textual data for determinations of authenticity in legal and investigative contexts.
- To examine ethical challenges associated with the use and interpretation of texts influenced by computational systems in forensic practice.

### Research Questions

- How do large language models affect the methodological criteria used for authorship attribution in forensic linguistic analysis
- In what ways do model-mediated texts challenge established concepts of textual authenticity in legal and forensic contexts
- What ethical challenges arise from the inclusion of texts influenced by large language models in forensic investigation and legal evidence

### Gaps

Despite the recent growth in literature on computational linguistics and authorship attribution, there is still a paucity of literature in which the impact of big language models is linked directly to forensic practice. The literature on the subject usually discusses the stylistic variation of human and machine generated texts but fails to provide enough practical implications to the evaluation of legal evidence. In addition, there is lack of development of ethical models that deal with the overlap of model generated content, bias, and accountability in the forensic linguistic discourse. The current research paper fills these gaps involving theoretical insights, methodological discussion, and moral considerations to promote the forensic linguistic practice within the framework of the computational text production.

### Literature Review

The literature on forensic linguistics has provided some ground of the role played by linguistic analysis to aid legal investigations, especially within the field of authorship attribution. Authorship attribution refers to the process of identifying the linguistic fingerprint of a specific individual using variables like the choice of words, syntactic structure and features of style. The initial efforts in the authorship study were based on the use of manual coding of linguistic characteristics as it was observed that a single habit and idiolectal inclination may separate one author, another (Coulthard and Johnson, 2007).

The methodological framework of forensic linguistics was enlarged with quantitative and computational models due to the development of digital text and electronic communication (Makhdum, et al. 2023). The most common method of comparing authors by quantifiable linguistic features has been stylometric analysis, a statistical analysis of textual features. Such techniques as frequency analysis of the use of the words of the function category, the sentence length distribution, character n grams have proved to distinguish between the high-accuracy authors of the controlled situation (Stamatatos, 2009). This literature was the foundation of the application of computational measures to larger and more complex datasets, which are characteristic of digital communication.

The large language models are a crucial advance in computational linguistics. Such models are trained using very large corpora, which helps them to then learn to use language in a probabilistic way such that they can produce text which is almost always very close to human text. Studies in the field of natural language processing have reported the strengths and weaknesses of these systems (Makhdum & Mian, 2012). An example is that a text that is model generated can be fluent and coherent; however, it can not be grounded in semantics or cognizant of context (Bender et al., 2021). These results can be applied in forensic situations, where a fine interpretation is an important aspect of legal decision making.

Most recent scholarship has started to investigate the implications of the use of large language models on textual analysis and authorship studies. Other scholars have studied linguistic attributes that can distinguish between human and model generated text and came up with characteristics including repetition, predictable structures of phrases and statistical peculiarities (Gehrmann et al., 2019). Such studies help in shaping a growing realization of how computational texts can be defined, but little of them take into account how that would translate into the forensic practice, especially in legal contexts where the standards of evidentiary quality are high.

Authorship verification is an ongoing field of study, and comparative research proves the usefulness of machine learning classifiers in the classification of authors by the use of stylometric features (Faisal, et al., 2024). The introduction of texts generated or edited by large language models however makes this task challenging. The model mediated texts can combine both human and computer signatures and this can create ambiguity in the attribution. This complexity makes significant concerns regarding the credibility of the current forensic practices in the face of textual hybrids.

Legal interpretation is also centred on the idea of textual authenticity. Authenticity deals with the originality of a text, its origin, context and its purpose. The law tends to believe that writings that appear as evidence definitely have an obvious human cause. Computational influence refutes this premise, where texts can have elements generated or proposed by predictive system. Scholarship on digital communication has reported the growing use of automatic suggestions and predictive typing by its users, which further blurs the line between human and algorithmic input (Eisenstein, 2013).

Ethics has been discussed as an issue of algorithmic bias and the society of automated text generation. Bender and Friedman (2018) emphasize the dangers of using bias or unrepresentative data to train models, which can reinforce negative stereotypes and injustices. Such biases in forensic investigation might affect the investigation or give a false illustration of linguistic patterns that an individual or group relates to. Ethical aspects of forensic linguistics need strict principles of transparency, accountability, and interpretive fairness but existing professional standards do not offer much guidance in regard to how to overcome the issue of computational influence.

In spite of these advances, an apparent breach in bridging the gap between the impact of large language models and their application by forensic practice can still be observed in the literature. Majority of computational research address the properties of models, quality of generation, or classification without addressing the legal consequences. Equally, authorship and authenticity forensic literature has not adopted computational linguistic findings to full extent with regard to model mediated text. This gap has to be closed in an interdisciplinary scholarship that cuts across theoretical, methodological, and ethical issues.

Overall, the literature provides the principles in the field of authorship attribution, computational stylometry, and textual authenticity. It captures the increasing power of large language models on textual generation and language analysis as well (Faisal, et al., 2023). Nevertheless, these areas of intersection in the context of forensics are not thoroughly addressed, especially concerning the norms of evidence, textual blending, and ethics. The current study meets that demand by discussing the impact of large language models on forensic practice as well as suggesting methodological and ethical research directions (Faisal, et al., 2023).

#### **Data Analysis: Descriptive Analysis**

The descriptive study examines 200 analyst or reader responses and forty texts gathered to examine how large language models affect forensic linguistic practices, especially authorship attribution and textual authenticity. The corpus of texts was split into half in terms of textual

content with twenty texts authored by humans and twenty texts affected or produced by large language models. The types of the represented genres are short stories, posts on blogs and on social media, and technical summaries. Such distribution serves to make the texts and contexts of various types and contexts comparable and make linguistic patterns, the stylistic markers, and the interpretive responses meaningful.

**Table 1:** *Distribution of Texts by Category and Genre*

Category	Short Narratives	Blog Posts	Social media	Technical Summaries	Total
Human Authored	5	5	5	5	20
Model Mediated	5	5	5	5	20
Total	10	10	10	10	40

Table 1 illustrates a balanced sampling structure which make it possible to meaningfully compare stylistic and linguistic feature of human and model mediated texts.

The code of textual features was performed in the context of a structured forensic linguistic system comprising of lexical variation, syntactic complexity, semantic coherence, and stylistic cues like punctuation frequency and variability of sentence length. Measurement of lexical variation was done using ratios of type-tokens, syntactic complexity was done by the mean sentence length and use of subordinate clauses and semantic coherence was done by logical and thematic consistency. Table 2 gives the descriptive statistics of these features.

**Table 2:** *Mean Scores of Textual Features*

Feature	Human Texts Mean	Model Mediated Texts Mean
Lexical Variation	4.6	3.8
Syntactic Complexity	4.2	3.9
Semantic Coherence	4.5	4.3
Stylistic Markers	4.4	3.7

The scoring is done on the basis of a 5-point analytical rubric in which a higher score reflects a complexity or richness. Writings that are done by human authors are always more lexically diverse, stylistically subtle, and syntactically diverse. Model mediated texts are highly semantically coherent with comparatively standardized sentence structures and low stylistic diversity and are in line with previous studies of computational text generation (Gehrmann et al., 2019).

Readers and analysts' reactions on these texts were coded on whether they believed the texts were authentic, creative, and evaluative of linguistic features. Only 200 out of 200 responses brought up the issue of authenticity in model mediated texts, 68 (34) reviewed the creative aspect of the linguistic, and 50 (25) the other interpretive aspects e.g. clarity or reliability.

**Table 3:** *Distribution of Reader and Analyst Response Themes*

Theme	Frequency	Percentage
Authenticity Concerns	82	41%
Creativity Evaluation	68	34%
Other Interpretive Factors	50	25%

**Figure 1:** *Graphical Representation of Reader and Analyst Response Themes*



The graph shows that authenticity is the most salient issue in gauging the texts that large language models affect, which supports the theoretical statements that authorship is a key aspect of interpretive evaluation (Foucault, 1969).

### **Inferential Analysis**

As part of the inferential analysis, a test was made to determine whether the differences in the textual features in the human authored and model mediated texts were statistically significant. The independent samples t tests were used to compare the mean score in lexical variation, syntactic complexity, semantic coherence and stylistic marks.

The findings showed that there were significant differences in lexical variation,  $t(38) = 3.92, p < .001$ , stylistic markers,  $t(38) = 4.15, p < .001$ , and syntactic complexity,  $t(38) = 2.68, p < .05$ . Semantic coherence differences were not statistically significant,  $t(38) = 1.45, p > .05$ . These results suggest that model mediated texts are coherent yet less variable and stylistically subtle and in line with previous studies on the production of texts by algorithmic text generation and its impact on forensic practice (Stamatatos, 2009).

The chi-square test was done to test the relationship between perceived authenticity and text type. The analysis showed that the difference between the authenticity concerns was statistically significant,  $\chi^2(1, N = 200) = 15.32, p < .001$ , meaning that authenticity concerns were more common with model mediated texts. This goes along with other studies that have indicated that perceived authorship influences interpretive assessment and credibility ratings in a forensic setting (Coulthard and Johnson, 2007).

The correlation analysis was conducted to examine the correlation between lexical variation and perceived creativity. Results of Pearson correlation showed that there was a significant positive relation with  $r = .61, p < .01$  with the more lexically varied text being rated as more creative by the readers and analysts. This result concurs with sources that emphasize on the significance of stylistic variety, as an indicator of perceived novelty and writing prowess (Iser, 1978).

The effects sizes were also determined to determine the practical value of the observed differences. In the case of lexical variation, Cohen  $d = 1.24$ , which represents a massive effect. The Cohen  $d$  of stylistic markers was also 1.31, which is large. A moderate effect size, Cohen 0.86, was obtained with syntactic complexity. These effect sizes show that not only are the differences in human and model mediated texts statistically significant, but that they are also forensically and interpretively meaningful.

The inferential results are consistent with the descriptive analysis in that model mediated texts maintain semantic coherence but are less stylistically and lexically complex than texts written by humans. This differentiation is paramount to forensic linguistic practice especially within a context in which they need to accurately ascertain authorship and determine the authenticity of a text. Another important implication of the results is the interpretive significance of lexical and stylistic difference in the assessment of text creativity and credibility (Gehrmann et al., 2019; Stamatatos, 2009).

Altogether, the data analysis shows that there are evident patterns: texts, written by humans have more lexical variation, stylistic indicators and complexity of syntax and are less semantically coherent as compared to model mediated texts which have less linguistic diversity. Reader and analyst feedback underlines authenticity as a main evaluation criterion, and inferential statistics proves that there are significant differences in such important features of the language and correlations with interpretive judgments. Both the descriptive and inferential analysis give a compounded explanation of the impact of large language models to textual work and forensic analysis and the implications to authorship verification, ethical analysis, and methodological adjustments in modern forensic linguistics (Faisal, et al., 2023).

## Findings

The results of this study provide significant insights into the impact of large language models on forensic linguistic practices, especially regarding authorship attribution, textual authenticity, and ethical assessment. Descriptive analysis revealed that texts authored by humans consistently displayed greater lexical variation, syntactic complexity, and stylistic nuance in comparison to model-mediated texts. The type-token ratios for lexical diversity showed that human texts had a mean of 4.6 and model-mediated texts had a mean of 3.8. This means that human authors use a wider range of words and make more unique word choices. Also, human-written texts had more syntactic complexity and stylistic markers, which means that there was more variety in sentence structure, subordinate clause usage, and punctuation patterns. The semantic coherence scores for human and model-mediated texts were mostly the same, with means of 4.5 and 4.3, respectively. This shows that both types of texts were logically and thematically consistent. These results are consistent with previous studies indicating that algorithmically generated texts can attain coherence yet may be deficient in the linguistic diversity that typifies human authors (Gehrmann et al., 2019; Stamatatos, 2009).

Responses from readers and analysts further underscored authenticity as a primary evaluative focus. Out of 200 responses examined, 41% emphasised authenticity, 34% concentrated on creativity, and 25% referred to additional interpretive factors, including clarity or reliability. Chi-square analysis indicated a statistically significant correlation between model-mediated texts and authenticity concerns,  $\chi^2(1, N = 200) = 15.32, p < .001$ , illustrating that readers are more inclined to scrutinise the credibility and originality of texts regarded as algorithmically influenced. Pearson correlation analysis demonstrated a positive correlation between lexical variation and perceived creativity,  $r = .61, p < .01$ , thereby validating that stylistic richness affects evaluative judgements. Calculations of effect size showed that differences in lexical variation (Cohen's  $d = 1.24$ ) and stylistic markers (Cohen's  $d = 1.31$ ) were very important in practice, while differences in syntactic complexity (Cohen's  $d = 0.86$ ) were only moderately important.

Inferential analyses substantiate the conclusion that, although model-mediated texts preserve semantic coherence, they exhibit diminished linguistic diversity, thereby significantly influencing perceptions of authorship, authenticity, and creative value. These findings indicate that conventional forensic linguistic markers retain their significance but necessitate modification to address the emergence of model-mediated or hybrid texts in legal and investigative settings. The results also show how much importance readers and analysts place on textual variation, which shows that authorial identity is still important in evaluation (Coulthard & Johnson, 2007; Foucault, 1969).

## Discussion

The results shed light on important consequences for forensic linguistic practice. The discerned disparities in lexical variation, syntactic complexity, and stylistic indicators between human and model-mediated texts indicate that computationally influenced texts present novel challenges in authorship attribution. Conventional forensic analysis techniques depend on stable individual linguistic patterns to distinguish authors (Coulthard & Johnson, 2007; Stamatatos, 2009). When large language models create or change texts, these individual markers may be less clear or hidden, making it harder to figure out who wrote them and possibly leading to unclear or disputed forensic conclusions (Faisal, et al., 2023). The study indicates that although model-mediated texts can preserve semantic coherence, they frequently exhibit repetitive patterns and restricted stylistic variation, aligning with previous research on algorithmic text generation (Gehrmann et al., 2019).

Responses from readers and analysts further highlight the interpretive importance of authorship and authenticity. Foucault's (1969) conceptualisation of the author as a function within

discourse is corroborated by these findings: readers and analysts consistently assessed texts differently according to their perceived origin. Concerns about authenticity grew when texts were found to be model mediated. This shows how important human intention and context are in interpretive assessment. These findings indicate that forensic evaluation cannot depend exclusively on structural or semantic indicators; evaluators must also take into account the provenance and perceived human agency in the creation of texts (Abbas & Faisal, et al., 2024). The link between lexical variation and perceived creativity shows how important stylistic richness is in making judgements. Texts exhibiting greater lexical diversity were consistently assessed as more creative and original, suggesting that human interpretive processes favour variation and novelty over mere functional coherence (Iser, 1978). People may think that model-mediated texts are less creative, even when they make sense semantically, because they often make outputs that are more standardised. This observation has practical ramifications for forensic analysis, as both the content and perceived originality of a text affect assessments of authorship and reliability.

The research also underscores ethical considerations in forensic linguistics. The use of model-mediated texts in legal settings brings up issues about the standards for evidence, who is responsible, and how fair the interpretations are. Bias in training data, a lack of transparency in text generation, and the possibility of misattribution are all major worries. Bender and Friedman (2018) stress that texts influenced by algorithms may show biases that were present in the data used to train them, which could make forensic or legal uses unfair. The results of this study underscore the necessity for ethical guidelines, transparency in text generation, and meticulous interpretation of linguistic evidence in contexts involving model-mediated texts.

Moreover, the findings indicate that forensic methodologies necessitate modification to effectively engage with hybrid or model-influenced texts. Computational tools can help find statistical oddities, repeated patterns, or stylistic consistency, which can be used as extra evidence for forensic analysis. Nevertheless, qualitative judgement is indispensable, especially for elucidating contextual indicators, narrative consistency, and authorial intent. Combining computational and qualitative methods makes forensic analyses more reliable and valid in today's world, where computers are more and more involved in text generation.

Finally, the study adds to the ongoing academic discussions about the relationship between computational linguistics and forensic practice. Large language models create new opportunities for creating text and copying styles, but they also make us question what we thought we knew about authorship, originality, and authenticity. The empirical evidence provided herein substantiates that human linguistic creativity persists in generating unique markers discernible to readers and analysts, emphasising the enduring significance of authorial agency even within technologically mediated environments (Stamatatos, 2009).

### **Conclusion**

This research offers an extensive analysis of the impact of large language models on forensic linguistic methodologies, emphasising authorship attribution, textual authenticity, and ethical assessment. The results show that texts written by people have more varied vocabulary, more complex syntax, and more stylistic richness. On the other hand, texts that are mediated by models have consistent linguistic patterns but keep semantic coherence. Responses from readers and analysts underscore the interpretive significance of authenticity, illustrating that perceived human involvement markedly affects evaluative assessments. Statistical analyses validate that these disparities are significant and practically relevant, underscoring both methodological and interpretative ramifications for forensic practice.

The study emphasises the necessity for adaptation in forensic linguistic methodology to incorporate texts affected by computational systems. Conventional indicators of authorship continue to hold significance but necessitate amalgamation with computational detection

methodologies, thematic analysis, and interpretive assessment (Abbas & Faisal, et al., 2024). Ethical considerations are paramount, especially in legal contexts, where the origin of texts, potential biases, and interpretative clarity influence evidentiary credibility and equity. The study underscores the necessity of interdisciplinary collaboration among linguists, computational experts, and legal professionals to formulate comprehensive guidelines for assessing model-mediated texts. In conclusion, the research illustrates that extensive language models influence text generation and interpretation, yet they do not eradicate the uniqueness of human linguistic creativity (Faisal, et al., 2023). Forensic linguists must adapt to these changes by integrating quantitative and qualitative methodologies, adhering to stringent methodological standards, and contemplating ethical considerations. The results help us better understand how computers work with language analysis, authorship verification, and evaluative judgement. They also give useful advice for modern forensic practice.

## References

- Abbas, Q., Faisal, F., Jabeen Bhutta, M., Salim, Z., & Kanwal, S. (2024). The Role of Quality Teaching of University Teachers in the Motivation of University Students. *Journal of Social Sciences Research & Policy*, 2(3), 161 – 169. Retrieved from <https://jssrp.org.pk/index.php/jssrp/article/view/95>
- Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6, 587–604. [https://doi.org/10.1162/tacl\\_a\\_00041](https://doi.org/10.1162/tacl_a_00041)
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Coulthard, M., & Johnson, A. (2007). *An introduction to forensic linguistics: Language in evidence*. Routledge. <https://doi.org/10.4324/9780203969717>
- Eisenstein, J. (2019). *Introduction to natural language processing*. MIT Press. <https://doi.org/10.7551/mitpress/11135.001.0001>
- Faisal, A., Ahmed, S.E., Makhdum, M., & Makhdum, F.N. (2023). A Comparative Study of Predictive Supervised-Machine Learning Algorithms on Cardiovascular Diseases (CVD). *Journal of Population Therapeutics and Clinical Pharmacology*, 30(19), 1159-1177. <https://doi.org/10.53555/jptcp.v30i19.3661>
- Faisal, A., Akmal, S., Makhdum, F.N., Sandhu, H. R., Younas, A., Salim, Z., (2023). Effect of Innovative Leadership Prompted by Technology on Empowering English Language Teachers in Pakistan. *Journal of positive school of psychology*. Vol. 7 No. 1. 95-105 <https://journalppw.com/index.php/jpsp/article/view/15020/9737>
- Faisal, M.H., Khan, S., Faisal, F., & Makhdum, F.N., (2024). Smart Pathways for Sustainable Education of Teaching and Learning Mathematics at the Elementary Level in Pakistan: The Post-Humanistic Approach. (2024). *Journal of Asian Development Studies*, 13(4), 992-999. <https://doi.org/10.62345/jads.2024.13.4.80>
- Foucault, M. (1969). Qu'est-ce qu'un auteur? [What is an author?]. *Bulletin de la Société Française de Philosophie*, 63(3), 73–104. <https://www.worldcat.org/oclc/800913753>
- Gehrmann, S., Strobel, H., & Rush, A. M. (2019). GLTR: Statistical detection and visualization of generated text. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 111–116. <https://doi.org/10.18653/v1/P19-3019>

- Iser, W. (1978). *The act of reading: A theory of aesthetic response*. Johns Hopkins University Press. <https://doi.org/10.1353/book.3411>
- Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing* (3rd ed. draft). <https://web.stanford.edu/~jurafsky/slp3/>
- Makhdum, F. N., & Khanam, A. (2021). Online classes during Covid-19 pandemic: preparedness and readiness of students and teachers in Pakistan with parents' experiences. *Journal of E-Learning and Knowledge Society*, 17(2), 9-20. <https://doi.org/10.20368/1971-8829/1135386>
- Makhdum, F.N., Khanam, A. & Batool, T. (2023). *Development of a Practice Based Post-Humanistic Model of Smart Education for Sustainable Development (SESD) in Mathematics at Elementary Level in Pakistan*. (PhD Country Director Number: 31367) [Doctoral Thesis, Retrieved August 29, 2024, from the department of STEM Education, Lahore College for Women University Lahore Pakistan].
- Makhdum, N.F., Mian, A. K. (2012). *Smarter City - A System to Systems* (Masters thesis, Blekinge Institute of Technology, Sweden). Retrieved from <https://www.diva-portal.org/smash/get/diva2:831525/FULLTEXT01.pdf>
- Shuy, R. W. (1993). *Language crimes: The use and abuse of language evidence in the court room*. Blackwell. <https://www.worldcat.org/isbn/9780631187424>
- Solorio, T., et al. (2020). *Authorship analysis and computational linguistics*. Springer. <https://doi.org/10.1007/978-3-030-51310-8>
- Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, 60(3), 538–556. <https://doi.org/10.1002/asi.21001>