Your assigned topic, required structure, outline for your work, and instructions for completing your assignment.

Important Note on Structure and Outline
You must follow the provided below required structure and outline for your assigned topic and work. This outline includes comprehensive details and a roadmap designed to make your task easy, informative, enjoyable, impactful, and scientifically robust. The structure and outline include:

- **Section and Subsection Titles** (including Abstract).
- **Required Information for Each Section and Subsection.**
- **Word Limits for Each Section and Subsection.**
- **Type, Numbers, and Placement of Figures, Tables, and Evaluation Formulas.**
- **Suggested References and Recommended Numbers for Each Section and Subsection (if applicable).**
- **Any Supplementary Material for Enhanced Impact.**

Ensure you adhere to this outline to produce a well-organized, comprehensive, and high-quality submission.

Use of Generative AI Tools
You are allowed and encouraged to use generative AI tools to gather reliable information for your work. However, you must instruct the tools to provide information drawn from **reliable and peer-reviewed sources**, such as academic journals and textbooks. Avoid non-peer-reviewed references such as Wikipedia, arXiv preprints, most websites, blogs, and forums.

Accepted Sources
You may cite federal, government, and company websites (e.g., Apple, Microsoft, IBM, hospitals, and medical centers) if the information is directly sourced from them.

Citation Requirements (IEEE Format)
- In-text citations must follow the **IEEE citation style**. This format uses numbered citations in square brackets, placed **inside punctuation marks**.
- Examples:
  - "The studies conducted by Smith and Jones [2], [3] support this hypothesis."
  - "Previous research has shown promising results [4]–[6]."
- Citations are numbered sequentially throughout your paper.

Reference List
- All sources cited in the text must be listed in detail under the section titled **"References"** at the end of your work.
- Example:
  - [1] A. B. Smith and C. D. Jones, "Optimization of neural networks using evolutionary algorithms," *IEEE Trans. Neural Netw.*, vol. 23, no. 2, pp. 245–255, Feb. 2024. DOI: 10.1109/TNN.2024.1234567.
  - [2] J. R. Doe and M. L. Khan, "AI-based optimization techniques for data integration," in *Proc. Int. Conf. Machine Learning (ICML)*, Paris, France, 2024, pp. 567–572.
- If a generative AI tool provides incomplete or fabricated references, use **Google Scholar** to verify and obtain accurate references. Copy the complete reference as demonstrated below under **IEEE Citation Style Guide**.

Creating Illustrations, Figures, Tables, and Equations
- You may use tools like **draw.io (preferred)**, Figma, Microsoft Word, or any generative AI tool for creating **informative, high-resolution figures and diagrams**.
- For LaTeX or Mermaid Chart codes generated by AI tools, you can render them in **Overleaf (for LaTeX codes)** or **Mermaid Chart renderers** to save as images for insertion.
- **Tables:**

    - Can be created in Word or generated by AI tools. Ensure tables are editable in Word—screenshots or images of tables will receive a zero.
- **Equations:**
    - Formulate equations directly in Word or use AI tools to generate LaTeX code. Copy the code to Word to generate authentic equations. Images of equations are unacceptable and will result in a zero.

Formatting Guidelines

- **Figures and Tables:**
    - Numbered separately with descriptive captions that clearly explain their content without requiring reference to the text.
    - Ensure high-quality, well-labeled, and appropriately captioned figures.
    - Position tables and figures within the text at relevant points.
- **Text Style:**
    - Ensure consistency in section headings, subheadings, captions, and references.
    - Follow IEEE citation style throughout.

Proofreading and Submission

- Thoroughly proofread your work for grammar, spelling, punctuation, and formatting errors.
- Ensure clarity and coherence across all sections with appropriate transitions.
- Ensure adherence to formatting guidelines, citation style, and figure/table formatting.
- **Word Count:** The main content (from Introduction to References) should be between **3,500–4,000 words**. You may exceed this limit if necessary but should not fall below **3,500 words**. Exclusions include:
    - Your name and affiliation.
    - Abstract.
    - Captions.
    - References.
    - Supplementary materials.

IEEE Citation Style Guide

The IEEE citation style is commonly used in engineering, technology, and related disciplines. It uses **numbered citations** in the text, which correspond to **full references listed at the end of the document**.

1. In-Text Citation

In IEEE, citations are **numbered sequentially** throughout your paper. The numbers are enclosed in **square brackets** and are placed inside punctuation marks.

Format:

- **Single Source:** [1]
- **Multiple Sources:** [1], [2], [3] or [1]–[3] for consecutive references.

Examples:

- As stated in [1], the algorithm performs better under noisy conditions.
- The studies conducted by Smith and Jones [2], [3] support this hypothesis.
- Previous research has shown promising results [4]–[6].

2. References List

The references list appears at the end of your paper, **ordered numerically** by citation number **(not alphabetically)**. Each reference should be **single-spaced with a hanging indent**.

Format:

[#] Author's Initials. Last Name, “Title of the paper,” Abbreviated Journal Name, vol. x, no. x, pp. xxx–xxx, Month, Year. DOI (if available).

3. Common Reference Types

a. Journal Article

[1] A. B. Smith and C. D. Jones, “Optimization of neural networks using evolutionary algorithms,” IEEE Trans. Neural Netw., vol. 23, no. 2, pp. 245–255, Feb. 2024. DOI: 10.1109/TNN.2024.1234567.

b. Conference Paper
[2] J. R. Doe and M. L. Khan, “AI-based optimization techniques for data integration,” in Proc. Int. Conf. Machine Learning (ICML), Paris, France, 2024, pp. 567–572.

c. Book
[3] R. J. Wilson, Introduction to Graph Theory, 5th ed. London, U.K.: Pearson, 2022.

d. Book Chapter
[4] M. Brown, “Deep learning approaches for natural language processing,” in Artificial Intelligence Handbook, G. A. Stone and P. Richards, Eds. New York, NY, USA: Springer, 2023, pp. 120–145.

e. Online Sources
[5] National Institute of Standards and Technology (NIST), “Quantum computing roadmap,” NIST, Washington, D.C., USA. [Online]. Available: https://www.nist.gov/quantum-computing. [Accessed: Apr. 2, 2025].

Example Usage
In-Text Citations:
Generative AI techniques have been applied to various domains including text generation, image synthesis, and music composition [1], [2]. Additionally, recent advancements in transfer learning have improved model performance across tasks [3]–[5].

References:
[1] A. B. Smith and C. D. Jones, “Optimization of neural networks using evolutionary algorithms,” IEEE Trans. Neural Netw., vol. 23, no. 2, pp. 245–255, Feb. 2024. DOI: 10.1109/TNN.2024.1234567.
[2] J. R. Doe and M. L. Khan, “AI-based optimization techniques for data integration,” in Proc. Int. Conf. Machine Learning (ICML), Paris, France, 2024, pp. 567–572.
[3] R. J. Wilson, Introduction to Graph Theory, 5th ed. London, U.K.: Pearson, 2022.
[4] M. Brown, “Deep learning approaches for natural language processing,” in Artificial Intelligence Handbook, G. A. Stone and P. Richards, Eds. New York, NY, USA: Springer, 2023, pp. 120–145.
[5] National Institute of Standards and Technology (NIST), “Quantum computing roadmap,” NIST, Washington, D.C., USA. [Online]. Available: https://www.nist.gov/quantum-computing. [Accessed: Apr. 2, 2025].

Captioning Figures and Tables in IEEE Style
In IEEE style, **figures and tables** are numbered separately and labeled **with descriptive captions**. The captions should be **brief but descriptive enough** to explain the content without requiring the reader to refer to the text.

1. Figures
📌 **Rules:**
- **Numbering:** Figures are numbered consecutively (e.g., Fig. 1, Fig. 2, etc.) throughout the document.
- **Placement:** The figure caption is placed **below** the figure.
- **Text Style:** Centered, italicized, with the figure number in bold.
- **Citation:** Refer to figures in the text as “Fig. 1” (do not use “Figure 1”).

Example of a Figure Caption
Fig. 1. Architecture of the proposed neural network model.

In-Text Reference Example:
As shown in Fig. 1, the architecture consists of multiple convolutional layers followed by fully connected layers.

2. Tables
📌 **Rules:**
- **Numbering:** Tables are numbered consecutively (e.g., Table I, Table II, etc.) throughout the document.
- **Placement:** The table caption is placed **above** the table.
- **Text Style:** Centered, uppercase, with the table number in bold.
- **Citation:** Refer to tables in the text as “Table I” (use Roman numerals).

Example of a Table Caption

TABLE I. SUMMARY OF DATASETS USED FOR TRAINING

Dataset	Samples	Features	Source
MNIST	60,000	784	LeCun et al.
CIFAR-10	50,000	3,072	Krizhevsky
ImageNet	1.2M	150,528	Deng et al.

In-Text Reference Example:

The datasets utilized in this study are summarized in Table I.

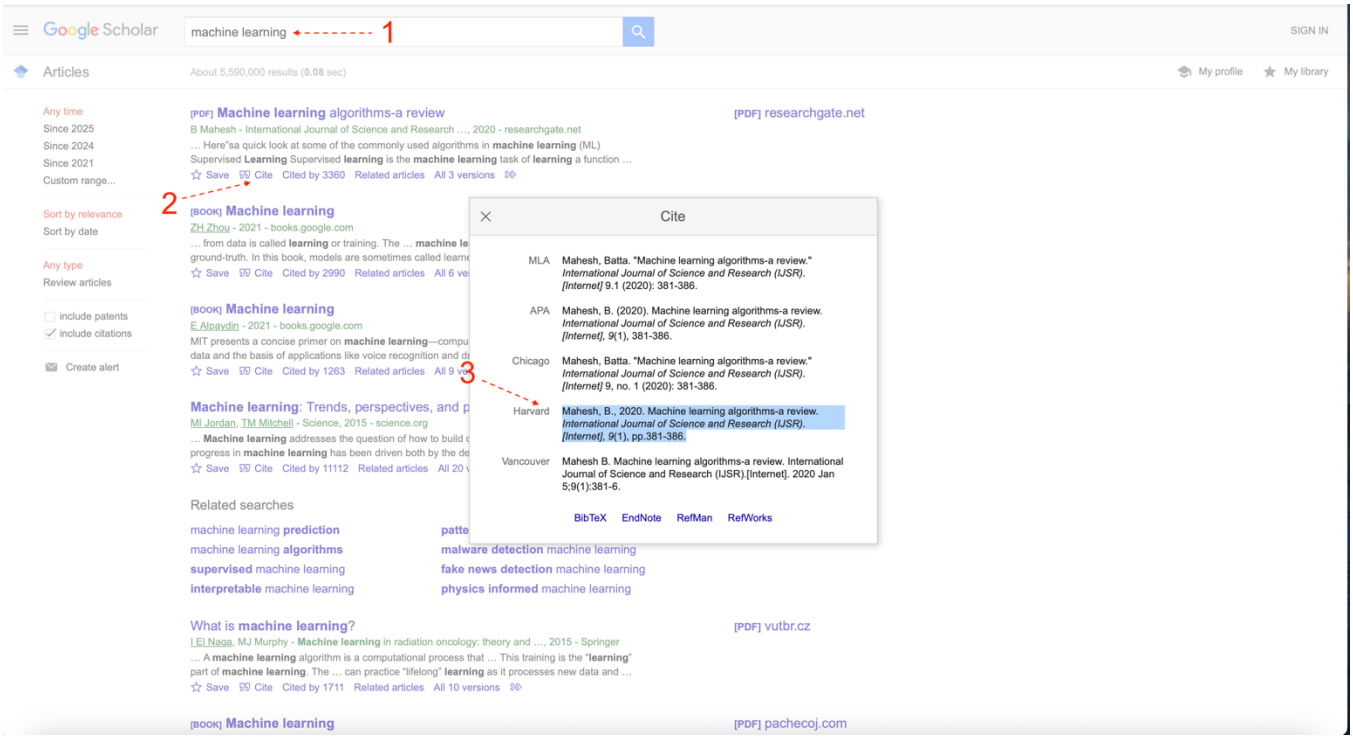
3. Combining Figures and Tables

If you are working on a detailed review article, you might need to **cite multiple figures and tables**. Here's how to do that:

Generative AI models have shown varying performance across different datasets (Table I). Additionally, the architecture design (Fig. 1) plays a crucial role in model robustness and efficiency.

IEEE Style Checklist for Figures and Tables

1. **Figure captions** are **below** the figures, italicized, with the label “Fig.” in bold.
2. **Table captions** are **above** the tables, all uppercase, with the label “TABLE” in bold.
3. All figures and tables are **numbered sequentially** throughout the paper.
4. All figures and tables are **cited in the text** before they appear.



Important Note

- If you have any questions, please **direct them to me**. I created these assignments and will be grading them myself. Teaching assistants will not be able to answer questions related to these assignments.

Summary

- You are encouraged to use generative AI tools to gather information.
- Focus your effort primarily on creating **informative, high-resolution figures and diagrams**.
- Tables, equations, and references can be efficiently handled using generative AI tools as well.

Good luck, and feel free to reach out if you need any guidance or clarification!

Sincerely,

Mohammed Aledhari, Ph.D.
Assistant Professor
Computational Healthcare and Biotechnology Lab's Director
Department of Data Science
College of Information
Department of Computer Science and Engineering
College of Engineering
University of North Texas
Discovery Park, NTRP E236E
3940 N Elm St.
Denton, TX 76207
Phone: 940-565-3641
Fax: 940-369-7600

1. Overview/Summary/Abstract (150 words maximum, unreferenced)

Purpose:
Provide a succinct summary that encapsulates the core aspects of the proposal.

- **Objectives:** To establish comprehensive benchmarking frameworks for evaluating feature engineering and selection techniques in Machine Learning (ML) and Large Language Models (LLMs).
- **Significance:** Current feature selection methodologies lack standardized metrics for robustness, efficiency, and generalization across various domains and architectures.
- **Innovation:** Introducing cross-domain benchmarking protocols, novel evaluation metrics, and open-source tools for feature selection in LLMs.
- **Approach:** Develop, implement, and validate benchmarking frameworks for feature selection across multiple tasks and datasets, with an emphasis on interpretability, robustness, and computational efficiency.
- **Impact:** Enabling systematic assessment of feature selection methods to enhance LLM performance, efficiency, and robustness across diverse domains.

---

2. Introduction (800–1,000 words)

Purpose:
Establish the context, highlight the knowledge gap, and define the objectives.

A. Background and Motivation

- Overview of feature selection techniques in ML and LLMs.
- Growing demand for effective feature selection mechanisms to improve efficiency, robustness, and performance.
- Challenges in evaluating feature selection methods across diverse tasks and architectures.

B. Knowledge Gap

- Lack of standardized benchmarks for feature selection in LLMs.
- Limited understanding of how feature selection affects robustness, generalization, and interpretability.
- Inconsistencies in evaluation methodologies across various research studies.

C. Objectives

- Develop standardized protocols for assessing feature selection methods in ML and LLMs.
- Design benchmarks that evaluate robustness, interpretability, and computational efficiency.
- Create open-source tools for reproducible benchmarking of feature selection algorithms.

---

3. Preliminary Data (500–600 words)

Purpose:
Demonstrate the feasibility and validity of the proposed approach.

- Summary of initial experiments on benchmarking feature selection methods in ML and LLMs.
- Comparative analysis of various feature selection techniques (e.g., L1 regularization, mutual information, principal component analysis).
- Preliminary evaluation of robustness, accuracy, and computational efficiency.
- Figures and tables illustrating comparative performance across different datasets and architectures.
- In this section, you are required to create a table that lists and compares at least 10 available solutions, highlighting their features, weaknesses, and limitations in contrast to your proposed solution and preliminary data/results.

---

4. Significance (200–300 words)

Purpose:
Articulate the broader impact of the proposed research.

- **Scientific Relevance:** Establishing standardized benchmarks will enable systematic evaluation of feature selection techniques, facilitating advancements in ML and LLM performance.
- **Societal Impact:** Enhancing feature selection techniques improves model interpretability, robustness, and efficiency, benefiting critical applications such as healthcare, finance, and natural language processing.
- **Broad Applicability:** Developed benchmarks will provide guidelines for optimizing feature selection methods across diverse domains and architectures.

---

5. Innovation (200–300 words)

Purpose:
Highlight the novelty of the proposed approach.

- **Standardized Evaluation Metrics:** Developing universal metrics for assessing feature selection performance, including accuracy, robustness, interpretability, and computational efficiency.
- **Cross-Domain Benchmarking:** Designing benchmarks that evaluate feature selection methods across multiple domains, such as NLP, computer vision, genomics, and recommendation systems.
- **Comparative Analysis Frameworks:** Establishing protocols for comparing various feature selection techniques in the context of LLMs.
- **Open-Source Tools:** Providing accessible platforms for reproducible benchmarking and dissemination of best practices.

---

6. Approach (1,800–2,200 words)

Purpose:
Provide a detailed description of the proposed work, structured into Specific Aims.

Specific Aim 1: Developing Standardized Evaluation Metrics

- **Methods:** Create metrics that assess feature selection based on accuracy, robustness, interpretability, and computational efficiency.
- **Expected Outcomes:** Comprehensive metrics for comparing feature selection techniques across various tasks and architectures.
- **Evaluation Method:** Apply metrics to diverse datasets (NLP, computer vision, genomics) and compare performance.
- **Potential Pitfalls and Alternatives:** Address overemphasis on accuracy by incorporating robustness and interpretability metrics.

Specific Aim 2: Designing Cross-Domain Benchmarking Frameworks

- **Methods:** Develop protocols for evaluating feature selection techniques across multiple domains, including NLP, computer vision, and recommendation systems.

- **Expected Outcomes:** Robust benchmarks that provide insights into the generalization capabilities of feature selection methods.
- **Evaluation Method:** Compare performance across different datasets and architectures using standardized metrics.
- **Potential Pitfalls and Alternatives:** Address inconsistencies in benchmarking protocols by establishing clear guidelines.

Specific Aim 3: Creating Open-Source Tools for Reproducible Benchmarking

- **Methods:** Develop a user-friendly, open-source platform for evaluating and comparing feature selection techniques.
- **Expected Outcomes:** A reproducible framework accessible to the wider research community.
- **Evaluation Method:** Validate the platform through collaboration with external researchers and evaluation on diverse tasks.
- **Potential Pitfalls and Alternatives:** Address challenges related to scalability and interoperability.

Specific Aim 4: Establishing Guidelines for Practical Implementation

- **Methods:** Provide protocols for integrating feature selection techniques into real-world applications.
- **Expected Outcomes:** Comprehensive guidelines for deploying feature selection algorithms in practical scenarios.
- **Evaluation Method:** Assess the impact of developed techniques on industry-relevant tasks.
- **Potential Pitfalls and Alternatives:** Address discrepancies between academic benchmarks and industry requirements.

---

7. References (Up to 100 citations)

Purpose:

Provide a comprehensive list of relevant literature.

- **Foundational Works (10–15 references):** Seminal papers on feature selection techniques, benchmarking protocols, and LLM architectures.
- **Recent Advances (60–70 references):** State-of-the-art research on feature selection in ML and LLMs.
- **Contrasting Studies (5–10 references):** Highlighting limitations and gaps in current approaches.

---

8. Additional Requirements

- **Figures and Tables:** Include diagrams illustrating benchmarking frameworks, comparative analysis protocols, and performance metrics.
- **Formatting:** Adhere to agency-specific guidelines for structure, citation style, and layout.