UNIVERSITY OF CAPE COAST
COLLEGE OF HUMANITIES AND LEGAL STUDIES
SCHOOL OF ECONOMICS
DEPARTMENT OF DATA SCIENCE AND ECONOMIC POLICY
TERM PAPER 2024/2025
DATA CURATION AND MANAGEMENT PLANS

ATTEMPT ALL QUESTIONS

1. Explain how metadata and data preprocessing can work together to enhance the efficiency of data curation and management. Provide real-world examples to support your explanation.

2. Identify two global open data sources and describe how data can be accessed from each. What are the benefits and challenges of using open data in research and data-driven decision-making?

3. Discuss the importance of data preprocessing in data warehousing. Outline a step-by-step advocacy plan for an organization focusing on "data piling" without proper preprocessing techniques.

4. Using the article "A Survey of Large Language Models" by Zhao et al. (2023) [*], discuss the evolution of language models from statistical methods to large-scale neural models. Explain the importance of pre-trained language models (PLMs) and how these advancements will impact the field of data curation and management plans.

Submission Details

Deadline: Submit by 20th October 2024.

Mode: Add a folder to your repository on GitHub named "Term Paper" and commit the submission to the folder.

Article Reference

[*]Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*.