Plan

1) Data Management Plan

1.1 Public Good

This study examines whether Exeter students' use of AI tools (e.g., ChatGPT) relates to academic performance. Insights can guide course policy, assessment design, and the responsible use of AI in learning. Representing a clear educational and societal benefit.

1.2 Data Description

Design: short online survey (≤5 minutes), anonymised, with no identifiers.
Planned sample: approximately 300 students (simulation uses N = 300).
Variables (aligned with code):

- Performance: Average_Mark (0–100, with ~3% MCAR missingness introduced), Coursework_Mark, Exam_Mark (0–100).
- AI use: AI_Frequency (Never / Monthly / Weekly / 2–3× per week / Daily), AI_Level (minutes per session, numeric).
- Study context (confounders): Study_Hours (weekly), Attendance (%), Sleep_Quality (1–5), Prior_Attainment (band).

Data will be stored in a CSV file (*simulated_exeter_ai_study.csv*) with accompanying documentation:

- README.txt — overview of collection and preprocessing steps.
- data_dictionary.csv — variable names, types, and permitted values.

1.3 Collection

To ensure the dataset meets both ethical and statistical quality standards, data will be collected using Microsoft Forms or Qualtrics via University mailing lists, societies, and seminars. Participation is voluntary, open to students aged 18 or above. A pilot with five participants will confirm question clarity, timing, and accessibility. Questions are single-click or Likert-based to reduce cognitive burden. Guidance follows the UK Data Service (2024) recommendations for research data management.

1.4 Storage and Security

Working copies will be stored on University OneDrive/SharePoint with multi-factor authentication and restricted access. Local files will be encrypted using AES-256 and deleted after analysis. In handling data, the project will follow the UK GDPR, Data Protection Act 2018, and UK Data Service (2024) guidance on secure storage and access management.

1.5 Documentation, Versioning and FAIR Principles

File naming will follow the pattern *ai_exeter_vN.csv*. Version control will be managed through a private GitHub repository. Metadata and provenance documentation will ensure reproducibility. In line with the FAIR principles (Wilkinson et al., 2016), data will be made *Findable* through descriptive metadata and standardised file names, *Accessible* via secure University repositories, *Interoperable* by using open formats (CSV, TXT), and *Reusable* through detailed documentation.

1.6 Preservation

The anonymised dataset, Jupyter notebooks (*DataGenerator.ipynb*, *DataAnalysis.ipynb*), figures, and metadata will be archived within the ELE submission repository for at least 12 months to ensure reproducibility and align with UK Data Service (2024) preservation standards.

1.7 Sharing and Licensing

The de-identified dataset and accompanying code will be shared in the coursework ZIP file under a CC BY-NC 4.0 licence. No personally identifiable data will be distributed. Data sharing adheres to UK Data Service (2024) tiered-access principles.

---

2) Ethics and Privacy Review

2.1 Ethical Framework and Risk

This project poses low ethical risk. It collects no special-category data and ensures full anonymity. It complies with the University of Exeter Research Ethics Policy and the UK Data Service (2024) framework for ethical and legal data management. This proposal also aligns with the University of Exeter's Education and Digital Strategy (2024) for the responsible integration of AI in higher education.

2.2 Consent and Participant Rights

Participants will view an information sheet outlining purpose, anonymity, and withdrawal rights. Consent will be captured via a checkbox confirming understanding and agreement. This satisfies GDPR's requirement for consent to be freely given, specific, informed, and unambiguous.

2.3 Anonymisation and Data Minimisation

No names, emails, IDs, or IP addresses will be collected. Grades are recorded as continuous 0–100 values but only reported in bands to prevent re-identification. Demographic details are restricted to broad categories (programme, year). Data collection follows UK Data Service (2024) recommendations for data minimisation and anonymisation.

2.4 Legal Compliance and Safeguards

Lawful basis: informed consent.
Data security: encryption, access control, and removal of small subgroups from reports.
Bias mitigation: descriptive summaries will test representativeness; potential sampling bias acknowledged.

2.5 Transparency and Accountability

All processing and analysis steps will be documented within Jupyter notebooks using deterministic random seeds for reproducibility. AI tool prompts will be listed in *Prompts.pdf* following the module's Generative AI policy.

3) Analysis Plan (Aligned with Notebook Implementation)

Exploratory Data Analysis (EDA)
- Summarise variables (mean, median, variance, IQR).
- Detect outliers using IQR; visualise distributions.
- Handle missingness (~3% MCAR) using complete-case analysis (*df_clean*).

Statistical Analyses
1. Group differences: One-way ANOVA testing *Average_Mark* across *AI_Frequency* groups (Never–Daily). Post-hoc pairwise tests (Holm correction) will be applied if significant.
2. Non-linearity: OLS regression with quadratic term:
   Average_Mark ~ AI_Level + AI_Level² + Study_Hours + Attendance + Prior_Attainment + Sleep_Quality.
3. Assessment type comparison: Compare *Coursework_Mark* vs *Exam_Mark* across AI usage categories using grouped bar charts with 95% confidence intervals.
4. Correlations: Pearson and Spearman coefficients among study variables, visualised via a heatmap.

These analyses will test whether moderate AI use is associated with improved performance while accounting for study habits and prior attainment, providing evidence of both correlation and practical significance.

Visualisations (Four Figures)
1. Box/violin plot — *Average_Mark by AI_Frequency*.
2. Scatter plot with quadratic fit — *AI_Level vs Average_Mark*.
3. Grouped bar chart — *Coursework vs Exam marks by AI_Frequency*.
4. Correlation heatmap — study and performance variables.

All plots will follow accessibility and accuracy principles from ONS (2023) and Wilke (2019). Figures will be exported to a relative directory (*./figures/*) for reproducibility.

References
- **UK Data Service (2024).** *Research Data Management – Learning Hub.* **Available at: https://ukdataservice.ac.uk/learning-hub/research-data-management/**
- **UK Data Service (2024).** *Ethical Obligations for Researchers.* **Available at: https://ukdataservice.ac.uk/learning-hub/research-data-management/ethical-issues/ethical-obligations/**
- **University of Exeter (2023).** *Research Ethics Policy.* **Available at: https://www.exeter.ac.uk/research/integrity/ethics/**
- **Wilkinson, M.D. et al. (2016).** *The FAIR Guiding Principles for Scientific Data Management and Stewardship. Scientific Data*, **3(1).**
- **Wilke, C.O. (2019).** *Fundamentals of Data Visualisation.* **O'Reilly Media.**