



Individual Project — 2025

LLMs and DSLs

1 Introduction

Recent years have seen a huge amount of innovation in the form of Large Language Models and the use of these models for performing a wide variety of tasks. Domain-Specific Languages are “little” programming languages which target a particular domain to solve very specific problems [1]. Examples of DSLs include: regular expressions, SQL, and spreadsheet formulas.

Your task for this project is to investigate an aspect of LLM use in the context of *one* of the above-listed DSLs.

2 Requirements

You will need to conduct a small experiment in order to evaluate an aspect of LLM use with respect to the chosen DSL. You may choose to investigate any aspect that you like and you may use any LLM or LLMs that you want. For example, you could investigate the correctness of LLM-generated DSL code. Importantly, you may not *directly repeat* an existing experiment (by using the same dataset, the same LLM and the same experimental method). You can, however, use an existing dataset for a different purpose or use existing approach with different data or with a different LLM.

You will need to either find, and reference, an existing dataset, or you can create your own. If you create or derive your own dataset then this must be submitted separately. You must list the prompts given to the LLM/s and the corresponding code that was generated. Importantly, you need to consider how you evaluate the output generated by the LLM/s.

In particular, you must

1. Clearly explain which LLM aspect you are exploring and motivate why.
2. Provide a relevant background and literature review, and position your work with respect to existing work.
3. Describe the experimental method that was adopted.
4. Give your findings, and discuss and evaluate them.
5. Consider the internal and external validity [2] of your study.

3 Deliverables

This is an individual assignment — each student is required to work *entirely on their own*.

3.1 Project Report

Submit via Ulwazi a technical paper of no more than six (double column) pages in length, excluding appendices. The report must address the requirements as per [Section 2](#).

Your report must include the following appendices:

- a screenshot illustrating the input of one prompt to the LLM being used as well as the LLMs output,
- files or listings containing any additional scripts/programs that are used in analysing the generated output, and
- sample screenshots or file logs of the output produced by any analysis scripts.

You also need to upload a file containing a complete listing of all prompts used in the study as well as the corresponding output(s).

4 Deadline and Submissions

The submission date for all deliverables is specified in the “4th Year Submission Dates 2025” document.

All submissions must be in strict accordance with the guidelines contained in the *School’s Blue Book* and the rules contained in the *School’s Red Book*.

5 Assessment

5.1 Criteria

Your report, and other deliverables be assessed according to a rubric which will accompany this brief. This rubric will be made available on the course website.

5.2 LLM Report Writing Policy

All sources that you cite should be:

- credible,
- original,
- and verifiable.

LLMs such as ChatGPT fail on all three counts and must not be cited — the original sources should be cited instead. Note, this does not mean that ChatGPT and other LLMs, such as paraphrasing LLMs, cannot be used; rather, they should be used judiciously and acknowledged in a separate acknowledgement section in an appendix. Ultimately, you take full responsibility for the content that you submit.

5.3 Plagiarism

Refer to the *School’s Blue Book* for an explanation of what plagiarism is and how to avoid it. All instances of plagiarism from either the internet or within the class will be severely dealt with. No two students may have identical or overly similar reports.

References

- [1] M. Fowler, “Domain-specific languages guide,” <https://martinfowler.com/dsl.html>, 2019.
- [2] S. Easterbrook, J. Singer, M.-A. Storey, and D. Damian, *Guide to Advanced Empirical Software Engineering*. Springer, 2008, ch. 11: Selecting Empirical Methods for Software Engineering Research, pp. 285–311.



INDIVIDUAL PROJECT ASSESSMENT FORM

v 1.1

	Unacceptable	Poor	Acceptable	Good	Excellent
Background and Literature Survey	<ul style="list-style-type: none"> research background is extremely poor to non-existent 	<ul style="list-style-type: none"> research background is inadequate or lacks sufficient detail literature referenced largely lacks relevance and/or sources are mostly of poor quality an inadequate number of sources are cited 	<ul style="list-style-type: none"> research background is adequate, and has sufficient level of detail fairly relevant/appropriate literature and/or prior solutions are referenced the quality/number of citations is reasonable 	<ul style="list-style-type: none"> research background is detailed mostly relevant/appropriate literature and/or prior solutions are summarised, organised and interpreted a large number of predominantly high quality, up to date sources, are cited 	<ul style="list-style-type: none"> research background is extensive, in-depth, and captures the nuances of the issue at hand relevant and appropriate literature/prior solutions are very well summarised, organised and interpreted a large number of predominantly high quality, up to date sources, are cited
Research Question / Problem / Hypothesis	<ul style="list-style-type: none"> is not explicitly stated, or is minimally stated, so that it lacks real meaning 	<ul style="list-style-type: none"> is stated in a confusing manner or lacks sufficient clarity is far too broad has little consideration of context not linked to literature survey lacks motivation 	<ul style="list-style-type: none"> is reasonably clear and understandable is reasonably focused the scope is made explicit informed by literature survey adequately motivated 	<ul style="list-style-type: none"> is very clear and understandable is focused and specific the scope is carefully considered informed by, and explicitly linked to, literature survey findings well motivated and clearly identifies contribution 	<ul style="list-style-type: none"> is exceptionally clear, understandable and well articulated is focused and specific the scope is extremely well considered strongly informed by, and explicitly linked to, literature survey findings motivation is excellent, contribution clearly identified
Research Methodology / Approach	<ul style="list-style-type: none"> is not, or is, hardly described cannot answer the research question scope of work done is completely insufficient 	<ul style="list-style-type: none"> lacks sufficient detail to repeat methodologically unsound problematic in how it addresses the research question scope of work done is insufficient 	<ul style="list-style-type: none"> mostly detailed enough to repeat - reasonable discussion of datasets, methodology and data analysis methods reasonable methodological choices is appropriate for answering the research question scope of work done is sufficient 	<ul style="list-style-type: none"> repeatable: described in detail, the datasets, methodology, data analysis methods, etc. are well described reasonable methodological choices is appropriate for answering the research question scope of work done is more than sufficient 	<ul style="list-style-type: none"> repeatable: comprehensively described in all aspects: datasets, methodology, data analysis methods, etc. good methodological choices is appropriate for answering the research question scope of work done greatly exceeds expectations
Research Findings	<ul style="list-style-type: none"> research is trivial data is mostly questionable conclusions are entirely unsupported 	<ul style="list-style-type: none"> research has little meaning: lacks depth / findings restate well-known facts / deals with a trivial problem data of questionable quality gathered conclusions are weakly supported by the data 	<ul style="list-style-type: none"> research is meaningful, deals adequately, and in appropriate depth, with a suitable problem adequate presentation of findings some visualisations reasonable interpretations of results and conclusions 	<ul style="list-style-type: none"> research is interesting and meaningful presentation of findings is fairly comprehensive good visualisations strong, well-supported interpretation of results and conclusions validation of findings is adequate 	<ul style="list-style-type: none"> research is interesting and meaningful extensive and comprehensive presentation of findings excellent visualisations strong, well-supported interpretation of results and conclusions validation of findings is strongly argued
Technical Communication (penalty only)	<p style="color: red; text-align: center;">-20%</p> <ul style="list-style-type: none"> report deviates significantly from the school's standards use of language, style or tone is unacceptable 	<p style="color: red; text-align: center;">-5%</p> <ul style="list-style-type: none"> report does not conform to the school's standards use of language, style or tone is quite poor poor abstract/report structure incorrect citation practices two or more references are missing required details 	<p style="font-size: 2em; font-weight: bold;">0%</p>		

Notes:
 All categories are weighted equally in calculating the overall mark, except for Technical Communication which represents a percentage-point penalty (given in red) applied to the overall mark.
 If any category receives a rating of "unacceptable" then the student's mark is capped at 40%