

Assignment 1: Project Primer

1 Instructions

Three crucial (yet sometimes overlooked) aspects of scoping a project involve **formulating a clear and well-motivated problem, understanding prior work on this problem, and accurately evaluating the effectiveness of possible solutions**. This semester, let us not make this mistake. Through this assignment, you will practice developing a problem that you could use as the starting point (though you are free to change your project direction) for your term project by preparing a 5-slide presentation in Quarto markdown.

1.1 Content Requirements

Before you begin, please choose one of the three project tracks you intend to explore in your submission.

1. **(Slide 1)** A concrete problem on which you can make meaningful progress through the course of the semester. (one resource on defining a good problem is the [Heilmeier Catechism](#) that DARPA uses)
 - a. Ensure that it is **well-defined**. Your problem statement should be clear and specific, avoiding vague language. It should be easy to understand what the problem is and why it matters.
 - b. Ensure that it is **insightful**. Your project should be a meaningful step toward a larger *interesting* and *useful* goal. Note it may be a very small step towards this goal but it needs a larger goal that is clear. Answer the question: “Who cares? If the long-term goal is successful, what difference will it make?”
 - c. Ensure that it is **feasible** – it should not require more time or compute than reasonable within the course of a semester.

2. (Slides 2-4) Three research papers (must be from a **peer-reviewed venue** or posted on **arXiv**).
 - a. Guidelines for paper selection
 - i. Well-known AI/ML peer-reviewed conference venues include: NeurIPS, ICML, ICLR, AAAI, CVPR, IJCAI, AISTATS, UAI EMNLP, ICCV, CoRL.
 - ii. *Note: Papers on arXiv have not been peer-reviewed but most frontier research is published on arXiv. Many articles on arXiv are published in top-tier conferences so see if there is a published version if possible.*
 - iii. You should not cite websites or news articles.
 - iv. You may present **at most one survey paper**. Survey papers can provide helpful exposure to current methods and challenges.
 - b. For each paper, you **must**:
 - i. Briefly summarize the main contribution of the paper.
 - ii. Briefly explain how it specifically relates to your project.
3. (Slide 5) You must propose a metric that objectively and accurately evaluates approaches to your proposed problem. You must do the following three things:
 - a. **Propose a Metric:** Your proposed metric must be **computable** and **quantifiable**. It should be concrete and specific rather than vague.
 - *Multiple Metrics:* Sometimes a single metric is not enough. For example, a solution may need to be evaluated on both accuracy and fairness, or accuracy and latency, as there is often a tradeoff.
 - b. **Justify the Metric:** Performing well on your method of evaluation must imply that the problem is solved effectively. Conversely, a strong solution to the problem must not fare poorly on your evaluation metric.
 - c. **Contextualize the Metric:** You must define and justify what ranges of your metric constitute “**Poor**,” “**Okay**,” “**Good**,” and “**Excellent**” performance in the context of your problem.
 - For example, 51% accuracy is excellent for predicting stock price movements, but it would be poor for diagnosing a disease where a 90% accurate baseline already exists.
 - *Bounds:* While not required, providing upper or lower bounds for your metric (e.g., based on theory, a random baseline, or an oracle model) can help contextualize your results.

1.1.1 Track-Specific Examples & Guidance

- For TinyReproductions:

- **Problem:** The problem is to demonstrate a key insight from a specific research paper, not to solve a new problem. For example, the problem could be to show that a specific loss function leads to more stable training than an earlier method.
 - **Related Work:** Your literature review should focus on the single paper you intend to reproduce, its essential baselines, or a relevant survey paper that puts its claims into context.
 - **Evaluation:** The metric is a measure of how well your reproduction matches the original paper’s findings. You can use quantitative metrics like percentage change in performance or qualitative metrics like the visual similarity of plots.
- **For ProductPrototype:**
 - **Problem:** The problem is to solve a real-world issue for a target user or client. An example problem is automating the classification of GitHub issues for a software team.
 - **Related Work:** Your literature review should explore and justify the choice of AI/ML techniques and tools that could solve your problem. For instance, you could compare different time-series models for a load forecasting tool.
 - **Evaluation:** The metric should be tied to user needs and the effectiveness of the prototype, such as classification accuracy, inference latency, or other relevant software engineering metrics.
 - **For NovelResearch:**
 - **Problem:** The problem is to identify a gap in existing research and formulate a novel hypothesis to address it. An example problem could be a systematic analysis of reasoning failures in large language models.
 - **Related Work:** Your literature review should target a specific research area to clearly identify a gap or limitation in prior work that your project will address.
 - **Evaluation:** The metric should be a rigorous, quantitative measure of your proposed method’s effectiveness, which can be compared to strong baselines in the field.

1.2 Format

Plain text in [Quarto RevealJS markdown format](#) **without the YAML header** containing exactly five slides with a very simple example below. You can use more advanced RevealJS markdown like two column layouts if helpful. Any images used must have absolute URLs so that I can copy and paste into my lecture slides.

```
## Slide Title 1
- Point 1
  - Subpoint 1.1
- Point 2

## Slide Title 2
[Slide Content]

## Slide Title 3
[Slide Content]

## Slide Title 4
[Slide Content]

## Slide Title 5
[Slide Content]
```

1.3 Character Limit

7500 (Any content after the character limit will be truncated during grading)

2 Rubric for Grading

Criterion	Excellent (5)	Good (4)	Satisfactory (3)	Okay (2)	Poor (1)
Problem Formula-tion	The problem is exceptionally clear, insightful, and feasible. It represents a creative and meaningful step toward a larger, interesting research goal.	The problem is well-defined, insightful, and feasible. It is a reasonable step toward a known research goal.	The problem is proposed but may lack some clarity or insight. It might be a minor extension of a well-known problem.	The problem is too broad, unfeasible, or lacks a clear purpose. It may have been well-studied with little room for new insight.	The problem is trivial, nonsensical, or not proposed at all.
Literature Review & Synthesis	Cites three appropriate peer-reviewed or arXiv papers. Demonstrates an excellent, intuitive, and concise understanding of each paper's contribution and its specific relevance to the problem.	Cites three appropriate papers. The summaries of the papers' contributions and their relevance to the project are solid and well-reasoned.	Cites the required number of papers, but some may not be from appropriate venues. The analysis is present but lacks depth, or the connection to the problem is unclear for one or more papers.	The literature review is superficial or the analysis is weak. Papers may be missing or from inappropriate sources like websites.	There is no meaningful analysis. The cited papers are unrelated to the problem, or no papers are cited.

Evaluation Plan	Proposes a computable, quantifiable, and well-justified metric uniquely suited to the problem. The performance tiers (“Poor,” “Okay,” “Good,” “Excellent”) are thoughtfully defined and justified.	Proposes a computable and appropriate metric. The performance tiers are defined with reasonable justification, and the connection between the metric and solving the problem is clear.	A metric is proposed and is generally accurate but may not be well-justified. The performance tiers are present but not well-justified.	The proposed metric is vaguely defined, lacks justification, or the required performance tiers are missing.	The metric is inaccurate, not quantifiable, or is not proposed at all.
------------------------	--	--	---	---	--