Frugal AI - Project guidelines

Eva Feillet

October 2025

The goal of the project is to explore an active research topic connected to frugal AI.

1 Deliverables

⋆ a short report summarizing your readings and discussions (3-5 pages)

⋆ a minimal code base (preferably shared via a `git` repository)

The content of the report reflects the main steps of the project:

— Introduction : framing the topic, context, motivation, one-sentence research question. Define the task to solve (inputs/outputs).

— Related work : (short) what is the state of the art, short summary of your readings.

— Evaluation : How to benchmark the task ? Describe how to proceed. What is feasible on a personal computer or free server ? Define relevant constraints for the evaluation (compute, data, latency, energy).

— Baseline : basic method to address the problem (or popular legacy method). Explain the principle. Discuss its advantages and limitations.

— Method : Either (i) focus on a particular method of the state of the art, explain its principle, OR (ii) explain how a new method could be like.

— Implementation : minimal working example to showcase the task, the problem, and a baseline method. Depending on your choices, also evaluate an alternative method.

— Discussion and perspectives : with more time, how would you improve the implementation ? the method ?

2 Suggested step-by-step schedule

NB: this suggested schedule is generic and suited to a longer project. **In this activity, your goal is to understand how to set the basis for a real research & development project.** It is expected that you don't have the time to implement or test everything.

2.1 Framing the topic

Objective. Select a narrowly scoped task and define relevant constraints (compute, data, latency, energy).
Actions.

1. **Choose a task** based on first readings.

2. **Set constraints relevant to your chosen frugal AI topic**

3. **Define the problem statement:** inputs/outputs, and a one-sentence research question.

Outputs. 1–2 paragraph topic brief; list of constraints; initial bibliography (3–5 seed papers/blogs).

2.2 Identifying research questions

Objective. Turn the topic into testable hypotheses.
Actions.

1. **Draft 2–3 hypotheses:** e.g., "A distilled model at 30% params meets latency target with $< 2\%$ accuracy drop."

2. **Identify relevant metrics:** task metrics (e.g. accuracy/F1/BLEU/...), *and* efficiency metrics (e.g. latency, peak RAM/VRAM, model size on disk, energy/CO_2 estimate).

3. **Define success criteria for a new method** a small Pareto improvement (accuracy vs. cost) or meeting a fixed budget with minimal loss.

Outputs. Hypotheses list; metric definitions; success criteria. NB: you won't have time to test all the hypotheses.

2.3 Expectation vs Reality in the State of the Art

Objective. Contrast claims with what is feasible under your constraints.
Actions.

1. **Triage the literature:** e.g. create a 1-page matrix (method, idea, claimed gains, compute/data preconditions, code availability).

2. **Pick a *baseline*:** a simple, strong, reproducible method (e.g., uncompressed model, straightforward training, or a popular legacy approach).

3. **Select one SoTA/modern method *or* propose a new lightweight variant:** easier if an open-source code exists.

4. **Repro readiness check:** can you run it on your hardware? what to simplify (smaller backbone, fewer steps, subset of data)?

Outputs. Related-work summary (half page); scope note highlighting any gap between paper claims and what you can really use to reproduce the experiments.

2.4 Experimental evaluation

Objective. Design a minimal benchmark plan aligned with frugality / the constraints you identified.
Actions.

1. **Datasets:** choose small/medium public sets; predefine tiny dev/test splits; document preprocessing and licensing.

2. **Metrics:** for reporting both *effectiveness* and *efficiency*. What would you measure ?

3. **Protocol.** What to test ? What would you need for ablation ? e.g. ablation for one key hyperparameter. Baselines, method ?

Outputs. Evaluation plan. Credible outline.

2.5 Prototyping

Objective. a minimal, reproducible codebase.
Actions. Either (A) or (B)

1. (A) your own implementation (focus on the set-up and on a baseline method)

2. (B) reproducing experiments from a public repo (focus on comparing a baseline with an alternative method)

Outputs. Public git repo; README file with setup commands and an example run that finishes in < 10 minutes on a personal computer.

2.6 Take-aways

Objective. Synthesize findings, limits, and next steps.
Actions.

1. **Answer the research question(s):** include both task and efficiency outcomes.

2. **Discuss advantages and limitations**. With more time or resources, what would you do differently? How would you change the evaluation protocol? How would you improve the method?

Outputs. Final report (3–5 pages) + minimal code.