Master's Thesis in ML / NLP

The one-page version — contribution, evidence, writing, what it becomes

Konstantinos Skianis
Assistant Professor of NLP · Dept. of Computer Science & Engineering
University of Ioannina · y3nk0.github.io

THE BAR

Between an undergraduate project and a PhD chapter: a **contribution**, but a bounded one, finished in two semesters. Expected beyond the undergrad standard: a *positioned* related-work chapter, ablations, several seeds, significance where claimed, and an output that could become a workshop paper with a month's work.

> **The one-claim test.** *For [setting], [what you did] achieves [effect] compared with [baseline], because [mechanism] — shown by [the experiment].* Cannot fill every slot by month two? You have a project, not a thesis. The empty slot is your next task — usually *because* (numbers, no explanation) or *compared with* (no honest baseline).

FINDING A CONTRIBUTION

Pick a question that is **open**, **answerable with your compute**, and **interesting whichever way it falls** — then a failure is still a finding.

1 Extend one paper Most reliable. Push *one* dimension: new language or domain, distribution shift, a tenth of the compute or labels, longer inputs, different evaluation. The contribution is where it breaks and why.

2 Reproduce and interrogate Rebuild a published result, then ask what the authors did not: other seeds, datasets, scales, prompts.

3 Shared task with an angle SemEval and workshop tasks give data, metric, leaderboard. The thesis is the analysis of *why* systems place where they do.

4 Applied / industry A real problem — read the NDA column first.

5 Analysis or theory Probing, robustness, efficiency measured honestly, a metric critique with a small human evaluation, a bounded proof.

Where gaps are written down Limitations sections · task overview papers · survey taxonomies (thin boxes) · workshop CFPs · plateaued leaderboards · **your supervisor's grant** (funded, data exists, attentive supervision).

TWO-PAGE PROPOSAL, BEFORE ANY CODE

Question · **claim** (the sentence above) · why now (2–3 refs) · data (name, size, licence, in hand?) · baselines (trivial + published + source of its numbers) · method (3 bullets + the knob you vary) · evaluation (metric, splits, seeds, significance test) · **compute budget** · milestones (data, baseline, first result, freeze, draft) · risks · **fallback** (the thesis you still have if the idea fails).

> **Scope sinks theses, not difficulty.** Can you name the dataset and metric today? Run one baseline end to end in three weeks? If the idea fails completely, is there still a thesis? Any "no" — narrow it now.

INDUSTRY / NDA THESES

Settle **in writing, before you start**, with your academic supervisor in the room: can the thesis be published and defended openly · is there an embargo (6–12 months is common) · what may appear (method yes; raw data often no; relative results usually negotiable when absolutes are not) · who owns code and IP, and what may go on GitHub · who supervises what (degree requirements win) · **fallback if access is withdrawn**.

Writing around confidential data Describe data by statistics and structure, not content, and reproduce the core experiment on a public dataset so an examiner can check your reasoning. Do it from the start. GDPR and ethics approval: ask in week one, it takes weeks.

RELATED WORK: THE CHAPTER THAT SEPARATES DEGREES

Undergrad summarises; master's **positions**. By theme, not by paper, with a taxonomy figure, ending where nobody has been.

- 5–10 anchor papers from your supervisor; steal their vocabulary.
- Same queries across Scholar, Semantic Scholar, Anthology, OpenReview, DBLP — **log query, date, hits**.
- Snowball backward (references) + forward (cited-by, *search within citing articles*).
- Screen 200–400, keep 60–100; extract into a table.
- Cluster rows until groups stop changing = your taxonomy.

Read in 3 passes: 5 min decide → 1 h figures+method → hours for the five papers the thesis depends on.

TOOLS, WEEK ONE

Zotero + connector + Better BibTeX (pinned keys, auto-export to `refs.bib`) · fastest entry: Scholar *Settings* → *Bibliography manager* → BibTeX links, then **Cite** → **BibTeX** — **check it**; cleaner from Anthology / DBLP or `curl -LH "Accept: application/x-bibtex" https://doi.org/<DOI>` · Overleaf + `git/latexmk` locally, one sentence per line, `latexdiff` for drafts (**Prism closes 1 Jan 2027**) · one git repo, commit daily · compute sorted in week one · track every run (config, metrics, commit).

EVIDENCE THAT SURVIVES AN EXAMINER

1. Read 50 examples by hand; look for leakage between splits.
2. **Trivial baseline** first — it anchors everything.
3. **Reproduce the published baseline**; if you cannot, finding out why is itself a result.
4. Your method, one change at a time, each measured.
5. **Ablate** — start early, they always take longer.
6. Error analysis, 50–100 cases by hand = your *because*.

Claiming a difference 3+ seeds, **mean ± std** · **paired bootstrap or permutation test** on test predictions · many systems on one dataset inflates false positives — say so · report compute and the size of the hyperparameter search (a win after 100 trials against a baseline tuned once is not a win) · splits fixed before you look; test set touched once.

> **Costs a grade:** leakage · unfair tuning · one seed · accuracy on imbalanced data · a method chapter nobody could reimplement · a related-work chapter that is a list · writing started in the last month · results you cannot regenerate.

Negative results count — if the comparison was fair and you verified your implementation reproduces a known number first.

AI ASSISTANT: FIVE RULES

> **1** Explain every line you submit — the defence tests exactly this. **2** Never cite a paper you have not opened. **3** Never let it write your results or analysis. **4** Follow your university's policy and declare use. **5** Nothing confidential into a cloud tool.

WRITING & SUPERVISION

Start month one. Chapters: abstract 1 · intro 4–6 (contributions as a numbered list) · background 6–12 · related work 8–12 · method 8–12 · experiments 10–18 · analysis 6–10 · conclusion 2–4. Figures and tables first, then the text. Captions self-contained. Limitations honest. Complete draft to your supervisor **three weeks** before the deadline.

Supervision Decide yourself: libraries, code structure, plots · bring to the meeting: anything that changes the claim or scope · **escalate immediately**: lost data or compute, ethics/NDA trouble, a deadline you will miss. Five bullets the evening before:
did / found / blocked / propose / need.

INTO A PAPER

Plan it in month five, not after the defence. Cut background and most related work; keep claim, method, key experiments, analysis. Target a workshop, **Findings**, the Student Research Workshop, or a system paper. Agree authorship early. Submit after the thesis, using examiner comments as free reviews. Release code + checkpoints + `CITATION.cff`.

THE DEFENCE

20–30 min + questions, before a committee that has read unevenly. Problem (2 slides) → why it is hard + what exists (2–3) → **your claim** (1) → method (3–5) → results (4–5) → analysis + limitations (2–3) → conclusion. One idea per slide. Rehearse out loud, timed, twice. Know your numbers: data sizes, your result, the baseline, spread across seeds, training cost. **Expect:** why this baseline? is it significant? what under distribution shift? why does it fail on class X? six more months? "I do not know, but here is how I would find out" is a fine answer; guessing is not.

WHAT IT BECOMES NEXT

A PhD application The thesis is your evidence that you can do research. Give referees the abstract, results and plans — ask while it is fresh. A workshop paper moves you from "promising" to "has published".

A job The repository is the portfolio: a README that reproduces your main table, honest limitations, clean code. Interviews ask what examiners ask.

Both Not exclusive — an industry year before a PhD is a good path.

NINE MONTHS

M1 proposal signed, data in hand, 15–20 papers, trivial baseline. **M2** published baseline reproduced, tracking, taxonomy drafted. **M3–4** method, first results, claim sentence complete. **M5** ablations, seeds, significance, publication plan. **M6–7** error analysis, results frozen, tables by script. **M8 complete draft**, artefacts released. **M9** revisions, citations checked, two rehearsals, submit.