

# Checklist for Thesis & Papers

Roman Kern, 2022-01-28  
Prior to submission of your work

## A. Planning

Actions related to the planning of the research work, typically conducted prior to writing.

### **Research gaps**

Identified gaps in the existing work, which warrant (our) research. The identified gaps are actually existing, significant gaps. (e.g., talk with other ppl. about the perceived gap)

### **Research questions**

Formalized the main research question, which one being associated with the main contribution of the paper.

### **Weight (quality?, impact?, citeability?)**

Have a good understanding of the weight of the contribution. For a thesis, does it justify a Bachelor's or Master's degree? Is it a conference or journal, is it A\*, A, B, workshop? Should it be a short paper, long paper, poster?

### **Motivation**

Have story line that makes it clear why the presented research has been conducted/initiated – distinguish between the motivation to conduct the paper (**author perspective**) vs. the motivation of the readers to pick up the paper (**reader perspective**). Demonstrate the problem's relevance. Note: The motivation might not be explicit in the written text.

### **Contribution**

The main contributions of the work are clearly defined (and justified).

### **Scope**

The scope of the contribution (i.e., where it applies) is clearly defined.

### **Impact**

The impact of the contribution (i.e., what is now possible) is clearly defined.

### **Novelty**

It is clear that the research provides a novel (= trendy, hot topic) topic.

**“Spin” of the paper**

How to sell the paper to the reader/supervisor/reviewers. Tailored towards the target community/audience.

**Outline for work**

The structure of the work has been planned and discussed, which content goes in what section. Storyline and the main arguments. See Easterbrook.

**Peer check**

Read the papers from the research groups/supervisor. Identify **terminology** and **expected structure** of the target community, identify key papers/works to cite.

**Target venue selection (journal, community)**

Where/when to submit the paper.

**Target venue analysis & Peer check**

Read the **call for papers**. Read the papers from the target venue,. Identify **terminology** of the target community, identify key papers (from the venue) to cite. Identify **potential reviewers**. Read the **authors guidelines** (for layout, blinding, ...), which English (British/American) for text & key words. Talk to people (internal), who already submitted to the journal.

## B. Content

### **Introduction Story**

Typically the introduction gives a motivation why the research has been conducted, especially important for novel problem settings.

### **Related work (with storyline)**

Are all relevant prior works being cited in the introduction and background/related work section. The “shortcomings” of prior work are made obvious to the reader and the expectation is clearly set (to close these **research gaps**).

### **Problem definition/setting**

Formally defined (optionally with notation) problem definition that the work tackles.

### **Theoretical framework**

Guiding theoretical underpinning to prevent the impression that just existing Lego bricks have been combined to solve the problem without reasoning (just try out).

### **Problem solution**

Statement that make it clear to the reader, how well the problem has been solved.

### **Running example (optional)**

Single example throughout the work, starting the in the introduction (with a figure).

## C. Evaluation

### □ **Data sets described**

Reader understands the data sets (size, main characteristics and statistics). Optimally with an URL. And optionally also indicate the current SotA on these data sets.

### □ **Variation in the data sets**

The paper should have different data sets, ranging from synthetic data sets to real-world data sets. More is better.

### □ **Evaluation methodology**

Explanation of the type of experiment, what is being done. Also maybe include **expectations**.

### □ **Competitive baselines**

Typically two types of baselines: 1) upper and lower bound, 2) state-of-the-art baseline. Often also a human baseline is interesting, or simple rule-based baselines.

### □ **Results for comparable methods**

Also include results for methods that can be considered to be similar. They provide a point of reference.

### □ **Results beyond SotA**

Highlight which results go beyond SotA. It should be clear why the proposed methods achieved the good results.

### □ **Statistical test**

If it possible to conduct a statistical significance test, then the results should be included. If not, it should also be made clear to the reader.

### □ **Error analysis**

If possible, try to identify cases where your approach did not work and highlight what the (potential) reasons are and how they could be rectified in future work.

### □ **Ablation studies**

If applicable, running the method and selectively taking out parts and observe the behaviour. This gives an insight into the importance of the individual aspects of the method.

## D. Presentation

### I. Language

#### **Consistency**

The whole work should be in one single writing style (e.g., only British English). The terminology should be consistent (i.e., the same term used in the paper), also including the figures. Prevent repetition of e.g., explanations or definitions. Abbreviations should be consistently used (typically defined once per chapter).

#### **Scientific Writing Style**

Conform to the agreed scientific writing style, e.g., we instead of I. Avoid overly complex language. Proper use of hedging (avoid impression of claims that cannot be backed up). All claims need to be backed up, either with an explanation, or a reference.

#### **Proper use of tenses**

The individual sections should be written using the proper tense.

#### **Flow of the Text**

Always refer to what you previously said, i.e. don't include isolated sentences. (E.g. Sentence1: Concept A is novel. Sentence 2: *This* concept is therefore ...)

### II. Layout

#### **Professional layout**

Make it look professional, including proper typesetting and positioning of the tables/images/.... Stick to the template if available. Try not using Microsoft Word/similar word editors.

### III. Notation & Pseudocode

#### **Consistency**

Consistency within the work and across existing literature (e.g., same letters/glyphs for the same concept).

#### **Readability**

Keep it as simple as possible/necessary.

#### **Explanation for symbols**

Each symbol (e.g., variable) that is introduced in a formula needs to be explained.

#### **Pseudocode is self-contained**

All variables need to be initialized and the flow should be clear (i.e., prevent side effects).

**Double-check correctness**

An error in notation alone can be a reason for a reject/revision.

## IV. Figures & Tables

**All text/content in figures should be readable**

The font size of the figures should be large enough to allow easy reading. Charts should be easy for the reader to grasp and their main meaning should be clear.

**Annotation in the figures (optional)**

Provide annotation for figures, e.g., a red line to indicate the baseline performance. Arrows to highlight important aspects.

**Self-contained captions**

The captions should not only summarize the figure/table, but also explain it. Long captions (4-5 sentences) are acceptable (but check with author guidelines). Address all questions a reader might have (e.g., if there is a visible outlier, explain why).

**All axis are named**

All axis of a chart should be named and explained.

**Accessibility**

Make sure your figures/tables can be read by color-blind people.

**Consistent across figures**

If there are multiple figures/charts that report similar aspects, try to align them, e.g., same axis scaling, consistent colours, consistent naming.

## V. References & Citations

**All references match the style of the target venue → consistency**

Depending on the target venue (journal, conference, ...) the style of the references might differ

**All references are complete (meta data)**

Are some information missing in the references, including year, name of the venue, etc.

**All cited papers are actually matching**

Do the cited papers actually back up a claim made in the paper? Are the citations just copied from another source without checking?

**All claims in the manuscript have proper citation**

Are there claims made in the paper that are not properly backed up?

**High quality source references**

Are a significant number of chosen references/sources from critical sources (qualitative high sources)

**References from the target venue/community**

Many of the chosen references/sources should be from the target value – since it is also expected that the potential reviewers are among the cited articles

**References should be recent**

The references should be up-to-date to show that the research is current/novel.

**Cite seminal papers**

For certain tasks there is one “defining” paper that everyone cites, and one should follow this by also including this.

**Check that papers really exist**

All paper need to really exist. Make sure that the citations have a DOI.

**Check for double entries / duplicates**

The reference section should not contain the same paper multiple times.

## VI. Reproducibility

**Are all data sets being available**

Optimally with links to the data sets? (e.g., Zenodo)

**Is the source code available**

Link to the repository and a clear license.

**Information in the paper sufficient for reproduction**

The details given in the paper allow readers to reproduce the results, i.e., critical model parameters are listed, used resources are references (e.g., which stop word list, which library). An appendix (or supplemental material) is often a good place to report such details.

**Script to reproduce exact results**

Is there a script to exactly reproduce the results (and optimally the figures) of the paper?

Are software/package versions listed?

**Include the original source of results**

If one copies the results from another paper, consider to make it clear.