

# Writing Scientific Papers and Reports

Petr Grinberg

*Faculty Of Computer Science, HSE, Moscow*

pgrinberg@hse.ru

*Last Edit Date: October 12, 2024*

**Abstract**—A critical part of scientific discovery is the communication of research findings to peers or the general public. Mastery of the process of scientific communication improves the visibility and impact of research. While this guide is a necessary tool for learning how to write in a manner suitable for publication at a scientific venue, it is by no means sufficient, on its own, to make its reader an accomplished writer. This guide should be a starting point for further development of writing skills.

## ABOUT THIS GUIDE

This paper aims to provide some guidelines on how to write a scientific paper with extra comments concerning Deep Learning (DL) oriented projects. Sources of this paper include materials from different authors (see Section VIII). We advise the reader to also have a look at these sources. The following Sections are organized in the standard order, which you can see in most of the scientific publications. Each Section has corresponding advice, information and examples.

Before you start writing a paper, think about your reader and his/her purpose in reading:

- Who: The reader of your project/lab report might be
  - knowledgeable or expert in your field, and/or,
  - interested but with no deep technical knowledge, and/or,
  - important for future contacts / developments, and/or,
  - a supervisor that needs to assess your work to give you a grade.
- Purpose: The reader is
  - interested in the results and discussion of your work and also in the credibility of your methods, and/or,
  - looking for inspiration for her/his own work, and/or,
  - interested in recruiting you.

Remember that an effective project/lab report begins with pre-project/lab planning. Usually, your supervisor will clarify the purpose of the project/lab and the procedures, but if not, then you need to think this through.

- What do you want to learn?

- What are the variables?
- What are the procedures?
- What materials and facilities will be used?

Remember that an effective project/lab report requires careful in-project/lab procedure and recording data accurately and completely.

Always read the template of the conference/journal you are submitting your paper to. Each template may contain its own rules for formatting your paper. For example, some conferences require you to put caption below the Table and some – above.

## TITLE

The title should be short (about 10 words), interesting, and it should describe the assigned task/project and/or what you found. Be careful and avoid overselling your paper in the title. For example, if your article verifies some hypotheses only for a particular model, do not claim in the title that these hypotheses hold for all systems, as your work does not contain any information about other architectures. This is valid for other sections too: do not claim something without any justification and reasons.

## ABSTRACT

In some but not all cases. Again, learn what your reader expects. The abstract is a very short summary (usually around 150-250 words) of what the question is, what you found, and why it may be important. Be sure to clearly and concisely highlight your main findings and novelty, how your approach differs from / is better than / enhances the existing material reported in the literature.

The importance of abstracts is increasing as more scientists/engineers/industries are using search engines to keep up with the literature. Since search engines can only search for words in a paper title and abstract, these may be the only parts that many people read. Consequently, a well written abstract is extraordinarily important. Think of an abstract as something that sells your paper to the readers, i.e., makes them interested so they want to explore the whole paper.

## I. INTRODUCTION

Introduce what your question is, explain why someone should find this interesting (what is a general problem that task tries to solve and why this problem has to be solved). Summarize what is currently known about the question, i.e., what is current state-of-the-art. Introduce a little of what you found and how you found it. You should explain any ideas or techniques that are necessary for someone to understand your results section but do not go into deep details.

- Context/Purpose/Objective(s) Why are we doing the project/lab? What questions are you trying to answer?
- Hypothesis. Which assumptions do you have? What do you expect the results to be? This should relate directly to the problem.

Sometimes, you may want to highlight your novel contributions again in the end of the paper. Use *itemize* and clearly indicate your main advances. See [1], [2] for the examples.

In the end of the introduction, indicate the order of the following sections to help the reader orientate. Usually, it looks like this: *The rest of the paper is organized as follows. Section II presents the existing solutions. In Section III, we propose our method. .... Finally, we sum up our findings in Section VII.*

If space is very limited, the article structure follows common rules, and the number of pages is not too big, the paragraph about contributions is preferred over the report organization and the latter can be omitted.

If you are publishing the code for your paper, you can indicate it here in the footnote<sup>1</sup>. The link can also be mentioned in the abstract or Section IV. If you want to include it in the abstract, avoid the footnote and write it directly: [https://github.com/Blinorot/pytorch\\_project\\_template/tree/report](https://github.com/Blinorot/pytorch_project_template/tree/report).

## II. RELATED WORK

In this section, introduce current state-of-the-art and additional information required to understand your approach. For the DL-oriented projects, this section may include: (i) baseline architectures and/or discussion on previous approaches for solving your task; (ii) definition/explanation of key DL-techniques and algorithms, which your own method is based on. Do not go into too deep details, as your paper is limited in size and the reader can always look at the corresponding references (do not forget to write citations). If the DL-technique is well-known (e.g. ResBlock [3]), you do not need to explain it, except for the case your paper improves this well-known technique (e.g. you propose a new ResBlock version, Res2Block [4]). Do not include the metrics and numbers here, you have Section V for this purpose. Saying "ModelName achieves state-of-the-art performance on DatasetName" is usually enough.

<sup>1</sup>[https://github.com/Blinorot/pytorch\\_project\\_template/tree/report](https://github.com/Blinorot/pytorch_project_template/tree/report)

Some related work is usually mentioned in the Section I. The "Introduction" is focused on describing the field in general, introducing the problem, listing current solutions, and limitations of existing work that your approach is overcoming (your novel contribution is proposing a method that deals with these limitations). The "Related Work", however, contains more detailed explanation and discussion of previous work. In some cases, you can merge Sections II and I together.

Always remember, that if your paper is not a survey, your aim is to explain your own novel approach and not to rephrase the existing material.

It must be noted that the number of details also depends on the type of your paper. For example, coursework / term / thesis papers usually require more in-depth discussion on existing literature and not strictly limited in maximum size. This is done to assess how well you explored the given problem/task. For conferences and journals, readers usually have enough knowledge in your research area and you may skip some unnecessary details, i.e., it is enough to include only information described at the beginning of this Section.

## III. METHODOLOGY

Sometimes this Section can be joined with the following one. Explain your novel approach here in details. Ideally, the paper should have enough information so the reader can reproduce it without any external source of details (code, for example). If you are limited in space, you may want to put some details in Appendix. A good idea would be to show a graphical Figure summarizing your method. Explain how your method differs from the existing ones (from those in Section II). Try to provide theoretical justification for the superiority of your approach (you may want to do it here or in Section V). If your method is good, but you do not understand why, ask yourself: maybe this good performance is accidental and your method is not that cool. Your paper will most likely be rejected in the absence of such justification. If you cannot come up with a theoretical argument, at least provide comprehensive analysis by conducting experiments on several datasets and doing ablation study.

## IV. EXPERIMENTAL SETUP

This Section and the previous one are like a cooking recipe. These two sections should be written in the past tense, since your experiments are completed at the time you are writing your report<sup>2</sup>. Include enough detail so that someone can repeat the experiment<sup>3</sup>.

Write which dataset are you going to use. Justify your choice: why do you use this dataset and not another one

<sup>2</sup>Actually, the scientific community sometimes does not agree on this point. Some researchers use present tense for describing experiments and results, some – past tense. The main rule is to be consistent. Discuss this with your supervisor.

<sup>3</sup>If space is limited and you provide code to reproduce your experiments, some details can be omitted.

(depending on the case, you may also do this in Sections I or II). The choice might be due to theoretical reasons, standards in the literature, proper comparison with other approaches. You might also include ablation experiments, in which you apply your method on different datasets. If you have enough space, it would be good to describe the dataset: number of objects per split, distribution, etc.

Provide the information about the training / evaluation scheme: hyperparameters, loss-functions, optimizers, number of epochs, batch size, schedulers, etc. Sometimes, these numbers can be written in the Appendix instead. Be sure to check that hyperparameters in your code and in your paper are the same. Information should be clearly written, detailed, and brief. Be careful not to mix results into this section!

If you want to do some specific measurements / experiments / analyses, explain how you do them here in details. For example, look at Sections IV.B and IV.C from [5].

In modern conference and journal rules, reproducibility becomes more and more obligatory. Make your paper as reproducible as possible. In code, try to make all the instructions clear, add Docker, and indicate all packages. In paper, write your hardware, i.e. which GPU and CPU were used.

## V. RESULTS

To write the results section, use the figures and tables as a guide. Start by outlining, in point form, what you found, going slowly through each part of the figures. Then take the points and group them into paragraphs, and finally order the points within each paragraph. Present the data as fully as possible, including stuff that at the moment does not quite make sense. Verbs in the results section are usually in the past tense. Only established scientific knowledge is written about in the present tense, “the world is round”, for example. You cannot presume that your own data are part of the body of established scientific knowledge, and so when you describe your own results, use the past tense, “a band of 1.3 KB was seen”, for example. There are, however, exceptions to this general rule<sup>2</sup>. It is acceptable to say, “Table 3 shows the sizes of the DNA fragments in our preparation”. It is also acceptable to say, “In a 1991 paper, Ebright and coworkers used PCR to mutagenize DNA”.

- Write a sentence that summarizes all your findings
- Develop your results section with concise text followed by graphics that show your data.
- Be sure to use units (if they exist) and be careful to make units readable!
- Tabulated data **must** be in the form of a table.
- Try to visualize data as much as possible<sup>4</sup>. For example, if you want to show a growing trend, it is better to plot an increasing line, rather than provide an array

<sup>4</sup>The choice of a plot type is also important. See some guidelines here in the lecture 3: <https://epfl-ada.github.io/teaching/fall2023/cs401/>

with numbers. Use this, if you need to show the trend/relation, in which the precise values are not that important.

- For each metric, indicate the best value (for example, “the lower the better”). Whenever possible, try to use/create metrics that have the same trend, for example, each of them is the lower the better.
- Include legends in every figure/table. Legends to the figures and tables explain the elements that appear in the illustration. Conclusions about the data are not included in the legends. Your figure legends should be written in the present tense since you are explaining elements that still exist at the time that you are writing the report. Follow the conference/journal/thesis template requirements for figures/tables and their captions.
- Avoid using screenshots as figures. Vector graphics are preferred. To improve readability of your paper, use proper color schemes that are colorblind-safe (and print-friendly if needed). There are nice tools in the internet for choosing colors<sup>5</sup>.

For the DL-oriented projects, start by introducing the metrics (this may be done in Section IV as well). Compare your model with the baselines. Do not forget about ablation studies; they will help you with the justification, which we talked about in Section III. For example, if you proposed two novel techniques, run your model without both of them, with only one of them, with another one, and, finally, with both of them. This way you will identify which of them is more important. DL models might be seed-sensitive, therefore, run your pipeline several times and indicate the mean / variance.

To follow modern trends of DL papers, you may want to include some efficiency metrics: number of parameters, floating-point operations (FLOPs) or multiply-accumulate operations (MACs), memory required for processing one batch, time needed for one epoch/step, inference time, real-time factor (in audio, ratio between GPU processing time and audio time), etc.

## VI. DISCUSSION

Sometimes you may not see “Discussion” section directly. Its content may be implicitly written in the “Results” section. It is up to you to choose whether you separate this Section from the Section V or not. You may also want this Section to be a part of the Section VII instead.

In this Section, you provide analysis of the figures and tables from Section V. Show your understanding/interpretation of the data. For the DL-projects, try to answer the questions: “Does your model perform better?”, “How much better?”, “What are the advantages and disadvantages of the model?”.

The “Discussion” Section is one of the most important ones. This is the Section that proves that your method indeed works and it is the important contribution to the field.

<sup>5</sup>For example, <https://colorbrewer2.org>

## VII. CONCLUSION

Sum up your findings. A good conclusion answers these questions:

- What did you do in the project/lab? Restate the purpose/problem, a brief description of how you tested it, what you used to gather data.
- What does your data say? Look at your data table or sketch and turn it into a sentence or two. Be sure to include both the control and experimental groups. Basically, you rephrase and sum up the Results/Discussion Section. In the DL-projects, this may be something like: *Our experiments showed that our method outperforms existing solutions by a large margin.* If you proposed several methods, give a short sentence about all of them: *MethodA achieved lower quality than SOTAModelFromTheLiterature, however, MethodB surpasses all other solutions by a large margin.*
- What did you learn? This should answer the question posed in the purpose/problem.
- (Optional.) What can be improved? Indicate the limitations of the work and your ideas on how to solve them or why do they exist.

Do not forget that conclusion should be short. Do not copy-paste text from Introduction and Results sections; you should summarize them instead.

We summarized the whole guide in Appendix A. We suggest you to have a look. To further improve your writing skills, we advise you to read as much literature as possible. While reading, pay attention to the paper structure, how authors explain everything, etc. This experience will allow you to become a better writer yourself.

## VIII. ACKNOWLEDGEMENTS

We thank Cheng Soon Ong [6] from ETH Zurich, Andrea Ridolfi from Bern University of Applied Sciences, and authors of [7] from MIT OpenCourseWare for making their “How to write a paper” materials available for students.

**P.S.** If someone helped you with the project or you had some funding, write it in this Section.

## REFERENCES

Include only those references that pertain to the question at hand. You should list the references alphabetically by the first author’s last name or according to their order of appearance in your report (check conference/journal requirements). Include all the authors, the title, the name of the journal/conference in which it was published, its year of publication, the volume number, and page numbers (use as complete metadata as possible, use BibTex or BibLaTeX).

If your claims are based on the previous work, do not forget to cite it. Add references for the datasets, models, algorithms (common ones like Adam [8] too). As a starting point, use Google Scholar, Scopus, or their alternatives.

However, they sometimes contain the pre-print version of a citation instead of a conference/journal one. In such a case, you have to find a proper metadata in the internet yourself. When you are writing a paper, add references immediately and not after the whole article is finished. Otherwise, you increase the risk of forgetting to cite something.

## REFERENCES

- [1] S. Pegg, K. Li, and X. Hu, “Rtfs-net: Recurrent time-frequency modelling for efficient audio-visual speech separation,” *arXiv preprint arXiv:2309.17189*, 2023.
- [2] T. Huang, T. Chen, M. Fang, V. Menkovski, J. Zhao, L. Yin, Y. Pei, D. C. Mocanu, Z. Wang, M. Pechenizkiy *et al.*, “You can have better graph neural networks by not training weights at all: Finding untrained gnns tickets,” *arXiv preprint arXiv:2211.15335*, 2022.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [4] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, “Res2net: A new multi-scale backbone architecture,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 2, pp. 652–662, 2019.
- [5] P. Grinberg and V. Shikhov, “Rawspectrogram: On the way to effective streaming speech anti-spoofing,” *IEEE Access*, 2023.
- [6] C. S. Ong, “Writing scientific papers and software,” <https://da.inf.ethz.ch/teaching/2022/DeepLearning/files/howto-paper.pdf>, [Accessed 12-04-2024].
- [7] Laboratory Fundamentals in Biological Engineering, MIT OpenCourseWare, “Guidelines For Writing Up Your Research,” <https://ocw.mit.edu/courses/20-109-laboratory-fundamentals-in-biological-engineering-spring-2010/pages/assignments/guidelines-for-writing-up-your-research/>, [Accessed 12-04-2024].
- [8] D. P. Kingma, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [9] C. Hardebolle, V. Macko, V. Ramachandran, A. Holzer, and P. Jermann, “Digital Ethics Canvas: A Guide For Ethical Risk Assessment And Mitigation In The Digital Domain,” 2023.

## APPENDIX

### A. Evaluation of the Report Quality

Table I provides you with the general basic idea of how to understand if your report/paper is good or not. Table II provides you with the writing style guide. You should aim to follow the “good” evaluation criteria.

### B. Ethics and Risks

There is a rising trend on assessing ethical risks, you can find information how to do this in [9].

Table I  
EVALUATION CRITERIA FOR EACH SECTION.

Section	Goal	Evaluation		
		Good	Ok	Not Good
Title	To give content information to reader.	Engaging.	Appropriate.	Not enough content information or too much.
Abstract	To concisely summarize the experimental question, general methods, major findings, and implications of the experiments in relation to what is known or expected	Key information is presented completely and in a clear, concise way. All information is correct. Organization is logical. Captures any reader's interest	Sufficient information is presented in proper format. Would benefit from some reorganization. Understandable with some prior knowledge of experiment.	Some key information is omitted or tangential information is included. Some information is misrepresented. Some implications are omitted. Incorrect format is used.
Introduction	To identify central experimental questions, and appropriate background information. To present a plausible hypothesis and a means of testing it.	Relevant background information is presented in balanced, engaging way. Your experimental goals and predictions are clear and seem a logical extension of existing knowledge. Writing is easy to read. All background information is correctly referenced.	Relevant background information is presented but could benefit from reorganization. Your experiment is well described and a plausible hypothesis is given. With some effort, reader can connect your experiments to background information. Writing is understandable. Background information is correctly referenced.	Background information is too general, too specific, missing and/or misrepresented. Experimental question is incorrectly or not identified. No plausible hypothesis is given. Writing style is not clear, correct or concise. References are not given or properly formatted
Related Work	To provide information about state-of-the-art solutions, baselines, and useful/required algorithms/techniques from the literature	All baselines are presented in a brief format. Required background algorithms/techniques are explained. All citations are provided and correct. Only useful information is written.	All baselines and algorithms/techniques are explained but have unnecessary too deep details.	Not relevant/not used baselines. None or missing references. Rephrasing of original paper instead of brief summary.
Methodology and Experimental Setup	To describe procedures correctly, clearly, and succinctly. Included a correctly formatted citation of the lab manual.	Sufficient for another researcher to repeat your experiment. Steps presented.	Procedures could be pieced together with some effort. Steps presented.	Procedures incorrectly or unclearly described or omitted. Steps not presented.
Results	To present your data using text AND figures/tables.	Text tells story of your major findings in logical and engaging way. Figures and tables are formatted for maximum clarity and ease of interpretation. All figures and tables have numbers, titles and legends that are easy for the reader to follow.	Text presents data but could benefit from reorganization or editing to make story easier for reader. Text includes interpretation of results that is better suited for discussion section. Figures and tables are formatted to be clear and interpretable. All figures and tables have numbers, titles and legends.	Text omits key findings, inaccurately describes data, or includes irrelevant information. Text difficult to read due to style or mechanics of writing. Text difficult to read due to logic or organization. Figures and tables missing information, improperly formatted or poorly designed. Figures and tables have inadequate or missing titles or legends.
Discussion	To evaluate meaning and importance of major findings.	Appropriate conclusions drawn from findings. Connections made between experimental findings. Connections made between findings and background information. Future directions considered. Writing is compelling.	Appropriate conclusions drawn from findings. Experimental limitations considered. Writing is clear.	Conclusions omitted, incorrectly drawn or not related to hypothesis. Relationship between experimental findings and background information is missing or incorrectly drawn. Writing style and mechanics make argument difficult to follow.
References	To give credit work on which your own is based.	Complete list of reliable sources, including peer-reviewed journal article(s). Properly formatted in body of report and in reference section.	Adequate list or reliable sources. With minor exceptions, properly formatted in body of report and in reference section.	List is incomplete or includes sources not cited in body of report. List includes inappropriate sources. List not properly formatted. References not properly cited in body of report.

Table II  
EVALUATION OF WRITING STYLE

Writing Style and Mechanics	Evaluation	
	Good	Not Good
Verb Voice	Appropriate for audience. Consistent passive or active voice.	Too simple or too advanced. Irregular use of passive and active voice.
Word choice	Concise. Says what you mean. Scientific vocabulary used correctly.	Verbose. Ambiguous or incorrect. Scientific vocabulary misused.
Fluency	Sentences and paragraphs are well structured. Punctuation is correct or has only minor errors. Grammar is correct or has minor errors. Spelling is correct.	Sentences are repetitive or awkward. Paragraphs are not logical. Periods, commas, colons, and semicolons are misused. Significant number of run-on sentences, sentence fragments, misplaced modifiers, subject/verb disagreements. Significant number of spelling errors.
Scientific format	Past tense for describing new findings. Present tense used for accepted scientific knowledge and figure legends. All sections are included and properly formatted. Formal language	Misleading verb tenses. Some sections are missing. Figures miss legends. References are not properly formatted. Informal language: contractions, slang, etc.