

INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE TOULOUSE

Optimizing and Adapting Language Models for Domain-Specific Task

End-of-studies Apprenticeship Report

Minh Duy Nguyen



Bachelor's Degree in Computer Science and Engineering

Supervised by: Milad Mozafari (Torus AI), David Bertoin (INSA Toulouse)

January 20, 2026

Abstract

[**INFO:** *The abstract is essentially informative in nature and should be written concisely (up to 200 words) in a way that captures the interest of the reader.*

The Abstract replaces reading the document and does not contain figures, tables, citations, etc. It should include the following topics: scope, objectives, methods, main findings, including results, conclusions and recommendations, if any.

For more information on how to write a good abstract, consult the online tutorial available on the Library website, “Publication Support Guide”, section: “Structuring Technical Report”.]

Write the Abstract, but only at the end.

Acknowledgement

[**INFO (optional element):** *Usually the contribution of other people or entities is mentioned, both for carrying out the study and for producing the report. They can be done on a separate page or included in the introduction.*]

Table of contents

1	Introduction	1
1.1	Context	1
1.2	Objectives and expected results	1
1.3	Report structure	1
2	Methodology used and main activities developed	7
2.1	Methodology used	7
2.2	Participants, roles and responsibilities	7
2.3	Activities developed	7
3	Solution development	9
3.1	Requirements	9
3.2	Architecture and technologies (or Design and Implementation)	9
3.3	Developed solution	10
3.4	Validation	10
4	Conclusions	11
4.1	Results achieved	11
4.2	Lessons learned	11
4.3	Future work	11
References		12
A	An Appendix	13

Todo list

Write the Abstract, but only at the end.	i
Remove the “Examples” section when no longer needed.	2
Figure: Insert the Christmas figure.	3

Figures and tables

bash

Bash is a *Unix shell* and command language written in 1989 by Brian Fox for the GNU Project as a free software replacement for the *Bourne shell*.

firewall

In computing, a *firewall* is a network security system that monitors and controls incoming and outgoing network traffic based on predetermined security rules. A *firewall* typically establishes a barrier between a trusted network and an untrusted network, such as the Internet.

Figures and tables

Figures and tables are essential elements in a report, providing visual representations of data and information. They should be clearly labeled and referenced in the text.

Chapter 1

Introduction

[**INFO:** *Succinct contextualization of the report subject, making reference to the scope and objectives. Here the motivation of the presented work is clarified and the adopted approach and its relationship with similar works is explained, from a generic perspective. Details about what is explained in the following sections should not be anticipated. If relevant, the target audience can also be indicated¹.*]

1.1 Context

[**INFO:** *Present the organizational context in which the project/internship took place (company, institution, research unit, research laboratory, etc.). Present the problem addressed and the motivation for the work carried out (what is the problem addressed and why is it important).*]

1.2 Objectives and expected results

[**INFO:** *Indicate the objectives of the work and the expected results.*]

1.3 Report structure

[**INFO:** *Briefly describe the structure of the report.*

It is expected that the report will have between 10 to 25 pages (in A4 format, single column, with a font size not exceeding 12pt in paragraph text), already including appendices.]

In addition to the introduction, this report is organized into 4 more chapters. In Chapter 2 . . .

1. For more information, consult the online tutorial “Publication Support Guide”.

Examples

Remove the “Examples” section when no longer needed.

[**INFO:** *Some parts of the document are illustrated below.*]

This section does not appear in the final document!]

Equations

[**INFO:** *This text is just an example that precedes an equation.*]

Simple equations can be inserted inline with the text: the line is $y = mx + b$.

More complicated equations should be separated into individual lines and numbered sequentially on the right within parentheses. This is the generic quadratic equation:

$$ax^2 + bx + cx = 0 \quad (1.1)$$

Where a is the quadratic coefficient; b the linear coefficient; c the constant coefficient independent of variable x , to be determined.

Equations should be referred to by their number. For example, Equation 1.2 solves problems formulated as shown in Equation 1.1.

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (1.2)$$

Figures and tables

All figures and tables must be captioned and numbered sequentially:

- figures should be captioned below;
- tables should be captioned at the top.

Keep figures centered and inline with text so that the caption always appears attached to the image.

[**INFO:** *Figures should float freely on the page and be referenced and described in the text, with sources properly stated, to avoid plagiarism.*]

As an example, Figure ?? shows the FEUP campus.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean

faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

[**INFO:** *Space can be reserved for placing a figure in the future; for example, Figure 1.1.*]

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

[**INFO:** *Tables should float freely on the page and be referenced and described in the text, with sources properly stated, to avoid plagiarism.*]

Table 1.1 serves to exemplify how to show some values that, in this case, are related to some numerical data associated with resources and investments of FEUP in the year 2011.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean

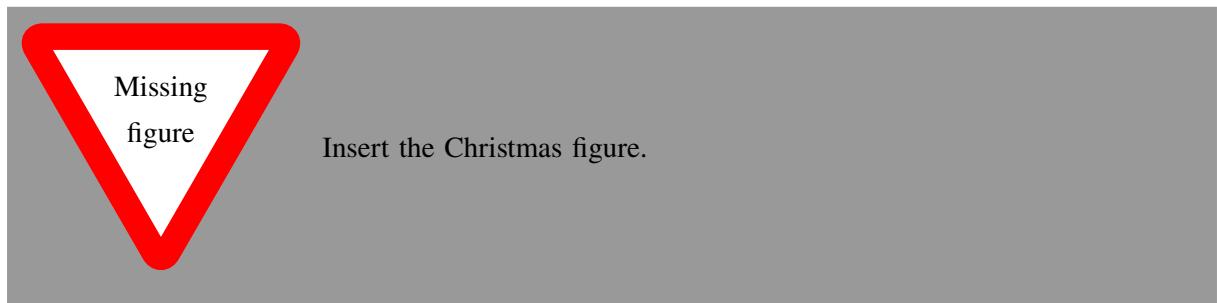


Figure 1.1: Christmas at the FEUP Campus.

Table 1.1: FEUP Physical Resources. Adapted from FEUP 2023

Description	Quantity
Total area of FEUP campus	93 918 m^2
Green spaces	23 000 m^2
Number of computers dedicated to teaching	1 815
Investment in laboratory equipment	1.46 M€

faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Citations

As you write the report text, you should indicate the works of other authors on which you are based, in the form of citations. This consists of briefly indicating the sources used from which you obtained additional information to develop the topic of your report.

There are two main ways to cite:

- by **paraphrase**: interpretation of the original content in words different from those of the consulted source, indicating the source immediately after; or
- by **transcription**: use of an excerpt of the original content presenting it in quotation marks, indicating the source immediately after.

Citations must follow a standardized style. Among the many that exist, the FEUP Library recommends the Chicago style (author-date format).

[**INFO:** *Below are some examples, at random, of citations (by paraphrase) according to this style.*]

The decision to choose a topic for an academic work can vary Bell 2002. The topic can be thought of and chosen by the student themselves, or from a list of topics already conceived, with potential interest for study F. Gomes, Lopes, and J. L. Palma 2014.

Each citation throughout the text should correspond to a reference indicated in the final list of bibliographic references Lipsum 2023; Mikkelsen 2012; Vasiljevic, Lopes, and Gomes 2018.

It is important not to forget that figures (images, tables, graphs, etc.) from works by other authors (for example obtained through the Internet) should always be cited, after the respective captions D. Gomes, Lopes, and Palma 2022.

For more information on this subject and to see examples, consult the guide “Avoiding plagiarism: good practices in the use of information”².

Duis non odio morbi ut dui sed accumsan risus eget odio ISO 2011; Ornelas 2016.

Code

2. <https://feup.libguides.com/plagio/citar>

[INFO: Below is an illustration of including code in the document.]

Listing 1.1 Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

Listing 1.1: Python example

```

1 # Take the user's input
2 words = input("Enter the text to translate to pig latin: ")
3 print(f"You entered: {words}")
4
5 # Break apart the words into a list
6 words = words.split(' ')
7
8 # Use a list comprehension to translate words greater than or equal to 3 characters
9 translated_words = [(w[1:] + w[0] + "ay") for w in words if len(w) >= 3 ]
10
11 # Print each translated word
12 for word in translated_words:
13     print(word)

```

Using macros

[INFO: Below is an illustration of using *L^AT_EX* macros defined in the preamble. Note the use of the \class{} macro for classes, methods and components.]

The WindsPT, taken from *Windscanner.PT*, uses SVG ... as in `Student.calculate-age()`.

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

[INFO: The subsequent component parts that make up the body of the text should be structured in sections, estimating that up to 3 levels is sufficient for this type of work.

For more information, consult the online tutorial “Publication Support Guide”. Note that the sections indicated there can be adapted according to the topic or depth of the study to be developed.]

[**INFO:** *It is not customary to have consecutive section headings without text.*]

The dash

About the use of the hyphen and the dash³:

1. The hyphen (Alt + 0045): without spaces (-);
2. The em-dash (Alt + 0151): in Portuguese, surrounded by spaces (—);
3. The en-dash (Alt + 150): without spaces (-);
4. The mathematical subtraction sign: without spaces (-).

Quotation marks

About quotation marks in L^AT_EX, either the glyph is used directly, or they are made with the backtick at the beginning and the apostrophe at the end, as in “example”⁴.

3. “Portuguese Language Questions”, <https://ciberduvidas.iscte-iul.pt/consultorio/perguntas/o-uso-do-hifen-e-do-travessao/31251>

4. What is the best way to use quotation mark glyphs? <https://tex.stackexchange.com/questions/531/what-is-the-best-way-to-use-quotation-mark-glyphs>

Chapter 2

Methodology used and main activities developed

In this chapter, the methodology followed is described, the main participants in the project are listed and the main activities developed are recorded.

2.1 Methodology used

[**INFO:** *Describe the methodology Despa 2014 followed (for example, iterative development with biweekly sprints and weekly follow-up meetings) and the resources used (for example, GitHub¹, etc.).*]

2.2 Participants, roles and responsibilities

[**INFO:** *Identify the project team, stakeholders and other participants with whom there was interaction; in the case of group work, clarify the roles and responsibilities of each group member.*]

2.3 Activities developed

[**INFO:** *Describe the activities carried out over time (including relevant events, such as presentations, meetings with clients, etc.) and respective deliverables, typically using a Gantt chart wikipedia 2023 (see Figure 2.1) and a brief description of each activity/deliverable. It can also be presented through a table with weekly progress.*]

1. <https://github.com/>

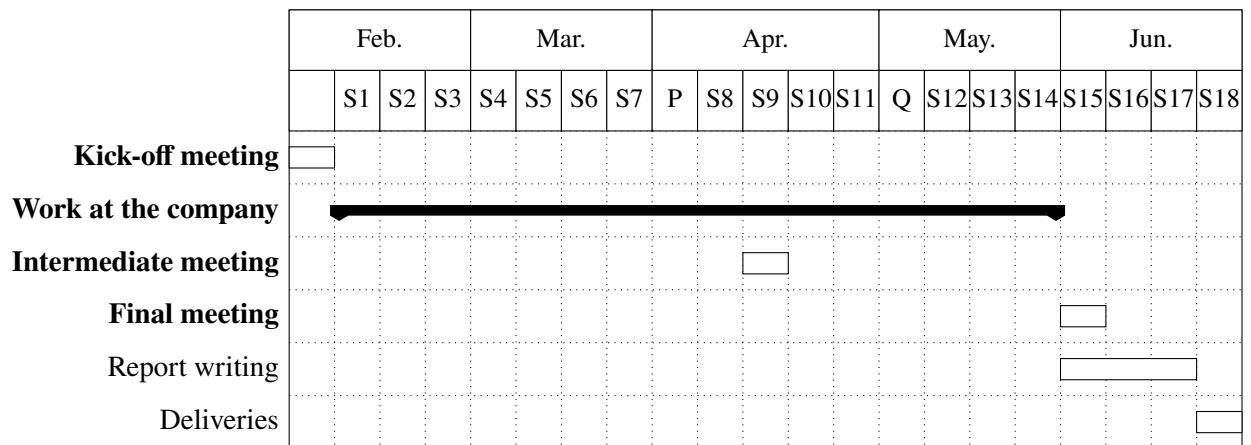


Figure 2.1: Gantt chart.

Chapter 3

Solution development

In this chapter, the work developed to achieve the expected results is described.

If it is the case of a software prototype, the requirements, solution architecture, development and validation of the prototype are presented.

3.1 Requirements

[**INFO:** Identify relevant functional and non-functional requirements and respective sources, as well as project constraints.

Below is an example of a table with user stories, which can be used if the number of rows is large. An additional column for priorities may be needed.]

User stories

Examples of possible usage scenarios are presented in Table 3.1, using *user stories* written according to the model: “As a <type of user>, I want <some goal> so that <some reason>”.

Table 3.1: Usage scenario

Identifier	Name	Description
US01	Register (High)	As a Visitor, I want to register a new account so that I have access to profile data

3.2 Architecture and technologies (or Design and Implementation)

[**INFO:** Architecture and technologies used and respective justification (taking into account requirements and existing alternatives), technical diagrams prepared (see Figure ?? taken from SPARX¹), technical difficulties encountered and their resolution, etc.]

1. <https://sparxsystems.com/resources/tutorials/uml2/component-diagram.html>

3.3 Developed solution

[**INFO:** *Present the developed solution from the user's perspective, with the help of screenshots.*]

3.4 Validation

[**INFO:** *Description of the validation of the developed solution, in relation to the identified requirements and constraints, and respective results (for example, experimental evaluation results, tests performed, feedback from users or specialists, etc.).*]

Chapter 4

Conclusions

In this chapter, the results achieved and lessons learned are summarized. Finally, the limitations of the work are presented and improvements and future work are proposed.

4.1 Results achieved

[**INFO:** *Summarize the results achieved and contributions (in relation to the objectives).*]

In the case of group work, clarify individual contributions, in qualitative and quantitative terms (percentage).]

Sed feugiat. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Ut pellentesque augue sed urna. Vestibulum diam eros, fringilla et, consectetur eu, nonummy id, sapien. Nullam at lectus. In sagittis ultrices mauris. Curabitur malesuada erat sit amet massa. Fusce blandit. Aliquam erat volutpat. Aliquam euismod. Aenean vel lectus. Nunc imperdiet justo nec dolor.

4.2 Lessons learned

[**INFO:** *Reflect on the lessons learned (taking into account the learning objectives).]*

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

4.3 Future work

[**INFO:** *Identify limitations of the work performed and ideas for improvements and future work.]*

Bibliographic references

- Bell, Judith. 2002. *Como realizar um projeto de investigação*. Gradiva. Lisboa.
- Despa, Mihai Liviu. 2014. “Comparative study on software development methodologies.” *Database Systems Journal* 5 (3): 37–56.
- FEUP. 2023. *FEUP em Números 2023*. Disponível em https://sigarra.up.pt/feup/pt/web_base.gera_pagina?p_pagina=258031.
- Gomes, D., J. Correia Lopes, and Palma. 2022. “WindsPT e-Science platform for wind measurement campaigns.” In *TORQUE2022 (EAW's Science of Making Torque from Wind)*.
- Gomes, Filipe, J. Correia Lopes, and José Laginha Palma. 2014. “WindS@UP: The e-Science Platform for WindScanner.eu.” *Journal of Physics: Conference Series* 524 (The Science of Making Torque from Wind 2014 (TORQUE 2014) 18–20 June 2014, Copenhagen, Denmark): 012006. ISSN: 17426596. <https://doi.org/10.1088/1742-6596/524/1/012006>.
- ISO. 2011. *Geographic information – Observations and measurements*. Standard ISO 19156:2011. International Organization for Standardization.
- Lipsum. 2023. *Lorem Ipsum*. Accessed November 15, 2023. <http://www.lipsum.com/>.
- Mikkelsen, Torben. 2012. “WindScanner.eu — a new Remote Sensing Research Infrastructure for On- and Offshore Wind Energy.” In *Proceedings of the International Conference on Wind Energy: Materials, Engineering, and Policies (WEMEP-2012)*. www.windscanner.eu.
- Ornelas, José Pedro Alves. 2016. “Platform for monitoring and treat depression.” Master’s thesis, Universidade do Porto (Portugal).
- Vasiljevic, Nikola, J. Correia Lopes, and Daniel Gomes. 2018. *e-WindLidar: making wind lidar data FAIR*. Technical report. DTU, Denmark. <https://doi.org/10.5281/zenodo.2478051>.
- wikipedia. 2023. *Gantt chart*. Accessed March 15, 2023. https://en.wikipedia.org/wiki/Gantt%5C_chart.

[**INFO:** *In the final list of references, the works of the authors cited in abbreviated form throughout the text should appear, obtained automatically with BibTeX. The bibliographic reference is the most developed way of indicating the sources of information on which it was based.*]

Appendix A

An Appendix

[**INFO (optional element):** *Appendices and annexes contain information that complements, supports and clarifies the report and whose inclusion in the main body of the report would interfere with a good order of presentation of ideas.*

There is an important difference between appendices and annexes: “In the appendix only documents authored by the author of the report are compiled, while in the annex documents authored by other authors than the report author are compiled.”¹]

1. For more information, consult the online tutorial “Publication Support Guide”.