# ANU Computing Internship

Statement of Work

Parsing custom templates for fake text generation.

Thesis….

Student project COMP3820 / COMP4820 / COMP8830

|  |  |  |  |
| --- | --- | --- | --- |
| Student Name: | Luke Wanless | Student ID: | U7120506 |
| Degree Program: | Bachelor of Advanced Computing (Research and Development) (Honours) | Course Code: | COMP3820 |
| Host Organisation: | Penten | Host Supervisor: | Isaac Jackson |
| Project Title: | Parsing custom templates for fake text generation | | |

*Note: Delete text highlighted in yellow. Update text highlighted in green.*

*The objective of this assessment item is to demonstrate your understanding and plan for your project in your own words. You should not copy from the host website or any other source. You should discuss each of these sections in detail with your host and project stakeholders where available. It’s not expected that you have all the answers – it’s your job to get the answers by talking to people and doing some research. Don’t be afraid to ask questions! Maximum word count is 2000 words (do not exceed). Use diagrams/images/charts/tables to illustrate where appropriate.*

# Project Vision and Objectives

*Penten are a cyber security company that provide security solutions to private sector and government organisations. Specifically the Applied AI department utilise modern AI tools to create Honeyfiles for cyber deception. These Honeyfiles are designed to alert system administrators to the presence of hackers within their filesystems. As apart of this customers need to be able to generate Honeyfile documents that appear realistic. This has motivated the addition of a feature which allows clients to upload internal organisation templates which can then be populated with tracers and fake generated content.*

*My project concerns parsing these templates in a robust manner to allow for automated generation with minimal supervision and avoid unwanted side effects. Specifically templates may be populated with stub text or unpopulated. Text will appear in headings, paragraphs, emails and in other locations. Similarly these templates may be entirely empty. In order to generate text in an automated and robust manner it is necessary to first classify these document elements such that the appropriate replacement text can be generated. My project aims to use large language model API to parse the XML of .docx files in an unsupervised manner and return a complete honeyfile.*

*This is an important extension to the hosts current functionality as it improves the realism of generated honeyfiles and removes the requirement of manually uploading and onboarding the template specifications of each new client; a process that is not scalable or compatible with widely distributed software.*

*The end goal for the project will be uploading an array of custom templates with random fields, parsing this template accurately and returning a contextually consistent Honeyfile using large language models.*

*[~300 words (suggested not required)]*

*This must be in your own words, showing your understanding of the host organisation and project. Do not copy from the host website or other sources.*

* Start by describing your host organisation – who they are and what they do?
* Describe your project and objectives – what is it about, why is it important to your host, what are the goals?

# Project Stakeholders

*The stakeholders for this project are Penten’s product management team, Applied AI development team and users of HoneyTrace. For security reasons the users of Honeytrace cannot be named other than Australian Defence and private sector.*

*[~300 words (suggested not required)]*

*This must be in your own words, showing your understanding of the stakeholders. Stakeholders are not just the people you work with, they are anyone who might be impacted by or interested in the work that you are carrying out (e.g., end users).*

* Who are the stakeholders in this project and why? What do they do? How do they interact?
* What are the expectations of all stakeholders for this project, including your host?

# Project Benefit

*[~200 words (suggested not required)]*

*This must be in your own words, showing your understanding of the project benefit.*

* What are the benefits of this project and to who? How will the project make things better for the host and other stakeholders?

# Project Plan

*[~300 words (suggested not required)]*

*This section demonstrates your plan for the project. You might want to use tables or diagrams to illustrate.*

* What are the key deliverables and outputs for the project? What/when are the milestones? Include your assessment deliverables.
* Create a schedule for your project, showing the tasks/deliverables/milestones you will complete and when.

# Project Resources and Tooling

*[~300 words (suggested not required)]*

*This section demonstrates that you have considered the logistics of your project*

* What are the resources available to your project and the potential costs? (e.g., people, tools, equipment, consumables) Where will these resources come from? Who will bear the costs?
* What tools will you use for your project? (e.g., development tools, task management)
* What equipment will you use / do you need for your project? Do you have this equipment?

# Project Constraints and Risks

*[~300 words (suggested not required)]*

*This section demonstrates that you understand the constraints and risks related to your project.*

* What are the technical or other constraints related to your project (e.g., reliability, security, safety)?
* What are the potential risks associated with your project? How will they be managed? Why might your project fail and what can you do about it?

Supervisor Approval

***Your supervisor must read and approve this document for it to be assessed.***

*Your supervisor must sign and the date this box.*

***Do not leave this step to the last minute.***

*You must give your supervisor a reasonable amount of time to review this document. You should also develop this document in close consultation with your supervisor. (They shouldn’t be seeing it for the first time when you ask for their approval).*

*I [Supervisor name] confirm that I have read and agreed to the scope of works set out in this Statement of Work document.*

|  |
| --- |
| *Supervisor signature Date* |