­­­­­Personal Development Report

Thomas van der Molen

S7-AI-A-RB

|  |  |
| --- | --- |
| **Project Information** | |
| Project members | Thomas van der Molen (4168003) |
| Project name | Personal Development Report |

Table of Contents

[Introduction 3](#_Toc157200616)

[1. Data Preparation 4](#_Toc157200617)

[2. Data analysis & model engineering 7](#_Toc157200618)

[3. Targeted interactions 10](#_Toc157200619)

[4. Future orientation 14](#_Toc157200620)

[5. Investigative problem solving 17](#_Toc157200621)

[6. Personal leadership 21](#_Toc157200622)

[7. Graduation Preparation 24](#_Toc157200623)

[Retrospective 26](#_Toc157200624)

[Conclusion 27](#_Toc157200625)

# Introduction

My name is Thomas van der Molen, I am 21 years old and have previously done Mavo and Havo before coming to Fontys to study HBO ICT & Software Engineering.

I am currently working for the software company Author-e where I lead several different software projects.

I have done Advanced Software in my previous semester, and while I still want to focus on Software I think there is a lot to learn within the domain of AI, hence I decided to do Advanced AI as well.

During this semester, I want to learn more about the underlying algorithms power large AI models.

All previous semester I have been able to achieve an Outstanding grading, so I will try to achieve Outstanding this semester as well. However, I did recently learn that getting Outstanding every semester does not actually directly affect graduating with Cum Laude and for this reason I have put less importance on it.

# Data Preparation

You can **enrich/extend** and **clean** datasets, and are able to investigate and argument the use of **data storing methods**.

Evaluation 1: I think this learning outcome is currently at an **Orienting** level.

Evaluation 2: I think this learning outcome is currently at a **Proficient** level.

Evaluation 3: I think this learning outcome is at a **Proficient** level, **Advanced** if I had more direct evidence.

**Evidence Evaluation 1**

During the start of the group project, I have spent a lot of time researching varying DevOps tools specifically for machine learning projects, especially DVC for which I have created a research document, with guide for fellow students and have implemented my findings for data / model / experiment -management in the group project.

Furthermore, I have started performing data analysis on the data of the group project together with my project group, and have done initial dataset exploration for smaller datasets used during exercises which I have extensively discussed with Technical tutor (Olaf).

**Evidence Evaluation 2**

Since the last evaluation, I have created together with the rest of the project group the final data preparation notebook, this notebook contains the data merging techniques I had utilized previously in my EDA for One-class SVM and shared with the rest of the group. In this merging process, the datasets are grouped on specific identifiers, effectively squashing the data into less records, this process also creates new features such as average/median, deviations, timespans etc. that could enrich data records with timeseries-like features that from my research should be good indicators for predicting anomalies.

In the pre-processing file, I also create many versions of the datasets, including final Train/Test split datasets that can be used to standardize the data used when training/evaluating so that the various models can be compared more closely with each other. All the saved models are stored in DVC that I had previously set-up, allowing group-members and cloud pipelines to easily fetch the datasets needed (which with these enrich and computed datasets are starting to reach the 1GB file size which is already larger than GitHub’s 100MiB file push limit).

Feedback from Olaf confirms that the processing done was very well (note this feedback is of the whole data-preprocessing notebook which includes work of other students Fig 1), but more information of the reasoning behind the processing steps (EDA) could still be added.

A screenshot of a phone

Description automatically generated

Figure 1 Olaf Feedback SUE data-preparation

Furthermore, I have also started looking into how to convert web pages (HTML/CSS/JS) into reinforcement learning environments, for my open-program/data challenge. This work was shortly covered in my data challenge plan, on which Qin gave positive feedback on (Fig 2).

A screenshot of a phone

Description automatically generated

Figure 2 Qin data challenge plan feedback

Due to the vast amount of work I have done on the data preparation for group project and my own projects, I think I have shown a sufficient level of knowledge and execution within this field.

**Evidence Evaluation 3**

During this final phase, ironically DVC came back to proof it’s usefulness and increasing value it has provided since I had integrated it into our workflow from the start of the semester, when our group decided to create a streamlit application to showcase our work done as an interactive presentation, to get this working, we required some of the larger datasets of our project (which cannot be stored directly in our GitHub repository), this was made extremely easy as the cloud app could automatically pull them using DVC with no further issues. The project stakeholders also started to understand the benefit of this use-case, as with the project transfer they were easily able to get all our datasets locally via the DVC and could immediately integrate it within their own cloud storage infrastructure if they wanted to. This sentiment can also end up being seen in the anonymous peer feedback I had received, with the other project members also valuing the tool (Fig 3).

A text on a white background

Description automatically generated

Figure 3 Anonymous peer feedback

Secondly, for the group project I had been exploring transformers as a possible anomaly detection technology, one problem with these complex models, was that the data could not simply be fed in as the csv data we are generally working with. To get the data to work nicely with transformers, I had to first normalize and then vectorize, for which I had to write a compact script that also included some slight unexpected data processing specifically for the used TranAD model such as it expecting a true label during for evaluation for every individual feature (as it is able to predict per feature). But this data processor ended up working and I was able to feed the data in with no issues (even though only ~1% of the data ended up being used).

Furthermore, as the forced upon role of resident state machine/TUDelft research paper expert within the group, I also helped Tymofii to get the preprocessor that was provided to us for the statemachine model working during his research.

Lastly, for my data challenge, I had developed a web environment that can provide state information to the reinforcement learning agent. Getting the data for a given state, many collection/processing steps had to be developed, as every feature was unique. The most important features that could not be queried directly via javascript was the interactable components, as generally you can find all interactable components by checking the eventlisteners on a page, but after digging around for a long time, and even asking Olaf for possible ideas (as I knew he had experience in front-end development), it turns out to be impossible to query all eventlisteners on a page without directly editing the source javascript of the application thereby breaking zero integration requirement of the project. The only way around this was to specifically run a chrome browser and do the specific eventlistener query command via the developer command line. After getting the listeners however, the fun wasn’t over, as pretty much all current RL models, do not handle changing size of action spaces (the amount of interactables) well, so instead based on an idea from Simona, I instead create an internal list of these elements, and allow the model to scroll through the list and interact with the currently selected element thereby setting my action space to a discrete 2 action, this however breaks the general reward structure that these models use, so I had to modify how my agent deals with rewards by instead of calculating rewards and processing them on every action, it is done per action phase which only completes when an element is attempted to be interacted with.

# Data analysis & model engineering

*You can* ***apply*** *different types of machine learning models and tune the* ***hyper-parameters*** *of these models. Furthermore, you can* ***evaluate*** *the results from a trained model and can perform* ***other data analysis techniques****.*

Evaluation 1: I think this learning outcome is currently at an **Orienting** level.

Evaluation 2: I feel that this learning outcome has improved to a **Beginning** level.

Evaluation 3: I think that I have proven an **Advanced** level of proficiency.

**Evidence Evaluation 1**

Besides the workshops offered during this semester, I have explored models of which I had limited understanding such as CNN, for which I had actually found very unique issues while evaluating the model, which I have extensively discussed and further explored based on feedback from the Technical tutor (Olaf).

**Evidence Evaluation 2**

To continue from Evaluation 1, I have spent even more time exploring issues I had with my ANN and CNN model exercises, however this time I used a more methodological approach with hypothesizing possible issues and utilizing XAI to prove these hypotheses, the effectiveness of this approach can be seen by the feedback left by Nico on my XAI exercise (Fig 1), with my ideas being earlier confirmed by Nico in my CNN exercise itself (Fig 2).

A screenshot of a test

Description automatically generatedA screenshot of a chat

Description automatically generated

Figure 2 Feedback (Highlighted) Nico CNN - Exercise

Figure 1 Feedback Nico XAI - Exercise

For my group project, I have also explored several kinds of outlier detection algorithms, mainly focusing on One-class SVM (including an entire research document) which I have discussed in great detail with my technical tutor (Olaf) and transformers for which I have done a significant amount of preliminary research and asking a lot of questions to Olaf about the topic after his NLP lectures, which from the feedback received during the meetings seems to be going in a good direction.

I also spent a lot of time working on the RL exercise, as I will need the knowledge and experience within this context for my open/data challenge project. During the RL exercise, I spent a significant amount of time exploring several environments and different types of learning models (notably Q-tables and deep learning models), investigating how certain techniques can be implemented and combined and why they work in the first place. All my efforts did result in some interesting/good reinforcement models that I fully understood and could explain, which Iman seemed to also be positive about according to his feedback (Fig 3).

A screenshot of a phone

Description automatically generated

Figure 4 Feedback (Highlighted) Iman RL – Exercise

From the evidence shown so far I think that I have proven a beginning and closing in on proficient level of knowledge within the data analysis and model engineering scope.

**Evidence Evaluation 3**

During this final phase, I ended up spending a lot of time looking into two kinds of modelling technologies, these being: Transformers and Reinforcement Learning model.

Within the group project, we wanted to see if anomaly detector transformers could outperform the at the time best state machines, sadly due to the specialized use-case of these models, you can’t simply import them from a magical python package like you would with sci-kit-learn models or tensorflow keras neural networks, and to make it more difficult; platforms like hugging face seemingly also do not have any of these illusive anomaly detection transformers, let alone pre-trained models, as I was afraid that even if we got a working model, training it would be near impossible.

I was able to find some models in research papers and public repositories however, but sadly almost all of them are abandoned or actively tell you they won’t help you when something goes wrong, which it of course it always did. I quickly learned that if I was to successfully implement a transformer for the group project I would need a fundamental understanding of the technology so that I could fix any issues along the way with using these models (as creating our own was definitely outside of the scope of this project, even though after looking at other people’s model’s source code for many hours I think I could give it a fair shot by now). During this time I often had meeting with Olaf to try and help figure out all the fundamental ideas that are implemented within these transformers, with him eventually stating that as far as he knows no one has really gone this deep before during the AI semester, and he even might not have all the answers, so I decided to use my Christmas vacation to read up on a bunch of research papers I had been bookmarking during my research.

All this work however, did have a positive outcome, as I was able to fork the TranAD repository and fix all the issues needed, as well as create a data processor that could vectorize our dataset correctly, and was able to generate some evaluation results for the performance of the model (even if it was only on ~1% of the dataset with no parameter tuning as I did not have time for this anymore, note: 1% dataset with 5 training epochs already takes 2+ hours). All this work was thankfully worth it, as the stakeholder of the project was happy with our wide coverage of different kinds of models (especially the deeper research into state machines, for which I also spent a significant amount of time, and the transformers), as well as a very small subset of people such as Iman (RL teacher) wanting to learn more specifically about this technology during the innovation insights event, which I now could provide.

Of course, all this work on transformers also bled into my NLP exercise, where I tried to already get a good idea of how models that utilize transformers work in practice, for which Olaf was very positive about my integration of it in one of my previous projects (Fig 4).

A text on a white background

Description automatically generated

Figure 5 NLP feedback – Olaf

Secondly, I also spent a lot of time researching and trying to implement advanced reinforcement learning models for my innovative data challenge for which I basically ran in to the exact same issues as with the transformers, with the only saving grace being the reduced level of complexity these models seemingly still have.

While I tried reaching out to experts within the domain such as Iman, the timings never seemed to work out as I was extremely busy during this time, but found some extremely good resources online for creating “rainbow” agents, which is what I was looking for.

I was able to use one of these resources to create a model and agent, for which I ended up implemented my own custom data preprocessing layer for feeding the complex environment data into the DQN, and implementing noisy layers into the network to replace the general epsilon greedy approach, as it would fit my use-case significantly better.

I was able to completely implement this model and even tune it to my custom environment, adding special functionality such as context aware exploration, where it will automatically determine if a cross-origin redirect has taken the agent outside of the testing scope or not. This model, ended up even performing better than I could have expected (depending on the complexity of the webpage), with Simona even mentioning in her feedback for the final data challenge report: “finally putting Reinforcement Learning to work for a serious application!” (fig 5)

A text on a white background

Description automatically generated

Figure 6 Data Challenge report feedback - SImona

# Targeted interactions

*You can* ***clearly communicate*** *with readers/audience, keeping into* ***consideration*** *the knowledge of set audience and the medium used to communicate information.*

Evaluation 1: I think this learning outcome is currently at an **Orienting** level.

Evaluation 2: I think that I have proven a **Beginning** or even **Proficient** level for this learning outcome.

Evaluation 3: I think that I have proven an **Advanced** level for this learning outcome.

**Evidence Evaluation 1**

During the early parts of the project, I have created a Business Proposal in collaboration with my project group, for which feedback from the Process tutor (Martijn) was very positive, keeping the audience in mind. Furthermore, I have created a DVC research intended to document my findings and be able to share this with my fellow students for them to use it as a guide, also providing a README in my group project with instructions of how to use DVC within our project.

During the initial meeting with SUE, I have also taken a leading position in communication both during the meeting and general communication via email/google chat in which I showed my ability to communicate clearly with stakeholder.

**Evidence Evaluation 2**

I am still taking a leading position in my group project meetings, making sure that I am well prepared and know what general topics will have to be discussed. This became especially useful when having a meeting with a researcher from TU/Delft about the paper he wrote, as I was the only one to have actually read it before hand, this allowed us to ask insightful questions and make communication significantly easier due to the fundamental understanding on the (quite advanced) topics discussed in the research paper.

I also generally take the lead in meetings with the group project stakeholder, which the stakeholder does not seem to have any issues with, as I try to ensure that everyone is on the same page and that the stakeholder knows what we are talking about (which up until now seems to have worked, as I have only gotten positive feedback from the stakeholder and never got any requests to try and further explain a topic), this was also put into the group feedpulse by Martijn (Fig 1)

**A white background with black text

Description automatically generated**

Figure 1 Martijn Group feedpulse week 8

Besides my group project, I have also created a lot more documentation (including documentation on DVC and using GPU with TensorFlow on windows which seemed to have been useful resources for fellow students). Based on feedback received from teachers on my exercises I have also tried to keep improving my documentation within notebooks, which for the most part seemed to have been received positive, as shown by the feedback received from Olaf on group project work (on which I spent a lot of time getting a good structure while including work from all group members) (Fig 2), and my data challenge plan, which I was able to effectively go through and discuss with Qin (Fig 3).

**A screenshot of a phone

Description automatically generated**

Figure 2 Olaf feedback Exploratory Data Analysis

A screenshot of a phone

Description automatically generated

Figure 3 Qin feedback Data challenge plan

From all my work done so far to improve upon feedback received for Targeted Interactions, I think that I have proven my proficiency within this learning outcome.

**Evidence Evaluation 3**

While I have taken a step back from leading the conversations with stakeholders during group project deliveries, I still ensure that all necessary information comes across correctly and in an understandable manner for the participants, during deliveries I would regularly jump in the clarify any confusion.

During the final delivery for the group project, me and my group had created a streamlit application that could be used to better convey information both to the stakeholder in the form of a recap of the entire project as for the innovation insights event to steer conversations towards the topics that the person I was talking to was most interested in.

During both the midterm event as the innovation insights event, I ended up being the person to talk about the project with most people walking by and getting/keeping people’s attention for a project that is not as visually interesting (which was especially a challenge during the innovation insights event as the knowledge between people could differ completely).

I think I was able to communicate with people especially during the innovation insights event extremely well, as all people seemed very engaged when the project was explained within the context of their own personal life’s such as stealing private information or Netflix not being available when wanting to watch a movie, the best example of this was actually when an employee of Fontys came by that was more interested in the project’s methodology and approach for which she was so amazed, that later a different Fontys employee was recommended to come to our booth and asked if I could be a part of a promotional video they would be making for Fontys (sadly, I wouldn’t be available during the days that this would be happening so referred her to Tymofii instead who graciously accepted).

I also attempted to formulate my research documents, such as for my data challenge in a clear and understandable way so that people could refer to it in the future, as I am planning on (and currently in the process of) getting my custom web environment that my RL agent acts within referenced on the official [Gym third-party environment page](https://gymnasium.farama.org/environments/third_party_environments/), for which having clear documentation should be important. This extra work seemed to have been done correctly as Simona also had positive feedback about the report made around this (Fig 4).

A screenshot of a computer

Description automatically generated

Figure 7 Data Challenge report feedback – Simona

Another place where I tried my best to clearly convey very specific information to a variety of people (most not technical within the context of the work), was the data visualization challenge, where I had to find a way to create a quiz about data from Eindhoven, that was both engaging and they hopefully would learn something from. I decided to do this by wrapping the whole idea in a neat looking jeopardy game as most people have heard of or seen this on TV at some point in their lives to get them engaged, then keep them interested by offering the challenge of getting the highest score possible and then pulling a switcheroo by having the questions, while still about Eindhoven, in reality be a way for people to learn how data visualization techniques such as histograms, boxplots, scatterplots, etc. can be used. This implementation was also received passively by Simona (Fig 5), while I could have perhaps made the design of it a bit better, which of course proves the classic statement of software engineers can’t design anything.

A text on a white background

Description automatically generated

Figure 8 Data Visualization feedback – Simona

Lastly, I spent quite some time creating presentations for my open program and data challenge delivery as I knew that due to the amount of presentation after one-another and the fact that my work might not be inherently interesting, I tried to make the presentations as engaging as possible, which I feel worked out as I could see people keeping interest when bringing up topics or scenarios I specifically knew other people would have personal experiences with, and for which I got questions about the project from seemingly interested people, with Timofey even asking for the custom RL environment I had made, which is the best sign you can get when actively trying to work through the proper channels to get that exact environment published.

# Future orientation

*You can* ***identify and explore*** *the* ***context*** *in which your project takes place, and you can approach this from* ***multiple perspectives*** *within the machine learning domain.*

Evaluation 1: I think this learning outcome is currently at an **Orienting** level.

Evaluation 2: I think this learning outcome is currently at a **Beginning** level.

Evaluation 3: I think I have shown an **Advanced** level of proficiency for this learning outcome.

**Evidence Evaluation 1**

For the set-up of the group project, I ensured that a Business proposal was created that could clearly show our understanding and vision of the project, to ensure that no unexpected issues will be encountered later on, including setting up effective communication channels with the stakeholders.

Secondly I have spent a lot of time researching and setting up a robust DevOps environment for the group project that the stakeholder company can adopt even after the project in finished, using best practices and tools explored in my DVC research.

**Evidence Evaluation 2**

Since the previous evaluation, I have spent more time exploring both my group project and data challenge. For my group project, I have reached out to a researcher at the TU/Delft who co-wrote the research paper a lot of our work is based on. To ensure that a meeting with the researcher would be worthwhile for everyone involved, I had read both the paper in question and (most of) the related papers, which he cited in his work or had written as well to extend the research done. This proved extremely useful as it has given me far greater insight into the topics discussed and allowed me to ask far better and targeted questions to help the group project further.

For the group project, I had also done more exploration into the data as I had discussed possible ways of enriching the data for specific models (such as adding time-series data to a model that normally does not account for time dependent data), which ended up in the group project’s EDA and data processing, which both technical tutors mentioned showed potential (Fig 1).

A screenshot of a phone

Description automatically generated

Figure 1 Feedback (Highlighted) Olaf Exploratory Data Analysis

Furthermore, I have also spent time during my open-program and start of data challenge to explore the required context of my RL-powered integration test project, this has culminated in a project plan for the data challenge, which after discussing with the data challenge tutors was received very positively (Fig 2).

A text on a screen

Description automatically generated

Figure 2 Feedback Qin data challenge plan

The evidence shown so far is a good beginning for proving proficiency for this learning outcome, possibly showing a proficient level with more evidence.

**Evidence Evaluation 3**

The feedback received from evaluation 2 was: “Close to PROFICIENT, but add more about ethics perspectives next time.”, so to make sure I could prove a proficient, or even advanced level, I first made sure to spend more time on the ethical part of my project.

A white text on a white background

Description automatically generatedFor my group project, I had created the TICT quick scan and made sure to get a meeting with the ethics tutor to discuss this quick scan and get feedback to later perform a deeper ethics analysis, this first meeting was very positive with the tutor stating that the quick scan done showed that we had a clear idea of the context we were working within and had an understanding of the potential problems.

From this meeting, the project group did extended research into a majority of the TICT topics, where I focused on the Data and Inclusivity parts, for which parts the tutor had no further feedback (Fig 3), while some other topics could be considered more. I made sure to explore these contexts further while working on the interactive ethics page for the final streamlit report we had made as part of the innovation insights event, where people interested in the ethics part could easily view what we had done.

I also explored the ethical concerns of me innovative data challenge into fairly large detail, which Simona had given positive feedback on (Fig 4).

Figure 3 group project ethics feedback – John (cropped)

A close up of a text

Description automatically generated

Figure 9 Feedback Simona, Challenge Report

Furthermore, a large part of the group project consisted of exploring what all was possible within the context of anomaly detection, for which I had spent a lot of time exploring, things such as the general DevOps approach in AI which led to the exploration of tools such as DVC, the various techniques of processing data and the development of specialized models within this context such as TranAD. All the exploration within the context of the group project came to a conclusion during the final delivery, in which the stakeholder summarized our work into a simple sentence: “Other projects look nice, your project **is** nice” emphasizing that while our project might not have been super visually appealing, we were able to explore it from many angles gaining invaluable results that can help SUE further on this project, and also in their own independent AI research.

This way of approaching whatever problem in an abstract way, thereby being able to explore it from many different perspectives I tried to apply into all my projects, which I feel was perfectly acknowledge in feedback received from Simona on my Data visualization challenge (Fig 5).

A text on a white background

Description automatically generated

Figure 10 Data visualization challenge feedback - Simona

# Investigative problem solving

*You can Identify a problem and create a* ***hypothesis*** *and* ***research question*** *and can do this using a sound* ***research methodology****.*

Evaluation 1: I think this learning outcome is currently at an **Orienting** level.

Evaluation 2: I think this learning outcome is at a **Proficient** level.

Evaluation 3: I think I have proven an **Advanced** level of proficiency for this learning outcome.

**Evidence Evaluation 1**

I have created a research document, exploring DVC. This research document is done using the DOT framework methodology, including the machine learning extension specifically added for the machine learning context.

This research document has been shown to the Process tutor (Martijn) who was positive about the methodology and research done.

**Evidence Evaluation 2**

After the first Evaluation, Martijn gave me a Beginning instead of my self-graded Orienting, due to the amount of work I had already done within the group and core program (Fig 1)

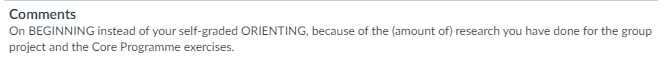


Figure 1 Martijn grade reasoning for Investigative Problem solving

Since that Evaluation, I have done a lot more research into many different topics such as the Reinforcement learning exercises, which also extend into my open/data challenge. Iman ended up giving me very positive feedback on my approach to the Reinforcement Learning exercise, specifically on my methodology for problem-solving and approaching a problem (Fig 2).

A screenshot of a computer

Description automatically generated

Figure 2 Iman feedback RL - Exercise

I used the same approach for all my other exercises, where if I ran into an issue, I would research the problem and create a hypothesis of what is causing it, with sometimes even being able to solve the issue afterwards. Two notable instances, were my CNN exercise and XAI (as I was able to prove some hypothesis of my other models during my XAI), my successful application of my approach can be seen in the positive feedback received from Olaf (during meetings in person) and Nico (See Fig 3/4 and in person discussions).

A screenshot of a phone

Description automatically generatedA screenshot of a text message

Description automatically generated

Figure 4 Nico feedback XAI – Exercise

Figure 3 Nico feedback (highlighted) CNN - Exercise

Furthermore, I had also created a plan for my data challenge, with contained a condensed and simplified version of my research/results done so far on the topic as a part of my approach, for which I also received positive feedback for by Qin (Fig 5).

A screenshot of a phone

Description automatically generated

Figure 5 Qin feedback Plan: Data Driven Innovation Challenge

Due to all the work done since my last evaluation, and the proficiency shown throughout many of my projects, I feel confident in my Proficient level for investigative problem solving.

**Evidence Evaluation 3**

During this final evaluation phase, I spent a lot of time on finishing up the various projects I had started, which also included discovering if my research approach had worked or not. The biggest showcase of this was my Innovative Data Driven Challenge, as this project always had a large elusive question mark hovering above it, never knowing for sure if the idea I had would end up working.

During the peer review of status update 1, while both peers did not leave any written feedback sadly, the ratings they gave were overall positive, indicating that the (almost exclusive) research I had done so far was on the right track (Fig 6 & 7).

A screenshot of a computer

Description automatically generatedA screenshot of a computer screen

Description automatically generated

Figure 7 Peer feedback status update 1

Figure 6 Peer feedback status update 1

This was later corroborated Qin’s feedback on status update 2 (Fig 8).

A green arrow pointing to a green line

Description automatically generated

Figure 8 Feedback Qin, Status Update 2

After status update 2 however, I was finally able to put all my research to the test and implement it into the web environment and RL rainbow model/agent, which based on the feedback received during feedback meetings, the final presentation, and the eventual report of my research (Fig 9) went as best as could be expected.

A screenshot of a text box

Description automatically generated

Figure 9 Data Driven Innovation Report feedback – Simona

The other project that I have finished during this final phase, is the group project for SUE, within this project I had spent a lot of time researching various topics, most notably, the state machines and anomaly detection transformers. The work I had done on this project’s research was received extremely well by the stakeholders (Nathan from SUE specifically), who ended our final delivery stating that the work we had done was more than he could have asked for and could be used as an exceptional stepping stone for future work, due to our expensive research.

Two smaller exercises that I had to do quite some problem solving on was the NLP exercise and the data visualization challenge, as for both of these exercises I wanted to create some kind of meaningful result that fitted within a given context.

Firstly the data visualization challenge, as this one’s goal would be for other people to check it out and learn something about Eindhoven and also data visualization along the way, my specific challenge was to make a quiz, but how do you make people interested in a quiz all around statistical data (that probably most people don’t know), so I decided to make a form of Jeopardy that dynamically generates the questions based on these statistics, which from the feedback received from Simona ended up being a perfect solution for this abstract problem (Fig 10).

A text on a white background

Description automatically generated

Figure 10 Data Visualization feedback - SImona

Secondly, for the NLP exercise I wanted to try and integrate some kind of NLP pipeline into a real world use-case, the use-case I chose was a procedural text adventure web game I had done for a previous semester, at the time I had created a composition algorithm to generate deterministic text, but now I wanted to explore the feasibility of using NLP/some kind of LLM for this, for which the research ended up in a proof of concept where it was directly integrated into the front-end’s game pipeline where it converts the deterministic text into natural text (a showcase video [can be found here](https://www.youtube.com/watch?v=Hgaf55zL9Xc)), this research/implementation was received positively based on Olaf’s feedback, while I probably could have spent a bit more time properly writing out the research I did (Fig 11).

A text on a white background

Description automatically generated

Figure 11 NLP feedback - Olaf

# Personal leadership

*You are* ***proactively*** *looking to improve your own* ***development*** *and* ***learning****, while taking a* ***leading position*** *in your projects.*

Evaluation 1: I think this learning outcome is currently at an **Orienting** level.

Evaluation 2: I think I have proven a **Proficient** level of proficiency for this learning outcome.

Evaluation 3: I think I have proven an **Advanced** proficiency for this learning outcome.

**Evidence Evaluation 1**

From the start of the semester, I have tried to ensure that my work can properly cover the learning outcomes set out for this semester. Secondly, I have taken a leading position within the group project, often leading meetings both within the group and with stakeholders.

Furthermore, I have spent significant time performing research and exploring machine learning within my own time to try and further understand topics that might not be covered by the normal curriculum (excluding the Open Program). I have also tried to be very proactive in asking for feedback on my work done, giving me valuable knowledge and ideas from both tutors and fellow students.

**Evidence Evaluation 2**

I have tried to have meetings with tutors and other stakeholders as often as possible to get feedback on my progress so far, which has helped me quickly adapt and learn when for example making incorrect assumptions about model performance.

During the group project, I am still taking a leading role in most meetings and outside of the meetings, including making sure assignments are submitted on time, and feedback/standups are done regularly, allowing the group project to run smoothly and make constant progress.

During this semester, there were two things that I wanted to improve on further with the first one being ethical concerns. As I knew this was a topic that would be covered later during this semester (and was covered for the Software domain during advanced software), I decided to get a jumpstart and already investigate potential ethical concerns in my open program, which was then transferred to my data challenge plan. I discussed the ethical concerns with Qin and received positive feedback on my approach (Fig 1) and was offered tools to look at further.

A close up of a text

Description automatically generated

Figure 1 Qin Feedback data challenge plan

One thing that I have always been bad at, is approaching people/companies which during my preparation for future after Fontys, was identified as one of the biggest potential failing points if I were to start working as a ZZP’er. To try and improve on this skill, I decided to go to the Career day at Fontys TQ, for this I had created a QR-code that linked directly to my CV and had created (sort of a minigame) challenge of getting as many companies as possible to scan the QR-code and look at my CV. As expected, at the start of the day, I was very hesitant to show it as I could not naturally work it into the conversations. But after talking with InfoSupport who were extremely excited when I mentioned I had a CV that they could get via a QR-code (instead of a piece of paper), I became far more confident in not only just asking people to view my CV, but in talking with companies in general (eventually leading 10+ companies to have scanned my QR-code and viewed my CV, which has now caused most companies to contact me to mention this fact meaning that it might have actually helped me stand out).

Due to all the progress and continued leadership, I have shown during this phase, I think I have shown a sufficient level of personal leadership.

**Evidence Evaluation 3**

Just as with the previous evaluation periods, I have continued to take a leading position within the group project. In fact, I feel that I have had to take on an even bigger role during this phase, being more mindful of deadlines and ensuring that the other group members had their work done on time, which I can understand, as a lower level of enthusiasm and motivation is expected towards the end of the semester, but this did cause me personally to spend significant effort ensuring that all deadlines, meetings and general work got done, which seemed to be noticed by the other group members (Fig 2 & 3).

A close-up of a message

Description automatically generated

Figure 12 Anonymous peer feedback

A red line between black text

Description automatically generated

Figure 3 Anonymous peer feedback (cropped)

During the evaluation period, we had two peer review moments, where from the first one I learned that due to my leading position during meetings with stakeholders, other team members did not feel the need to actively join in the conversation thereby causing them to struggle proving these learning outcomes which was not my intend (Fig 4). Based on the feedback received, I made sure to more clearly communicate with my team members how they would like to help prove their learning outcomes and how I could help, actively taking a step back from most meetings with stakeholders thereby giving other people more of a chance to speak.

A close up of a text

Description automatically generated

Figure 4 Anonymous feedback peer review - 1

My approach, seemed to have been accepted positively, as the other team members felt more confident in their ability to prove their learning outcomes, and did not have to repeat their sentiment on this in any later feedback.

Secondly, ethics was still a part that I had generally not spent too much time on, so I wanted to make sure to cover it more during this semester, this started with automated web testing leveraging RL, where I had explored potential ethical concerns, which Qin mentioned was a good starting point and could be explored further during my data challenge (Fig 5).

A close up of text

Description automatically generated

Figure 5 Feedback Qin, Data Challenge Plan

I later followed up on this in my eventual data challenge report, where my ethical analysis part was given positive feedback by Simona (even liking my slight tangent about a science fiction-y scenario) (Fig 6).

A close up of a text

Description automatically generated

Figure 6 Feedback Simona, Challenge Report

I continued my ethics journey in the group project, where I even had to track down the ethics tutor twice after he seemingly left while only meeting with half the groups, and a second time when he didn’t give feedback on our canvas submission even though stating during our meeting that he would and even announcing he had done so on canvas. I had done a majority of the TICT quick scan so that it could be discussed with the ethics tutor during our first meeting, which he responded to positively, stating that we had clearly explored the domain. After this meeting, I had done a more extensive exploration into the Data and Inclusivity TICT topics (as a part of a larger extensive exploration by the entire group), for which my part the ethics tutor had no further remarks during the submission, while any feedback he did give on other topics, I made sure to integrate into our final group project’s streamlit report with an interactable ethics exploration page (Fig 7).

A screenshot of a computer

Description automatically generated

Figure 7 Interactable ethics (cropped) page within streamlit, the TICT icons can be used to navigate between topics

# Graduation Preparation

*You create* ***opportunities*** *for finding a* ***graduation internship****, that aligns with your* ***personal needs*** *and the* ***school’s requirements****.*

Evaluation 1: I think this learning outcome is currently at a **Beginning** level.

Evaluation 2: I think I have achieved a **Proficient** level for this learning outcome.

Evaluation 3: I think that I have been able to reach and **Advanced** level, however I know only **Proficient** is given.

**Evidence Evaluation 1**

Before the start of the semester, I have already spent a lot of time building up an understanding of what I want to do after graduation, including interviews and domain explorations and improving my outwards appearance for companies by improving my LinkedIn, GitHub, CV and creating a portfolio website.

I already have created a full planning for finding a graduation internship, including fallback plans which can be found in my Open Program Plan.

The work done so far has already seemed to pay of with me getting semi-daily emails and LinkedIn messages from new companies and previously approach companies both for working opportunities and internships.

**Evidence Evaluation 2**

Since the first evaluation, I have done a lot of work on finding a graduation internship (most of it directly following the previously made planning).

First thing I did was create a spreadsheet containing a list of companies I am interested in, either from LinkedIn or the ASAM portal with all required information. I have been using this spreadsheet to keep track of the status and remind me of my previous opinion of a company (Figure 1).

Internship candidates spreadsheet


Figure 13 Internship candidates spreadsheet

From this spreadsheet, I had reached out to my T1 (green) companies, sadly these companies did not have (or want me for) an internship, so I moved on to my T2 (blue), while also responding to all companies that had contacted me on LinkedIn (even the ones offering me a job instead of internship).

During this period, I was still trying to get an internship with SpendLab (my nr1 candidate), as they were not responding to my emails making me contact them via other means and eventually getting into contact with the previous recruitment manager who had apparently told their successor to reach out to me (which he never did). After getting into contact with the new recruitment manager, he told me they had filled all internship spots, but forwarded me to a different company (Waves Process Intelligence) (while he mentioned he did that because of my excellent portfolio, I have a feeling that he might also feel bad that I could have gotten the internship if he had just replied to me earlier).

After corresponding with Waves Process Intelligence, and having an in-person interview I was offered 3 possible projects to do for them (which all of them are very interesting, and I even offered to share them with more Business/AI students), meaning that I now have a company I can do my graduation internship at. The company however, is very small so for now I am still looking for other companies knowing that I have a fallback company that I can always do the internship at.

I also went to the Career-day at TQ, here I talked with many companies (including some that did not respond to my emails…). As preparation for the Career-day, I had made my CV into a QR-code and shared it with companies during the day (companies were surprisingly extremely excited about this), which got me an interview with Info Support, BDO and Alten in the coming week, and correspondents with others.

After all this work, I have been able to find a fitting graduation internship that I would like to do, but have also been able to get into contact and be offered interviews with more companies (as I would still prefer to do an internship at a “large” company to get experience within that environment).

**Evidence Evaluation 3**

Since the previous evaluation, I have been offered internships by all companies whom I was offered an interview with.

After much debate and discussion with several teachers who had experience tutoring students of these various companies, I decided to do my internship at Info Support, as they offer the best growth and learning potential, especially when considering the freedom offered by their consultancy driven business model, alongside their well known reputation for quality (also Nico said I would be stupid not to choose them as apparently not many Fontys students are accepted).

Info Support, is offering me a very interesting project to solve a real-world problem they currently have with the billing process that has to take place from when their consultants register worked hours for a project, the client’s agreement on these hours and the eventual invoice. Besides the very real world problem I will get to solve, I will be set as the development lead for the project (due to my prior experience in this role at my previous software job) and will be allowed to manage the entire project’s structure (such as methodology, meetings, etc.) as I see fit, which I greatly appreciate and think will help me develop my skills even further.

Due to all my prior work done to try and find the best kind of internship for me, and the best company. I think that I have not just found a suitable graduation internship, but have found the best possible one for me.

*Note: I have signed an internship agreement contract, thereby fulfilling the set requirement for a proficient in this learning outcome.*

# Retrospective

At the start of the semester, I had no clue what I would be learning as I didn’t know if it would just be a glorified repeat of AI-4 with a bit of neural networks, or if we were going to learn the newest technologies out there, dissecting ChatGPT and whatnot.

Turns out, it was pretty much up to me which way the semester went, and I inadvertently created a self-fulfilling prophecy by dissecting anomaly detection transformers.

During my previous Advanced Software semester, while still challenging, I never felt I had too little time or that I was possibly falling behind, even though at the time I was also spending 12 hours a week during this school time working at a software company. During the previous semester, I could always reason that the semester itself wasn’t actually 40 hours a week, but I made it 40 hours by combining my working hours, and any extra time I would spend during weekends or vacation was just for extra credit, while I kept working the same amount of hours at that company during the AI semester, it became clear very quickly that this semester would actually be 40 hours, forcing me to work through most of my weekends just to keep the progression that I had planned to need to finish all my projects.

I had a lot of split goals during this semester, such as finding the perfect internship, fundamentally understanding the more complex modelling techniques that I did not cover during AI-4, create an innovative project to work on throughout the semester, etc. While I was able to achieve most of these goals to some degree, it might have saved me some stress and late nights by limiting what I had set out for myself.

For the group project, we had been given a research paper from the TU/Delft, as a good starting point for our project, like the other group members I only skimmed through the research paper as it seemed very daunting never fully understanding it. As preparation for the interview/meeting with the researcher who had written the paper, I finally fully read through the paper (and a few times after it), at which point I came to realize that the discoveries we had been making up until that point, the researcher had also made and was eluding to in the paper, making me question how much time we could’ve saved if everyone had sat down at the start of the semester and collectively read through the paper completely.

In retrospect, I would probably say that I should have dropped the transformer anomaly detection research as there was no feasible way to get the proper results within the timeframe of the semester, and this did end up being true to some degree (as I did not have time left to train the model on the whole dataset, or even tune any parameters), I do think that all the research I did has given me an extremely good understanding of the current most hot technology out there, which is invaluable.

One thing in retrospect I do regret however, is that in previous semester, I always asked teachers to create feed pulse moments after a meeting as this was the norm set by the teachers themselves. At the start of this semester, I had asked some teachers what they preferred, and most mentioned that they already keep notes themselves so they wouldn’t care about a feed pulse, so I never bothered writing them either, however this did make it significantly harder to get “real” proof for the PDR, and I would definitely do choose to do differently in retrospect.

# Conclusion

I had decided to do Advanced AI after I had done Advanced Software even though I knew I would be pursuing primary job paths in software after graduation; as there were no minors that seemed appealing to me for learning anything that I could possibly use in my future career, while AI could offer me a deeper understanding of the underlying algorithms of AI models.

One of my goals during this semester was to get an Outstanding final grade, as I have somehow gotten one every semester before this, and it would be a shame to stop now so close to the end. This in combination with what felt like genuine excitement by the various teachers on the topics they would cover, made me try to always go further than was expected.

This mindset did pay off in many ways I did not expect, as a surprising number of companies were interested in the work that I had been doing with AI. Such as one company that had been trying to do something similar to my data challenge, but had gotten stuck at a point I had already gotten past, at which point the interview switched to a consultancy session with me giving my opinion on various problems and ideas they had, proving the power in the combination of software and AI. I in fact later learned that I had become slightly infamous with the dad of an intern I had previously tutored contacting me as he was a manager at that company, after hearing I had chosen for a different company.

While I probably spent way too much of my time diving even deeper down rabbit holes than I had ever done before, I did end up having happy stakeholders for the group project and amazing results for my own projects that I could not be happier about.

I feel that throughout the semester, I have gone beyond what was part of the regular curriculum, and have personally accomplished my goal of reaching what could be considered an outstanding level of proficiency throughout the semester’s various learning outcomes.