**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Artificial Intelligence  Data Visualisation & Communication |
| **Assessment Title:** | L8\_CA\_2\_AI\_DV v5 |
| **Lecturer Name:** | David McQuaid  Sam Weiss |
| **Student Full Name:** | Rayen Bentemessek |
| **Student Number:** | 2021378 |
| **Assessment Due Date:** | 29/12/2024 |
| **Date of Submission:** | 28/12/2024 |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

Table of Contents

[Artificial Intelligence: 3](#_Toc186117184)

[Part 1: 3](#_Toc186117185)

[Scenario 1: 3](#_Toc186117186)

[Scenario 2: 4](#_Toc186117187)

[Part 2: 5](#_Toc186117188)

[How using Constraint satisfaction finds an answer to the problem: 5](#_Toc186117189)

[Difference between CSP and standard algorithms in finding solutions: 6](#_Toc186117190)

[Part 3: 7](#_Toc186117191)

[Data Visualisation: 8](#_Toc186117192)

[CSP Visualisations: 8](#_Toc186117193)

[Alternate scenarios – Interactive Visualisations: 10](#_Toc186117194)

[References: 19](#_Toc186117195)

# Artificial Intelligence:

## Part 1:

### Scenario 1:

In order to solve this problem, I made sure to choose the roles as my variables, then the combination of teams as my domains. This approach was helpful to avoid redundant solutions and get only unique ones and combinations of employees. Because the computer will count “Peter, Juan” and “Juan, Peter” as 2 separate solutions while it’s not as we only need 2 people for AI in scenario 1 for example. That is why we only get [“Peter”, “Juan”] as the only solution that can have these 2 persons.

Then I’ve created a function called “constraints” that will ensure I only assign 2 roles as a maximum per person because it’s a hard constraint. I did it through a dictionary called “employee\_counts” that will keep iterating for roles, adding employee counts and return true if maximum value of the dictionary is less than or equal to 2. Within the same function, I have ensured we only hire 4 people (excluding Ciara) which is our second hard constraint using a set of unique employees. The logic behind this is I treated hiring Ciara to be a soft constraint, because “she knows python “which doesn’t mean she has to be hired and it’s worth exploring solutions where she is not hired because she is the recruiter, and she may change her mind. I only focused on the fact that she has the funds to hire 4 people, and it was doable without hiring her in 2 solutions and she was present in 52 solutions.

A screenshot of a computer

Description automatically generated

To have a better layout, I created two lists, one where Ciara is hired and the other where she is not and added my solutions to the relevant lists, after that I printed my solutions in both cases, and I showed which person is hired for each position.

I have to mention that I have considered solutions where the same people are hired but in different roles to be different because I got different combinations in the roles which I considered to be different because it satisfies all the constraints that we have and it’s unique at the same time which means no other solution will have the same repartition but we may have the same people to hire as shown in the following picture.

A screenshot of a computer

Description automatically generated

### Scenario 2:

For scenario 2, it was pretty similar to scenario 1, because we needed one more AI Engineer and a security employee which is not critical (soft constraint) so I have added more combinations in my AI variable, and added a security variable with the possible employees (Mary and Maria) and an empty list [] which represents the option of not hiring security. I’ve also thought that Ciara knowing python to be in scenario 2 and I hired Juan at first ( which made me got 1014 solutions ) but then I changed it and realised that the 2 scenarios are independent, and Juan becoming a partner means he is no longer a candidate to consider so I’ve removed Juan from the possible employees and also removed Ciara from the Python variable. This resulted in 5 solutions, 4 of them not having a security and one where a security employee was hired as shown in the following picture.

A screenshot of a computer

Description automatically generated

## Part 2:

### How using Constraint satisfaction finds an answer to the problem:

The constraint satisfaction problem is basically a mathematical problem where it requires certain solution for certain variables, this solution must meet all the constraints. To do that we need first to start by defining:

**Variables**: These are the objects that need values assigned to them to satisfy the constraints. It can be using any type of variables such as Boolean, Integer, Strings etc. And they are finite as well (V1, V2…Vn). To be more specific, in our case where we need to assign employees to certain roles, I chose to define my variables as the roles: Python, AI, Web etc. It’s important to mention that this is not the only way to do that and there is not a “correct” way to do it as well, but whatever the choice of variables is, the domains must follow it (GeeksforGeeks, 2023).

**Domains**: The domains are basically the range of values that the variables can get and they cannot be empty because each variable needs one value at least, for example in my case Since one of the variables is Python, the values are the possible employees that can get the roles, like Ciara, Jane, Peter and Bruce. In other words, I’m telling my CSP that for Python you can only assign 3 of the 4 employees, I also chose to that in terms of combinations to avoid potential redundancy, for example: Ciara, Peter is the same as Peter, Ciara but the algorithm can’t tell so I forced certain combinations as shown in the picture below to avoid that potential issue (GeeksforGeeks, 2023).

**Constraints:** Constraints are basically the limitations or the conditions that give me meaning to the previous variables and domains. In our case, in scenario 1 for example, we can only hire a maximum of 4 people, and each employee can take a maximum of 2 roles, those are constraints that must be met to solve the problem, they’re also called hard constraints. There are also soft constraints where they’re necessary but it’s good to have which is for example, ‘Ciara knows how to code’ doesn’t mean she has to be part of the team or hiring a security employee in scenario 2.

Now after identifying the components of the CSP, we’re getting one step closer to highlighting how it works, so after we defined the variables, domains and constraints here comes the beauty of CSP as everything is done in the background using these techniques or algorithms:

* **Backtracking Algorithm:**

This algorithm plays a major role in the efficiency of the CSP and how fast it returns a solution, because as soon it faces an issue or solution that doesn’t work, it discards it and goes back or backtracks to the previous path. This is done for all the possible extensions of the solution and then it keeps going, if the solution meets the requirements, it is valid, if it does not, we don’t have to continue and we go back instead (GeeksforGeeks, n.d.).

Now let’s see how see this works in our case:

Say for the Python team we chose Ciara, Peter, Jane, we move to the AI team, and we choose Juan, Jim. remember that we already chose 4 people, so if we add Mary as a web developer, one of the constraints is violated so here the program backtracks and chooses Anita, but we have the same issue, so we must choose Juan as a web developer. Then for the database, we choose Jim, when we move to the Systems, if we choose Juan, we have another issue because Juan is hired twice, same thing for Jim. The backtracking algorithm will flag that this solution is incorrect, and the solution will be marked as invalid.

* **Forward-checking algorithm:**

This algorithm is basically an optimisation of the backtracking algorithm, as this algorithm will keep track of assigned values and dynamically remove them (temporarily ) from the domain list for the specific variable, in other words, in scenario 1, if Juan has already achieved his limit of 2 roles, we don’t assign and then backtrack like we do in backtracking, but we Just remove Juan from the possible role leaving less options but with less issues, in case no domains are left that means the solution is invalid and the algorithm backtracks (Cuni.cz, 2024).

* **Constraint propagation algorithms:**

Constraint propagation is a major concept in CSP, as it works by narrowing down the domains of variables based on constraints. This approach will help in reducing the search space as it ensures consistency across the different components, nodes consistency for example. So, it will eliminate the values that violate the constraints before doing the search. In our example it will initialise the domains for each role and constraints are applied simultaneously and iteratively to eliminate possible confusions, for example hiring someone for 3 roles which cannot be, and it keeps going until the solutions are found (GeeksforGeeks, 2024).

### Difference between CSP and standard algorithms in finding solutions:

In this discussion, we will talk about the differences between CSP and standard algorithms in solving problems, the main question is not can they do it? Because they both can. But in order for a CSP to find a solution, we need to define variables domains and constraints, then the algorithm will do the rest using -not limited to- the previous techniques and algorithms which will eliminate the invalid paths as soon as they’re faced which will save tons of processing power and time as well. However, the standard algorithm will require defining all the variables on their own, than implement tailored solution for the problem which can be good for some problems, but if it’s not implemented the right way, some disasters can happen. Speaking from a personal perspective where the running time for one cell took 4353 seconds to run as shown in the picture below and the bigger the problem, the less efficient it becomes. So, on one hand, CSP will guarantee reliable solutions if it’s implemented the right way with great efficiency which will avoid invalid solutions early, on the other hand, standard algorithms give more freedom in the implementation, but it may struggle in problems as it will explore the invalid paths unnecessarily which will make it slower. So CSP works much better in these kinds of situations the same way standard algorithms are amazing in most of the application cases but here it’s like comparing a Porsche 911 to a Land Rover Defender, they’re both amazing but one is a sports car and the other is an off-road SUV.

A screenshot of a computer

Description automatically generated

## Part 3:

In order to implement this algorithm, I started with A\* but the results were horrible, and I got too many solutions that I didn’t need, also the program was so heavy as shown in the previous picture and took ages to run and I found some issues defining the heuristics. So, I decided to switch to Depth First search, with its nature of exploring all possible combinations it was so helpful solving my problem. DFS was much more efficient in terms of memory usage and I found no issues with it while implementing because the time complexity is dependent on the roles and the possible combinations which was manageable because we only needed 6 roles maximum with limited predefined combinations which contributed in optimising its time efficiency.

I created 2 dictionaries for each scenario with the key being the role and the value as the possible combination. Then I implemented two functions, one for each scenario with a very similar approach to the CSP constraints functions to make sure no one is hired more than 2 times and respect the number of employees based on the scenario.

Then I implemented my DFS function, which will take roles from the combinations dictionary mentioned earlier, I’ve also iterated through a stack that starts with an initial state which is a tuple containing “index”, “current\_solution” and “employee\_counts” and then we use LIFO to work with the current state. Then we just start assigning the roles from the dictionaries with the combinations of employee making sure to get their employee\_counts and creating a new solution based on the combinations. Once we’re done, we create “next\_state” tuple with the updated components and we add it to the stack, if our index meets the length of our roles, we check it with the scenario function, if it is satisfied, we got a new solution. If not, we just keep iterating. Then we return the found solutions.

Everything here went according to our plan because we found the same number of solutions using DFS as we did with the CSP framework.

A screenshot of a computer code

Description automatically generated

A screenshot of a computer code

Description automatically generated

# Data Visualisation:

## CSP Visualisations:

In order to visualise the problem and effectively communicate the CSP scenario, I chose at first to display the role vacancies for the two scenarios using a bar plot as shown in the following picture.

A graph of different positions

Description automatically generated with medium confidence

Why a bar plot? Bar plot is a great chart to display the distribution of data points or comparing numeric values across different subgroups (Yi, 2024). In our case, the roles will be considered as different subgroups or categories let’s say, and each role has its own vacancies so for scenario 1 we will need 3 Python programmers, 2 AI engineers, 1 Web developer, 1 Database engineer and 1 systems engineer. The choice of bar plot can efficiently communicate the needs for Ciara’s company. In order to do the bar plot, I used the keys and values from my combinations of roles that I used in the AI code, so I just called them and then created a dataframe for each scenario with roles and number of vacancies being my columns and then used matplotlib for visualising them.

A screenshot of a graph

Description automatically generatedNow after visualising our vacancies, I chose to display the abilities of each employee, in order to do this, I decided to go for a heatmap as shown in the following picture.

The heatmap will communicate each employee with its own abilities, the data was literally taken from the assignment description as it is. In order to do heatmap, I used an abilities matrix for each scenario, where 1 means the employee can do that role, and 0 means they can’t. then changed the matrix into a NumPy array, I also created employees list and roles list for each scenario. Then I used matplotlib to visualise it, I’ve also added a colorbar for better understanding where dark colour means 1 and light colour means 0, which was inspired from chess board design. I tried to avoid extremely dark colour or bright colours and went for more neutral colours to make it distinguishable and more comfortable for the eyes to look at.

I also added annotation to my cells where values are also displayed in addition to the colour bar, and in order to do the annotated heatmap using matplotlib, I followed the matplotlib documentation (matplotlib.org, n.d.).

Then after discussing the vacancies, employees’ abilities and in order to fully deliver the CSP scenarios, I decided to visualize a random solution from each scenario.

A graph of different colored squares

Description automatically generated

This plot is simple and straight to the point, roles will be on the x axis, number of roles on the y axis, and each colour belongs to a unique employee, for example Bruce is green as shown in the legend beside it, so Bruce can take Python and Systems roles and same applies to all the employees. Since this was a random solution, we see that no bars in scenario 2 so that means this random solution didn’t include a security employee. The choice of colours here was based on diversity and staying within the same theme, so I tried to get different and distinct colours to avoid similarities, so all colours are unique and visually balanced as well making it easier to spot the differences without being too harsh on the eye by avoiding bright colours. In order to achieve this, I used Set3 palette to make my colours visually appealing and suits the context of a stacked bar chart.

The code was standard stacked bar chart with 2 subplots, I used random to generate the random solutions with a fixed seed, then a dataframe for each solution that takes roles as rows, employees as columns and then iterate through each employee assignment to a certain role and changes the value to 1 (all values are initialised as 0’s at first).

## Alternate scenarios – Interactive Visualisations:

For the interactive visualisation, the task consisted of allowing the user to explore alternate constraint scenarios, so it means I need to let the user choose which scenario they want to display (2 options; scenario 1 and 2 ) and then each scenario can be split into 2 groups as well. Scenario 1 can be split to solutions where Ciara is hired, and solutions where Ciara is not hired.   
Scenario 2 can be split to solutions where security employee is hired and solutions where security employee was not hired.

In order to do this, I had to research what kind of plot would be best, because I thought the common plots would not meet the task requirements and my expectations. Because I wanted to display relevant employees and their roles for each group and each scenario. An employee can have more than one role assigned to them, and some roles can have more than one employee as well which means we are in a many-to-many relationship. This is why I decided to go for a Sankey diagram using Plotly which will guarantee 2 things: interactivity and showing roles assignment with flow counts (python incoming flow count will be 3 for example, which means we hired 3 Python programmers). Following this principle, my nodes will be the roles and employees. In order to do the Sankey diagram, I followed the Plotly documentation to define my nodes, links and get the overall syntax that I adapted to my situation (plotly.com, n.d.). But then I faced an issue because I have a big number of solutions that I want to move through, so what I did is adding Dash which a framework built on Plotly that helped me add more interactivity and different scenarios exploration, once a user chooses a certain solution it dynamically updates it using callbacks. The process was challenging because I had 4 sub-scenarios to work on and many solutions to get my outputs from but I followed the Dash documentation especially for the dropdown menu which was basically HTML, but the concept of callbacks was a little bit complicated at first (dash.plotly.com, n.d.).

The following 4 visualisations are an example of my interactive visualisations, I decided to only use one type of visualisation because Sankey Diagram provides deep exploration of alternate constraint scenarios because it allows the user to choose the case they want to explore with all the possible solutions found.

A grey and white lines

Description automatically generated with medium confidence

**Scenario 1 - Solutions with Ciara**

A screenshot of a computer

Description automatically generated

**Scenario 1 - Solutions without Ciara**

**Scenario 2 - Solutions with Security**

A grey lines with white text

Description automatically generated with medium confidence

A group of grey lines

Description automatically generated

**Scenario 2 - Solutions without Security**

## Alternate Scenarios – GUI:

For my GUI, I wanted to keep my interactive visualisations and add them to the GUI, this was not possible for me using Tkinter because it offers very basic interactivity. Instead, I decided to modify my Sankey diagram code and make it usable, so I put it into functions that I can use later on. So, I created a function get\_layout that keeps the same layout of my Sankey diagrams, a Sankey\_diagram function that takes the dash app and solutions as arguments when it’s called. I also used very basic and simple bootstrapping, recreated the dash app that I have previously created with slight changes for the GUI (in terms of layout) then I run the app using a different port to make them run at the same time. Once that is done, I used threading for simultaneous running, and then I used Pywebview for my GUI which suits my case perfectly because it’s based on HTML, CSS and JavaScript so it will work perfectly with Dash being a web-based application. This will guarantee I keep the interactivity and functionalities of my visualisation and my GUI, and at the same time help me display them on GUI that has a basic and user-friendly interface. The webview will take the localhost URL once the app is running, <http://127.0.0.1:8051>.   
Pywebview implementation was also done by following the documentation, mainly the create\_window part (Flowrl.com, 2024).

A black and white line

Description automatically generated with medium confidence

And that is how my GUI looks like in the end. I wanted to keep the GUI as simple as possible following the requirements, so the user can choose any solutions they are looking for and then if they want to close it, they just hit the close button, and they will get a prompt asking if they’re sure to exit. I believe that based on my solutions and my scenarios that’s everything I would be looking for from a graphical user interface of visualisations. Just to see all the possible visualisations and then close it.

One last important note that needs to be mentioned, the assignment has isolated the interactive visualisation part from the GUI part. To make sure I follow the grading criteria, I decided to show the interactive visualisations on their own at first, and then show these visualisations in the GUI later on to make sure each section will be evaluated on its own. I could’ve merged them both in the same cell and avoid repeating some parts of code, but I prioritised following the grading scheme criteria for each section over code optimisation.

Also, the theoretical AI team part was implicitly included in each visualisation where I explained my rationale for charts choices, colours and layouts whenever necessary.

# References:

GeeksforGeeks. (2023). *Constraint Satisfaction Problems (CSP) in Artificial Intelligence*. [online] Available at: <https://www.geeksforgeeks.org/constraint-satisfaction-problems-csp-in-artificial-intelligence/>.

GeeksforGeeks. (n.d.). *Backtracking Algorithms*. [online] Available at: <https://www.geeksforgeeks.org/backtracking-algorithms/>.

Cuni.cz. (2024). *Constraint Guide - Constraint Propagation*. [online] Available at: https://ktiml.mff.cuni.cz/~bartak/constraints/propagation.html#FC [Accessed 13 Dec. 2024].

GeeksforGeeks (2024). *Constraint Propagation in AI*. [online] GeeksforGeeks. Available at: https://www.geeksforgeeks.org/constraint-propagation-in-ai/ [Accessed 16 Dec. 2024].

Yi, M. (2024). *A Complete Guide to Bar Charts*. [online] Atlassian. Available at: <https://www.atlassian.com/data/charts/bar-chart-complete-guide>.

matplotlib.org. (n.d.). *Creating annotated heatmaps — Matplotlib 3.5.2 documentation*. [online] Available at: <https://matplotlib.org/stable/gallery/images_contours_and_fields/image_annotated_heatmap.html>.

plotly.com. (n.d.). *Sankey Diagram*. [online] Available at: https://plotly.com/python/sankey-diagram/.

Plotly.com. (2024). *Dash Documentation & User Guide | Plotly*. [online] Available at: https://dash.plotly.com/dash-core-components/dropdown. [Accessed 26 Dec. 2024].

Flowrl.com. (2024). *API | pywebview*. [online] Available at: https://pywebview.flowrl.com/guide/api.html#webview-create-window [Accessed 26 Dec. 2024].

‌

‌

‌

‌

‌