

Assessing AI risk in assignment setting



Bernie Tiddeman
Computer Science
bpt@aber.ac.uk

Gen AI and the risk to assessments

- Gen AI tools pose a significant risk to the validity of assessments
 - And they are improving all the time!
- Such tools are excellent at performing some tasks
 - Creating plausible looking English prose
- But are not so good at other tasks
 - Making valid logical arguments, numerical tasks (“How many ‘r’s in ‘strawberry’?”), etc.

Revealed: Thousands of UK university students caught cheating using AI

Guardian investigation finds almost 7,000 proven cases of cheating – and experts say these are tip of the iceberg



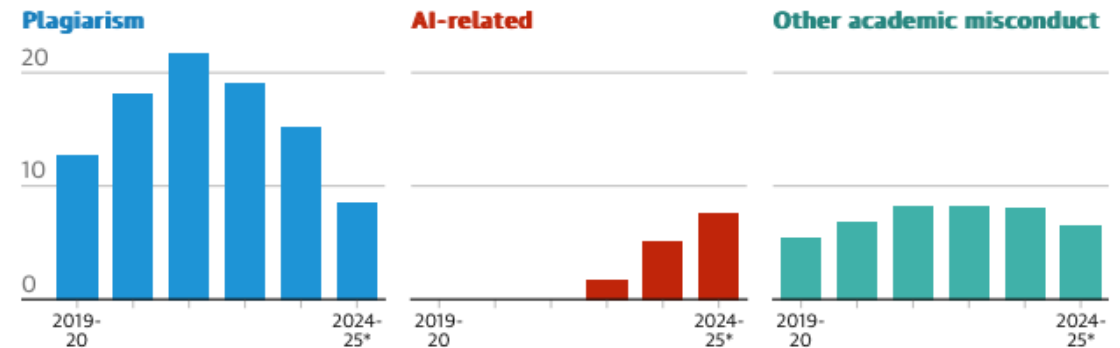
More than a quarter of responding universities did not yet record AI misuse as a separate category of misconduct in 2023-24. Photograph: Maurice Norbert/Alamy

Gen AI risks to validity of qualifications

- From the Guardian article:
 - “A survey by the Higher Education Policy Institute in February found 88% of students used AI for assessments.”
 - “Last year, researchers at the University of Reading tested their own assessment systems and were able to submit AI-generated work without being detected 94% of the time.”
- Better to avoid temptation!

As plagiarism falls, AI-related misconduct is rising

Proven misconduct cases per 1,000 students



Guardian graphic. Source: Guardian FoI requests . Notes: Between 80 and 125 universities responded depending on year and misconduct category. Rates are per 1,000 student population of universities which responded. 2024-25 is projected based on data to May.

Gen AI risks to education

- Preliminary results show that use of LLMs reduces brain connectivity:
 - “Brain connectivity systematically scaled down with the amount of external support: the Brain-only group exhibited the strongest, widest-ranging networks, Search Engine group showed intermediate engagement, and LLM assistance elicited the weakest overall coupling.”
- The “Brain muscle” needs exercise!

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task[△]

Nataliya Kosmyna¹
MIT Media Lab
Cambridge, MA

Eugene Hauptmann
MIT
Cambridge, MA

Ye Tong Yuan
Wellesley College
Wellesley, MA

Jessica Situ
MIT
Cambridge, MA

Xian-Hao Liao
Mass. College of Art
and Design (MassArt)
Boston, MA

Ashly Vivian Beresnitzky
MIT
Cambridge, MA

Iris Braunstein
MIT
Cambridge, MA

Pattie Maes
MIT Media Lab
Cambridge, MA

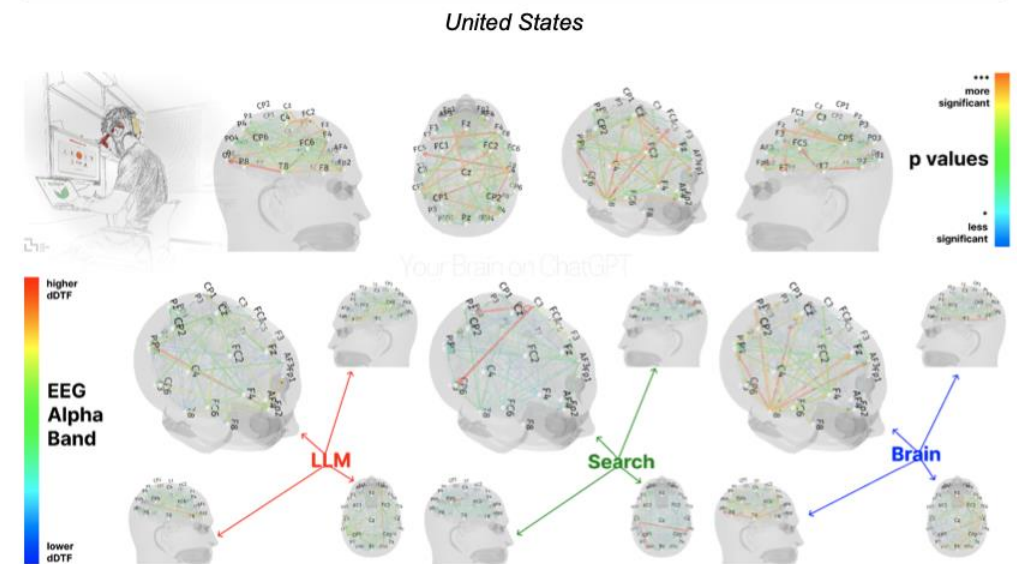


Figure 1. The dynamic Direct Transfer Function (dDTF) EEG analysis of Alpha Band for groups: LLM, Search Engine, Brain-only, including p-values to show significance from moderately significant (*) to highly significant (***).

The AI risk measurement scale (ARMS)

- Developed by Greenwich University in 2023
 - https://www.gre.ac.uk/_data/assets/pdf_file/0022/323590/ai-risk-measure-scale-guidance-and-resources-website-version.pdf
 - <https://www.youtube.com/watch?v=GEFMKYUCPJg>
- Developed in the Business School
- Guidance to rate assessments on 5-point risk scale
- “Programme Lead” rates all assessments and discusses with “Module Leads”
- Shares good practice with teaching team, and encourages MLs with higher risk assessments to redesign them
- Approx. 40% drop in UAPs

AI Risk Measure Scale (ARMS)		Description Download	Examples
1	Very low	It is highly unlikely that students can use AI to produce this type of assignment.	- Assignments that embed authenticity in the design (e.g. field trip + reflective report), assignments that allow establishing the identity of the person (e.g. presentations, in-person exams). - Subjective assignments that require personal reflection or creative thinking, such as personal narratives, or artistic projects. These types of assignments are typically based on the student's opinions and insights, which are difficult to replicate using AI.
2	Low	Students could potentially use AI to produce the assignment, but it is very unlikely to have a significant impact on the assignment's quality and/or originality.	- Assignments that draw on unique teaching material (e.g. novel cases produced by tutor). - Assignments that have clear guidelines, such as solving math problems or coding exercises, where AI could assist but the student's approach or solution is what is being evaluated as the main focus of the assignment.
3	Moderate	There is a moderate likelihood that students can use AI to produce the assignment, and it could have a moderate impact on the assignment's quality and/or originality.	- Assignments where AI could be used to assist students in completing the assignment, but the final work would still require the student's critical thinking, analysis, and interpretation. - Assignments that require a more complex analysis of a topic, e.g., critical analysis essay or a scientific report. Students may use AI tools to help with data analysis, visualisation, or interpretation in some areas, but the writing and argumentation are largely based on the student's understanding and critical thinking.
4	High	It is easy for students to use AI to produce the assignment, and it could significantly impact the assignment's quality and/or originality.	- Assignments that focus on well-published company case studies (e.g., Innocent, Apple, Bohoo, Starbucks etc.) and rather generic topics (e.g. advantages and disadvantages of FDI) which students can easily obtain through AI bots. - Assignments that involve sophisticated algorithms or complex modelling, such as financial forecasting, predictive analytics, or image recognition, where students could use AI to generate both, results and insights/commentary.
5	Very high	It is very easy for students to use AI to produce the assignment, and it will have a significant impact on the assignment's quality and/or originality.	- Assignments that require students to produce summaries or abstracts of published articles, reports, or research papers, this includes research proposals. These assignments require no input/modification from students and can be entirely produced by AI. - Assignments that involve large-scale data processing, such as machine learning projects or artificial intelligence simulations, where students could rely entirely on AI to create the work.

1. Very low risk

- **Description:** It is highly unlikely that students can use AI to produce this type of assignment.
- **Examples**
 - Assignments that embed authenticity in the design (e.g. field trip + reflective report), assignments that allow establishing the identity of the person (e.g. presentations, in-person exams).
 - Subjective assignments that require personal reflection or creative thinking, such as personal narratives, or artistic projects. These types of assignments are typically based on the student's opinions and insights, which are difficult to replicate using AI.
- **CS context?**

2. Low risk

- **Description:** Students could potentially use AI to produce the assignment, but it is very unlikely to have a significant impact on the assignment's quality and/or originality.
- **Examples**
 - Assignments that draw on unique teaching material (e.g. novel cases produced by tutor).
 - Assignments that have clear guidelines, such as solving math problems or coding exercises, where AI could assist but the student's approach or solution is what is being evaluated as the main focus of the assignment.
- **CS Context?**

3. Moderate risk

- **Description:** There is a moderate likelihood that students can use AI to produce the assignment, and it could have a moderate impact on the assignment's quality and/or originality.
- **Examples:**
 - Assignments where AI could be used to assist students in completing the assignment, but the final work would still require the student's critical thinking, analysis, and interpretation.
 - Assignments that require a more complex analysis of a topic, e.g., critical analysis essay or a scientific report. Students may use AI tools to help with data analysis, visualisation, or interpretation in some areas, but the writing and argumentation are largely based on the student's understanding and critical thinking.
- **CS Context?**

4. High risk

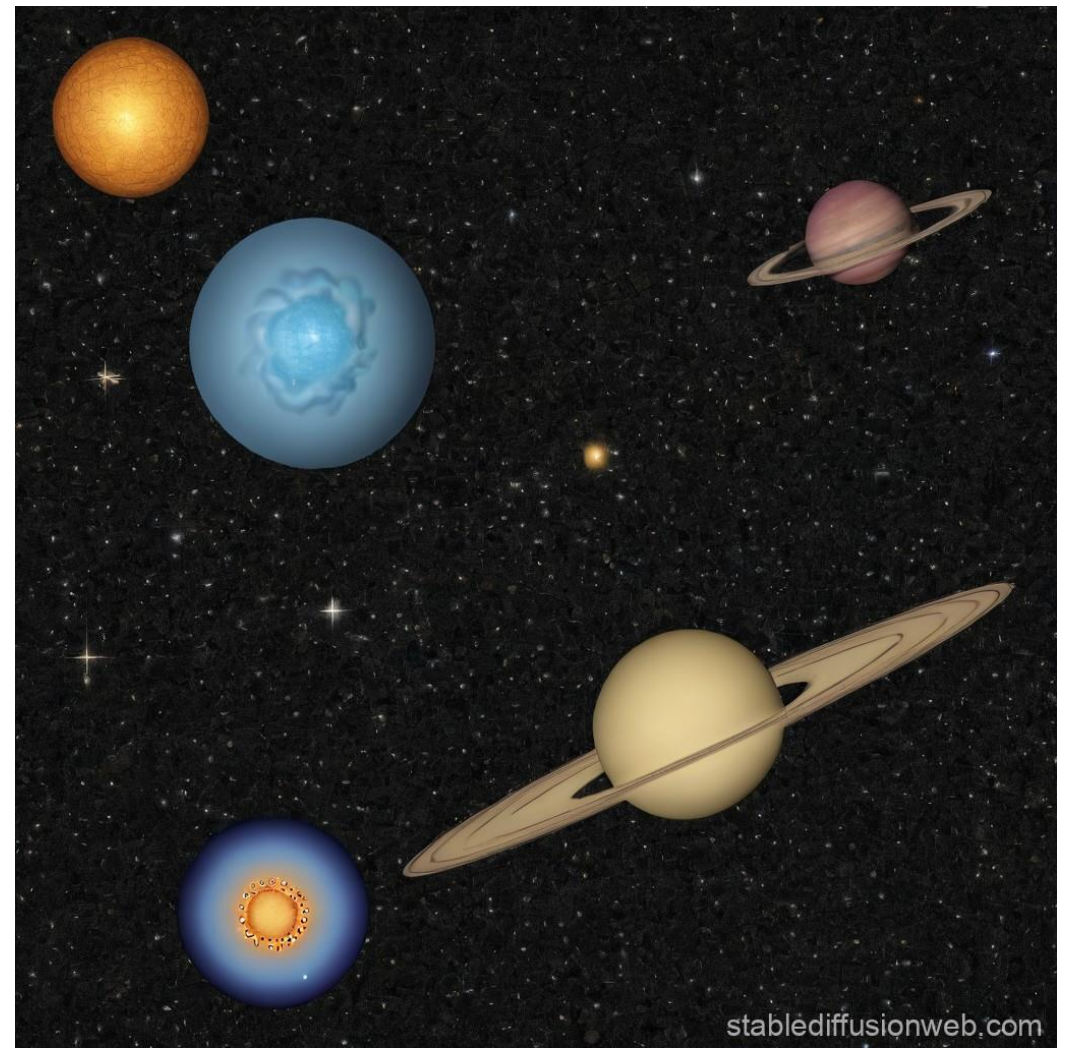
- **Description:** It is easy for students to use AI to produce the assignment, and it could significantly impact the assignment's quality and/or originality.
- **Examples:**
 - Assignments that focus on well-published company case studies (e.g., Innocent, Apple, Bohoo, Starbucks etc.) and rather generic topics (e.g. advantages and disadvantages of FDI) which students can easily obtain through AI bots.
 - Assignments that involve sophisticated algorithms or complex modelling, such as financial forecasting, predictive analytics, or image recognition, where students could use AI to generate both, results and insights/commentary.
- **CS Context?**

5. Very high risk

- **Description:** It is very easy for students to use AI to produce the assignment, and it will have a significant impact on the assignment's quality and/or originality.
- **Examples:**
 - Assignments that require students to produce summaries or abstracts of published articles, reports, or research papers, this includes research proposals. These assignments require no input/modification from students and can be entirely produced by AI.
 - Assignments that involve large-scale data processing, such as machine learning projects or artificial intelligence simulations, where students could rely entirely on AI to create the work.
- **CS Context?**

You be the judge!

- Original Aber Online MSc assignment (Part 2) on Algorithms and Data Structures module:
 - “Write an essay of 1,250 words where you justify how you would sort the data in an application to manage information about thousands of celestial objects. Your application can process, i.e. find, delete, and insert information regarding astronomical objects. Justify and sketch the steps for utilising a sorting and searching algorithm.”



<https://imgcdn.stablediffusionweb.com/2025/7/6/405bee05-0e50-46d4-97bc-7369157d7be5.jpg>

Chat GPT Solution

- Does quite a plausible job
- Distills mostly quite generic information, widely available
- My rating?
 - High Risk / Very high risk

Efficient Data Management of Celestial Objects Through Sorting and Searching Algorithms

In the modern era of astronomy and astrophysics, software applications have become essential for managing vast datasets comprising celestial objects. With projects such as the Sloan Digital Sky Survey and NASA's Gaia mission cataloging millions of stars, galaxies, and other astronomical bodies, it is imperative for applications handling this information to employ efficient data organization, sorting, and searching techniques. This essay will justify the use of specific sorting and searching algorithms in an application developed to manage, search, insert, and delete information about thousands of celestial bodies. It will also outline the implementation steps of these algorithms to ensure fast access and optimal performance.

Nature of Astronomical Data

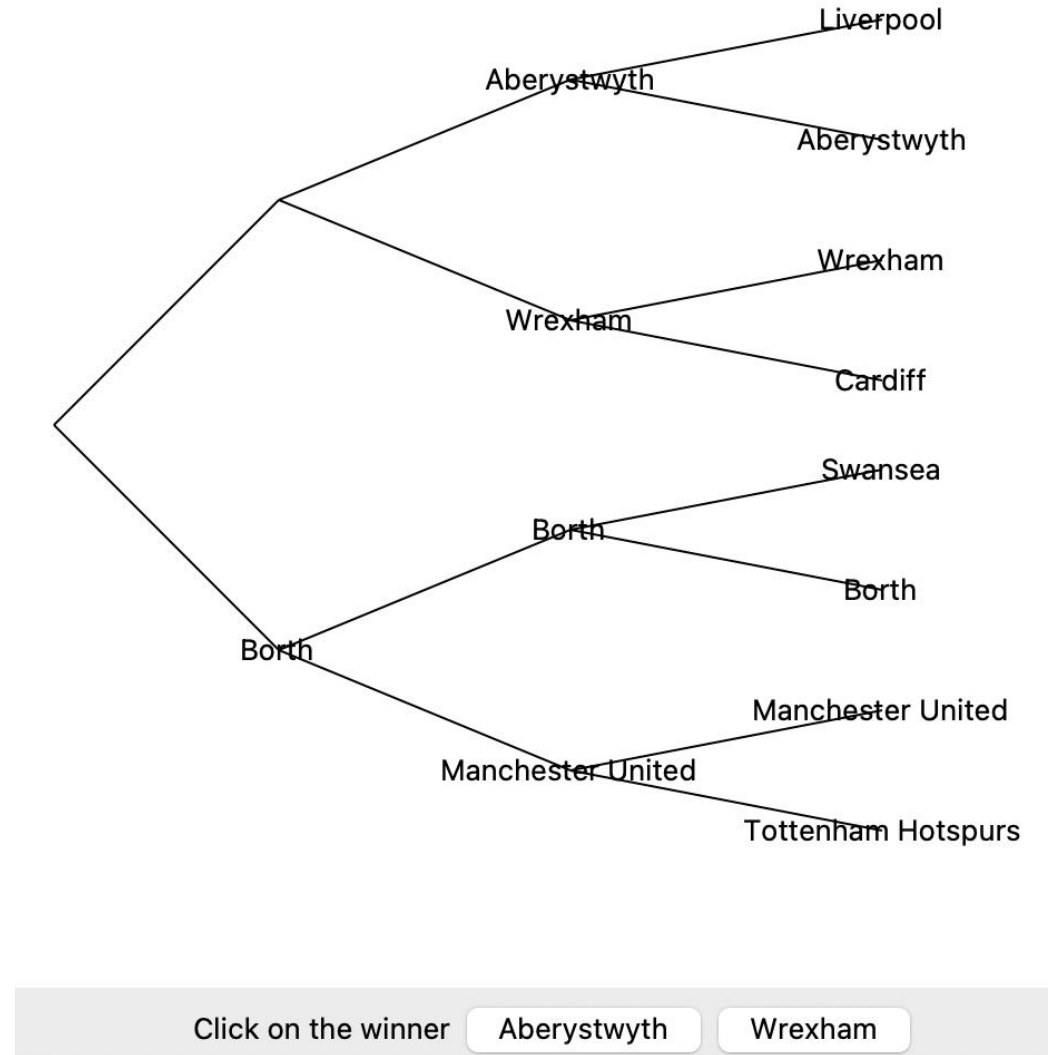
Astronomical datasets typically include structured information for each object, such as:

- Object ID
- Type (e.g., star, galaxy, asteroid)
- Position (right ascension and declination)
- Magnitude (brightness)
- Spectral type
- Redshift (for galaxies)
- Distance from Earth
- Date of observation

Given the heterogeneity and scale of this data, the application's backend must ensure efficient storage and retrieval. These requirements naturally lead us to examine the role of data structures and algorithms.

You be the judge!

- The revised assignment (summary):
- Write python code to support ***single elimination competitions***.
- You are provided with classes to display the competition tree, and a Node class that has a left and right child and some data.
- The implemented class should:
 - Include the root of the competition tree
 - Be iterable to iterate over matches
 - Provide a constructor ...
 - Include detailed comments ...



ChatGPT Solution

- Looks plausible at first, but:
 - Needed some work to get it to run
 - Tree constructed OK, but teams not shown on buttons
 - Added unnecessary constraint (power of 2 teams)
 - Behaviour not correct
- My rating?
 - Low risk
 - “Draws on unique teaching material”
 - (But didn’t try specialized coding AI)

```
class SingleEliminationTournament:
    """
    This class represents a single elimination competition using a binary tree.

    Attributes:
        tree (Node): The root of the binary competition tree.
    """

    def __init__(self, filename):
        """
        Constructor to read team names from a file and build a binary tree.
        Leaf nodes contain team names, internal nodes have data=None.

        Args:
            filename (str): File with one competitor name per line.
        """
        with open(filename, 'r') as f:
            teams = [line.strip() for line in f if line.strip()]

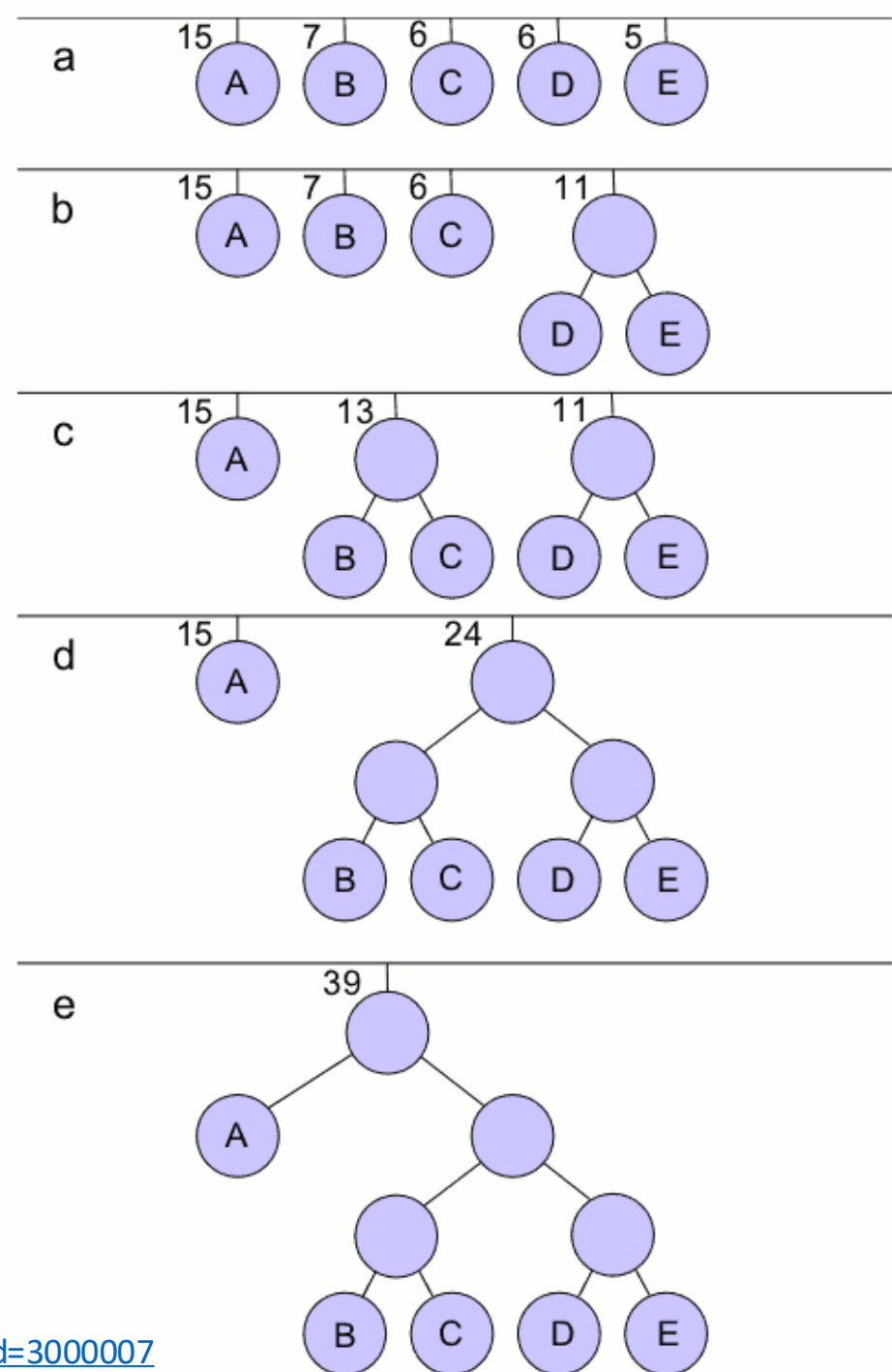
        if not teams or (len(teams) & (len(teams) - 1)) != 0:
            raise ValueError("Number of teams must be a power of two.")

        # Create initial leaf nodes
        nodes = [Node(data=team) for team in teams]

        # Build the tree bottom-up
        while len(nodes) > 1:
            next_level = []
            for i in range(0, len(nodes), 2):
                parent = Node()
                parent.left = nodes[i]
                parent.right = nodes[i + 1]
                next_level.append(parent)
            nodes = next_level
```

You be the judge!

- Implement the Huffman encoding algorithm in Java. It should:
 - take in a list of symbols and the probability of each as input.
 - use a priority queue to construct a binary tree by the process shown opposite.
 - Include a method to traverse the tree to get the encoding of each symbol (e.g. left = 1 and right = 0), and
 - Include a method for decoding an input stream of 0's and 1's.



ChatGPT Solution

- Looks plausible and runs out of the box
- Implementation of a well know algorithm (lots of examples on web)
- Not given existing code that it needs to work with
- My rating?
 - High Risk / Very high risk

```
import java.util.*;

public class HuffmanCoding {

    static class Node {
        char symbol;
        double probability;
        Node left, right;

        Node(char symbol, double probability) {
            this.symbol = symbol;
            this.probability = probability;
        }

        Node(double probability, Node left, Node right) {
            this.symbol = '\0'; // Internal node
            this.probability = probability;
            this.left = left;
            this.right = right;
        }

        boolean isLeaf() {
            return left == null && right == null;
        }
    }

    static class HuffmanTree {
        private Node root;
        private Map<Character, String> encodingMap;

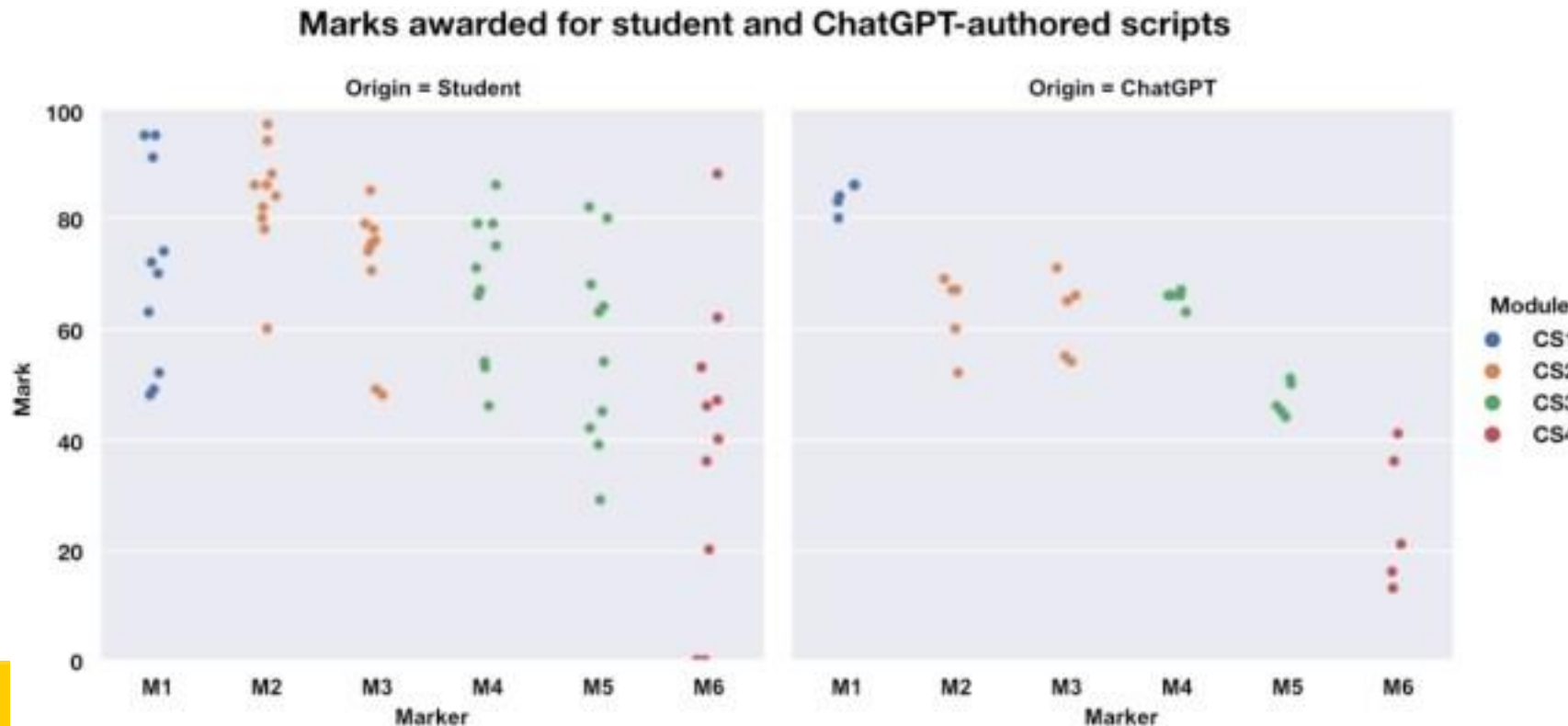
        public HuffmanTree(Map<Character, Double> probabilities) {
            buildTree(probabilities);
            encodingMap = new HashMap<>();
        }
    }
}
```

You be the judge!

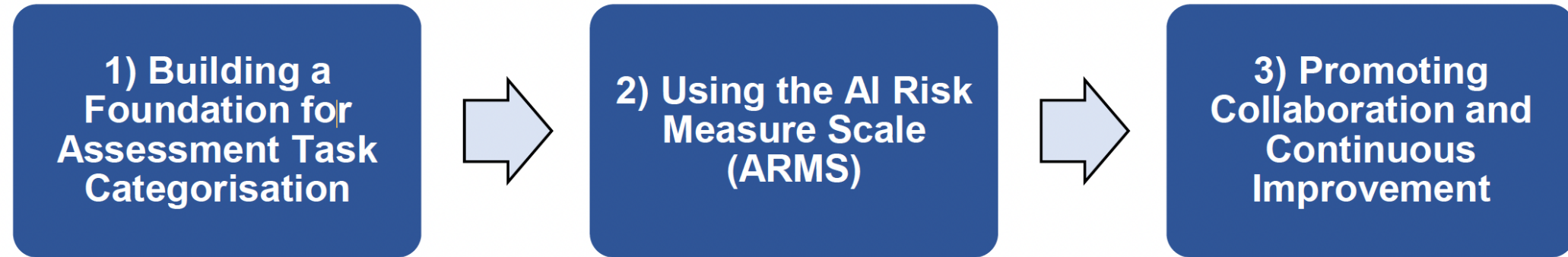
- Information security, MSc level
- “An extended report exploring a topic of the student's choosing related to a case study developed during the module. As appropriate for a postgraduate module, the report is expected to be written to a high standard in ‘academic language’ and to employ extensive referencing of module materials as well as external resources.”

MSc level may be less at risk from Gen AI

- **“Bob or Bot: Exploring ChatGPT's Answers to University Computer Science Assessment”, Richards et al., ACM Trans. On Computing Education, 24(1), pp1-32, 2024.**
- No ChatGPT MSc assignments reached the 50% pass threshold.



Aberystwyth exams processes



- In CS:
 - For (1) some training may be needed (summer show and tell, LTEU conference, ...).
 - For (2) module moderators obvious person to check assignment before release to comment on AI risk and rate
 - For (3) who would be best placed to look across modules and promote best practice?
 - Probably chairs of exam boards to look across modules (but extra work!)
 - L&T meetings and Summer Show and Tell to spread good practice
- Should it be combined with other UAP risks?

Summary

- ARMS – AI Risk Measurement Scale
 - Developed by Greenwich Business School
 - Simple risk scale, manual process
 - Some adaptation needed for different subjects, improvements in AI, etc.
- Suggested integration with Aber (CS) processes
 - Module moderators to rate and feedback
 - Feeds into exam board discussion through exam board chairs
 - Spread good practice through