University of Glasgow | School of Computing Science

Assessed Coursework

Course Name	**Safety Critical Systems H**			
Coursework Number	**1**			
Deadline	**Time:**	**4.30pm**	**Date:**	**3rd March 2017**
% Contribution to final course mark	**20%**			
Solo or Group ✓	**Solo**	✓	**Group**	
Anticipated Hours				
Submission Instructions				
Please Note: This Coursework cannot be Re-Assessed				

Code of Assessment Rules for Coursework Submission

Deadlines for the submission of coursework which is to be formally assessed will be published in course documentation, and work which is submitted later than the deadline will be subject to penalty as set out below. The primary grade and secondary band awarded for coursework which is submitted after the published deadline will be calculated as follows:

(i) in respect of work submitted not more than five working days after the deadline
   a. the work will be assessed in the usual way;
   b. the primary grade and secondary band so determined will then be reduced by two secondary bands for each working day (or part of a working day) the work was submitted late.

(ii) work submitted more than five working days after the deadline will be awarded Grade H.

Penalties for late submission of coursework will not be imposed if good cause is established for the late submission. You should submit documents supporting good cause via MyCampus.

Penalty for non-adherence to Submission Instructions is 2 bands

You must complete an "Own Work" form via https://studentltc.dcs.gla.ac.uk/ for all coursework

Safety-Critical Systems: Open Assessment 2016-17

Techniques for Supporting the Use of Artificial Intelligence in Safety Critical Systems (Level 4)

Prof. Chris Johnson
School. of Computing Science, University of Glasgow, Glasgow, G12 8RZ. Scotland.
johnson@dcs.gla.ac.uk, http://www.dcs.gla.ac.uk/~johnson

1 Introduction

Artificial intelligence and machine learning techniques pose particular problems for the development of safety-critical systems. In particular, they can introduce non-determinism because the behaviour of any software may change over time as the application learns to adapt its behaviour, for instance through changes in the environment. How can a regulator or designer be sure that those changes will not endanger the future safety of a complex system?

2 Tool Development

The aim is to provide a tool that will help assess and mitigate risks that arise through the introduction of artificial intelligence or machine learning techniques into a safety-critical system. The design of the technique is entirely open. You may choose to use one of the risk assessment techniques that are introduced during this course, such as Fault Trees or Failure Modes, Effects and Criticality Analysis. If you do this then in your report you need to explain why this particular approach is appropriate for the introduction of AI/ML into safety-related environments. Alternatively, you may propose one of several new approaches; for instance, the use of model checking and constraints on the adaptation of AI. In both cases, you must show how the technique can be used with a detailed AND specific case study based on significant research into existing plans by manufacturers/governments that have approved the safety-related use of AI in limited trials.

The key aim is to help organizations assess the likelihood and consequence of hazards that can arise from the use of AI in safety-critical systems. These include issues associated with testing and debugging, especially from the risk exposure associated with mass-market products. They also include the problems of interaction between the software, the processes under control and the environment in which it operates. The specific focus must be on helping mitigate those risks by appropriate planning before a system goes live.

You may choose to develop electronic tools that support the application of your technique using any programming methodology. The implementation of the tool could rely on simple web pages generated using HTML, PHP or any other associated technology. Your design may be realized using conventional programming languages or you could simply rely on paper-based support. However, the marking scheme will take into account both the strengths of the design for the risk assessment technique and the effectiveness of an implementation in terms of the support that they offer to the potential end users of AI technologies.

3 Evaluation

It is important that you evaluate your technique/tool for assessing the risks associated with AI deployment (hint: some systems have already been deployed so you can identify case studies from these to determine if your technique can model known risks). Other means of evaluation would be to ask a number of different users to try it out, exploiting an appropriate evaluation methodology. For example, you could ask one group to use your technique and another to use an alternate approach developed by someone else in the course. If you do this you MUST consider the relevant plagiarism

guidance on the School Learning and Teaching Committee web site and state the name of the person you worked with on your submission. You must develop your reports independent of each other. You also need to consider the level of existing expertise that the people you test will have in the development of AI technologies in safety-related systems.

If you evaluate with a friend and split your users into two groups, one for each tool, then this raises important methodological concerns. Firstly, how would you insure that both groups have the same level of expertise and background knowledge so that any comparisons are fair? Secondly, how would you go about assessing the accuracy of any risk assessments that are produced? Please consult with me before conducting your evaluation so that I can provide advice in answering some of these questions. You should also consult the course handbook and associated web pages that cover the ethical guidelines for user testing.

4 Transferable Skills

This exercise will provide a first-hand introduction to the challenges that face many large organizations as they try to innovate and at the same time ensure the safety of their products. There is little common agreement on the best approaches to adopt and hence you will be working in an area of active research, which is also a focus for public, government and commercial interest. The exercise will underline the uncertainty that often characterizes risk assessment in safety-critical engineering – for example, credible attempts to use quantitative techniques will attract high marks especially if you can validate assessments of the probability and consequence of particular hazards. You should consider the role of regulators in the development process; this is covered in the early part of the course including the use of process based software standards. Recall also that regulators must protect safety but also, where possible, enable companies to develop new markets.

5 Assessment Criteria and Submission Details

This exercise is degree assessed. It contributes 20% to the total marks associated with this course. The body of the report should not exceed fifteen A4 pages. The report must be printed out and must be submitted in a secure binder (something that keeps the pages together and does not have sharp edges). It must include: A title page containing your contact details (metric, email etc); a table of contents and appropriate page numbers; a section on the tool that you developed; a section on the evaluation method that you used; a results sections and some conclusions.

In addition to the fifteen pages in the body of the report, you may also include appendices. These should contain the listing of any code used during the study together (this can be included on a CD) with suitable acknowledgements for the source of code that has been borrowed from other programmers. The report should be handed in by 16:30 on 3rd March 2017 using the submission box outside the teaching office in Lilybank Gardens. Please make sure that you keep back-up copies of all of your work and submit a plagiarism statement using the standard on-line form. The following marking scheme will be applied: 30 for the method; 20 for the results; 30 for the conclusion; 20 for the technical documentation. All solutions must be the work of the individual submitting the exercise and the usual lateness penalties will apply unless I am given good reason in advance of the deadline. You must state your name and the title of the exercise on the front of your submission – this topic is only for level H students. Failure to answer the correct question will jeopardise your marks.

6 Hints

You will need to do considerable reading first into the background of AI technology so please do not delay starting this assessment.