Last updated: April 26, 2021

Course information

Course name:	Logics for Safe AI
Course code:	INFOLSAI
Credits:	7,5 ECTS
Teaching period:	2020-2021 period 4
Level (1, 2, 3 or M)	M
Schedule:	Lectures / Q&A sessions: • Mondays, 11:00-12:00 • Wednesdays, 10:00-10:45 Practical sessions: • Wednesdays, 11:00-12:45
Lecturers:	Lectures are taught by Natasha Alechina and Brian Logan
Contact person for the course:	Natasha Alechina (n.a.alechina@uu.nl)

Course goals

The course is an **advanced** course on verification and synthesis of provably correct AI systems. The course will make students familiar with the logical and computational techniques for specifying, verifying and synthesising the behaviour of AI agents and multi-agent systems, and the the computational aspects of these formalisms. On completing the course, students will be able to use appropriate logical formalisms, techniques and tools for verifying and synthesising the behaviour of AI agents and multi-agent systems.

Course contents

This course is about ensuring the safety and reliability of autonomous AI agents and multi-agent systems. In order to guarantee that the behaviour of a system achieves its objectives, we use formal proofs rather than empirical studies, and either formally verify that the system behaves in accordance with the specified objectives, or automatically synthesise provably correct behaviours from specifications of the system objectives. The formal techniques for doing this include epistemic and temporal logics and their combinations, and constitute the main technical content of the course. The emphasis is on mastering these techniques, their computational aspects, and the use of tools implementing them to verify and synthesise AI agents and multi-agent systems. The course also prepares students for undertaking research on formal aspects of artificial intelligence, and provides the foundation for undertaking Masters projects on developing safe and reliable AI systems. Lab sessions will introduce students to relevant specification and modelling techniques and the use of tools such as MCMAS and SynKit for the verification and synthesis of AI agents and multi-agent systems.

Assumed background

The course assumes the following background in logic: classical propositional logic, formal logical languages, parsing a formula, inductive definitions. The logic content of INFOMAIR is probably sufficient, but if you feel you need additional background reading, the book

Modelling Computing Systems: Mathematics for Computer Science by Faron Moller and Georg Struth is available online through the UU library at: https://link-springer-com.proxy.library.uu.nl/book/10.1007/978-1-84800-322-4 The first few chapters contain a very gentle introduction to these topics

For computer science, the following background is assumed: exposure to algorithms and computational complexity. If you don't have this background at all, please contact the course lecturers, as you may struggle with the course. If you need a refresher, you can read the first three chapters (and possibly also Appendix B on sets, trees and graphs) of *Introduction to Algorithms* by Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein (also available online through the UU library).

Required materials

The recommended textbook for the course is *Logical Methods for the Specification and Verification of Multiagent Systems* by Wojtek Jamroga, which is available as a free download at: https://home.ipipan.waw.pl/w.jamroga/papers/jamroga15specifmas-20200411.pdf For specific topics, we will also provide references to open source papers and software tools.

For the practical sessions, you will need to download and install the model checker MCMAS. MCMAS is open source, and runs on Windows, MacOS, and Linux. For information on how to download and install MCMAS, see the *MCMAS installation guide* on Blackboard.

Course format

Due to the COVID-19 pandemic, lectures, practical sessions and assessments will be online.

The first lecture will be given live on MS Teams (a recording will be placed on Blackboard). Other lecture hours will be used for live question and answer (Q&A) sessions on MS Teams. The lectures for these sessions will be provided on Blackboard as recordings at least 2 working days before the corresponding question and answer session (so on Wednesday for Monday Q&A, and on Monday morning for Wednesday Q&A). The practical sessions will be used to discuss the coursework assignments, which are both theoretical and practical (model checking assignments). Practical sessions will also be on MS Teams. For a detailed schedule of lectures and practicals, see below.

Week	Time	What	Where	Topic
17 - Apr 26	Mon 11:00-12:00	live lecture	Teams	Introduction; temporal logic
	Mon 12:00-12:45	live student meeting	Teams	Social
		CW1 released	BB	Temporal logic
	Wed 10:00-10:45	live Q&A	Teams	Temporal logic
	Wed 11:00-12:45	live practical	Teams	Installing MCMAS; CW groups
18 - May 3	Mon 11:00-12:00	live Q&A	Teams	Temporal logic; CW1
	Wed	UU non-teaching day		Replaced by group Q&As on CW1
	Fri 23:59	CW1 deadline	BB	
19 - May 10	Mon 11:00-12:00	live Q&A	Teams	Epistemic logic
	Mon 12:00	CW2 released	BB	Epistemic logic
	Wed 10:00-10:45	live Q&A	Teams	Epistemic logic

	Wed 11:00-12:45	live practical	Teams	CW2
20 - May 17	Mon 11:00-12:00	live Q&A	Teams	CTLK; epistemic planning
	Mon 23:59	CW2 deadline	BB	
	Wed 10:00-12:45	practice exam	BB	Temporal & epistemic logic
21 - May 24	Mon	UU non-teaching day		
	Wed 10:00-10:45	live Q&A	Teams	Coalition logic
	Wed 11:00-12:45	live practical	Teams	feedback on practice exam
22 - May 31	Mon 11:00-12:00	live Q&A	Teams	ATL
	Mon 12:00	CW 3 released	BB	Coalition logic; ATL
	Wed 10:00-10:45	live Q&A	Teams	ATL
	Wed 11:00-12:45	live practical	Teams	CW3
23 - Jun 7	Mon	UU non-teaching day		
	Wed 10:00-10:45	live Q&A	Teams	ATL; ATEL
	Fri 23:59	CW3 deadline	BB	
24 - Jun 14	Mon 11:00-12:00	live Q&A	Teams	ATL synthesis
	Mon 12:00	CW4 released	BB	ATL synthesis
	Wed 10:00-10:45	live Q&A	Teams	ATL synthesis
	Wed 11:00-12:45	live practical	Teams	CW4
25 - Jun 21	Mon 11:00-12:00	live Q&A	Teams	BDI logics
	Wed 10:00-10:45	live Q&A	Teams	BDI logics
	Wed 11:00-12:45	live practical	Teams	CW4; advice on exam revision
	Fri 23:59	CW4 deadline	BB	
26 - Jun 28	Mon 08:45-13:45	exam	BB	
28 - Jul 14	Wed 08:45-13:45	resit exam	BB	

Assessment

Coursework

The coursework is a mixture of practical and theoretical exercises that will prepare you for the exam. Coursework assignments will be issued bi-weekly. Students are expected to work on courseworks in groups of two or three people. Please make sure that at least one member of the group is comfortable with installing MCMAS. Submission and feedback for the coursework will be done using Blackboard. Coursework does not contribute to the final mark, but you must pass 50% of the assignments to qualify for the exam.

Exam

There is one exam at the end of the course. As this year's exam has to be online, it will consist of problem solving questions and is 'open book', that is, you can use whatever information you have from the textbooks, other sources, and the internet. You are **not** allowed to consult other people about the answers to exam questions.

In order to be allowed to sit the remedial exam, you should either get 4 or AANV for the original exam (e.g. miss it for unavoidable reasons).

Deadline policy

Deadlines are strict. If, however, there are personal or covid-related circumstances for missing a deadline, you can ask for an extension or a replacement coursework assignment

by contacting us prior to the deadline. Extensions of more than one week will only be granted after consulting the study advisor.

Fraud and plagiarism

Please make sure that you are aware of the UU rules about these subjects:
https://students.uu.nl/en/practical-information/policies-and-procedures/fraud-and-plagiarism