IITB-Monash Research Academy

An Indian-Australian research partnership

Project Title:	Automated Verification of Achievable Properties in Mechanism Design	
Project Number	**IMURA1035**	
Monash Main Supervisor (Name, Email Id, Phone)	Julian Gutierrez, Julian.Gutierrez@monash.edu	*Full name, Email*
Monash Co-supervisor(s) (Name, Email Id, Phone)		
Monash Head of Dept/Centre (Name, Email)	Jianfei Cai, jianfei.cai@monash.edu	*Full name, email*
Monash Department:	Department of Data Science and AI	
Monash ADGR (Name, Email)	Guido Tack, Guido.Tack@monash.edu	*Full name, email*
IITB Main Supervisor (Name, Email Id, Phone)	Prof. Swaprava Nath, swaprava@cse.iitb.ac.in	*Full name, Email*
IITB Co-supervisor(s) (Name, Email Id, Phone)		*Full name, Email*
IITB Head of Dept (Name, Email, Phone)	Prof. Varsha Apte, head@cse.iitb.ac.in	*Full name, email*
IITB Department:	Computer Science and Engineering	

Research Clusters:

Research Themes:

Highlight which of the Academy's CLUSTERS this project will address? *(Please nominate JUST **one.** For more information, see www.iitbmonash.org)*		**Highlight which of the Academy's Theme(s) this project will address?** *(Feel free to nominate more than one. For more information, see www.iitbmonash.org)*	
1	*Material Science/Engineering (including Nano, Metallurgy)*	***1***	***Artificial Intelligence and Advanced Computational Modelling***
2	*Energy, Green Chem, Chemistry, Catalysis, Reaction Eng*	2	*Circular Economy*
3	*Math, CFD, Modelling, Manufacturing*	3	*Clean Energy*
4	***CSE, IT, Optimisation, Data, Sensors, Systems, Signal Processing, Control***	4	*Health Sciences*
5	*Earth Sciences and Civil Engineering (Geo, Water, Climate)*	5	*Smart Materials*
6	*Bio, Stem Cells, Bio Chem, Pharma, Food*	6	*Sustainable Societies*
7	*Semi-Conductors, Optics, Photonics, Networks, Telecomm, Power Eng*	7	*Infrastructure*
8	*HSS, Design, Management*		

The research problem

Mechanism design is considered the engineering approach to economic theory and in recent times it has emerged as a solution for multi-agent AI/ML. Classical AI/ML solves an optimization problem (for learning or making decisions) from the viewpoint of a single agent and finds an *optimal* solution using several techniques (e.g., convex optimization, deep learning, etc.). For multiple agents, who may have possibly conflicting objective functions, an optimal solution for an agent depends on the actions of the other agents. Therefore, there is no longer a single optimization problem for which a solution needs to be found, rather the question becomes how to find a *game theoretic equilibrium* of such an interaction from which no agent would like to deviate. This is precisely the question asked in *mechanism design*.

The study of mechanism design dates back to the 1950s, which started the evolution of various foundational results due to Arrow, Gibbard, Satterthwaite, Myerson, and Nash, and over the seven decades after that numerous mechanisms have been developed that find applications in *auctions, voting, fair division, matching, crowdsourcing, social networks*, and many more. In the past three decades, with the proliferation of computing technologies, it has also been an interdisciplinary area of research to find (a) how to design mechanisms that also guarantee computational efficiency (since to take a real-time automated decision using computers, the mechanisms/algorithms need to be fast) and be applicable to various computer science application areas, and (b) how the techniques from computer science, e.g., approximation guarantees, can be used to develop mechanisms where it is impossible to satisfy a given set of properties. Both approaches made several landmarks in the landscape of artificial intelligence where it is possible to design computationally and economically efficient algorithms (or mechanisms) that run a *fair division of divisible or indivisible items*, *match students to universities or organ donors to patients in a robust way, help taking efficient decisions on public goods with monetary transfers*, or *help design socially optimal participatory democracies*.

Though many mechanisms are great in their innovation and the impact they deliver on society, one major obstacle in the practical use of these mechanisms is their explainability, i.e., the ease of understanding, for the end users. This is particularly because the mechanisms often tend to be complex algorithms and it requires training in mathematics or computer science to understand that these algorithms satisfy the properties that they are claimed to satisfy. Also, if a set of properties are infeasible to satisfy together by any mechanism, there is also little work to understand what is a minimal set of relaxed properties that is possible to satisfy together and can be found using low complexity algorithms. This project addresses these two fundamental questions in detail as explained below.

The intellectual merit of the project

The proposed project
1) will develop formal models of the mechanisms in a given setup, e.g., in auction, voting, or fair division, where the economic properties are mathematically well defined and known existing mechanisms satisfy many of these properties, and try to develop an automated method of verifying these mechanisms for their claimed properties in a computationally inexpensive manner, and
2) will develop a formal method to identify the minimal set of properties that is achievable efficiently by any mechanism. For instance, properties like strategyproofness (lying is not beneficial for any player) and unanimity (if every player prefers some alternative the most, that alternative is chosen as the outcome) in the voting domain always lead to dictatorial mechanisms (where exactly one specific player's, called the dictator, top alternative is chosen as the outcome always). What are the 'nearest' relevant properties where there exists non-dictatorial mechanisms and can they be calculated via a 'low complexity' algorithm?

While there are some initial works in the direction of (1), e.g., by Barthe et al. (2016), Tadjouddine et al. (2008), Bordini et al. (2009), and Branzei and Procaccia (2015), the direction of (2) is mostly unexplored to the best of our knowledge, but this is currently little work done in the area.

The broader impact of the project on societies

The first part of the project helps give a framework for mechanism design to be automatically verified by users who may be unfamiliar with mechanism design and therefore may be sceptical about its claimed properties. By testing the formal model, they can be convinced to be sure of the mechanism's functions and participate in it without hesitation. This will imply that a larger population of users will adopt mechanisms with such verifiable guarantees and will be benefited by the provable desirable properties that the mechanisms satisfy.

The second part of the project aims to advance the theory of mechanism design using formal methods. It asks the question of which set of additional properties can be achieved by modifying the current setup or mechanisms. This can lead to an automated path of developing properties and mechanisms which can help societies to take collective decisions.

Project aims

- Development of formal models to represent mechanisms for their automated property verification.
- Development of concrete decision procedures for the automated study of complex mechanisms.
- Development of a software tool that automatically verifies commonly used mechanisms.
- Development of formal methods and proofs based on both Artificial Intelligence and automated verification techniques to certify the correctness of a selected number of complex mechanisms.

What is expected of the student when at IITB and when at Monash?

The student is expected to learn mechanism design, research on it while at IIT Bombay, and expected to develop skills in formal methods, verification, and logic side of the project while his/her stay at Monash. The interdisciplinary nature of the project requires the student to spend time with each group and therefore, their stay at both places is necessary. Both PIs are also expected to visit for a short time the partnering institution to maximize collaboration and productivity.

Expected outcomes

- A **new theoretical framework** for the automated verification of mechanisms.
- A powerful toolkit of **optimal algorithms** for the automated analysis of mechanisms.
- A **software verification tool** specifically designed to verify properties of complex mechanisms.
- A **collection of formal proofs** of desirable properties of selected mechanisms in the literature.

How will the project address the Goals of the above Themes?

Describe how the project will address the goals of one or more of the 6 Themes listed above.

<u>AI and Advanced Computational Modelling</u>: the project specifically aims to provide a *novel modeling and verification framework* for systems consisting of several autonomous and interacting AI agents. As such, the project deals directly with multi-agent systems, a well-known subarea of *distributed AI*. The project focuses on issues arising due to the presence of limited resources, uncertainty, and conflicting and stochastic behavior, making the proposed outcomes of the project closer to what is found in nowadays real-world AI systems. We note that the project squarely fits within the CSE, IT, and Optimisation cluster because of our focus of study centered around Mechanism design, a subfield of game theory, where solutions are, essentially, locally/globally optimal outcomes of a game. From an optimization perspective, solving a game is solving a multi-agent, multi-objective optimization problem. In this project, we will do that, and go further by developing formal techniques to verify the optimality of such desired outcomes.

Potential RPCs from IITB and Monash

Provide names of the potential research progress committee members (RPCs) and describe why they are most suited for the proposed project

RPCs from Monash and IITB: To be decided when a particular student is selected.

Capabilities and Degrees Required

List the ideal set of capabilities that a student should have for this project. Feel free to be as specific or as general as you like. These capabilities will be input into the online application form and students who opt for this project will be required to show that they can demonstrate these capabilities.

Applicants are expected to:

- have a background in Computer Science, AI, Logic, or Mathematics;
- have excellent written and verbal communication skills in English;
- be creative, organized, and have strong mathematical and critical thinking skills.

Necessary Courses

Name three tentative courses relevant to the project that the student should complete during his/her coursework at IITB (the student will require to secure 8 point in these courses)

CS 228 (Logic in CS)
CS 310 (Automata theory)
CS 6001 (Introduction to game theory and mechanism design)

Potential Collaborators

Please visit the IITB website www.iitb.ac.in OR Monash Website www.monash.edu to highlight some potential collaborators that would be best suited for the area of research you are intending to float.

At Monash, collaborations with various other members of the Department of Data Science and AI are very likely. For instance, with other experts in Game Theory (Julian Garcia and David Dowe), Multi-Agent Systems (Daniel Harabor), and Logic (David Ripley at the Department of Philosophy). We also currently work with researchers at the University of Oxford (Michael Wooldridge) precisely on similar topics and associated applications in AI, formal verification, and multi-agent systems. **At IITB**, we have a set of experts in formal methods (Supratik Chakraborty and Krishna S N) and multi-agent systems (Shivaram Kalyanakrishnan) who may be potential collaborators.

Select up to **(4)** keywords from the Academy's approved keyword list **(available at http://www.iitbmonash.org/becoming-a-research-supervisor/)** relating to this project to make it easier for the students to apply.

Data Science, optimization, algorithms (6)
Maths (8)

Research cluster: CSE, IT, Optimisation, Data, Sensors, Systems, Signal Processing, Control