Hippo DAI Lab: Hyper-heuristics for Interpretable Public Policy Analysis

We propose the Hippo DAI lab to perform cutting-edge research in the field of AI-based decision support for public policy design and analysis. The methodological breakthroughs from the Hippo lab will expand our insights on climate change, an urgent and a significant societal challenge. The lab will foster collaboration among the faculties of EEMCS, TPM, as well as other TU Delft and national research initiatives, and will create new educational opportunities IN and WITH AI.

1 Applicants

Jazmin Zatarain Salazar (PhD, Civil and Environmental Engineering, 2018) is an Assistant Professor in Policy Analysis (PA/MAS/TPM). Jazmin's goal is to advance methods in hyper-heuristics to enhance their value for public policy design. Jazmin is well suited for this task as demonstrated by her publications and international collaborations with pioneering experts and developers in multiobjective evolutionary optimization. She has theoretical and practical expertise to design AI-based decision support tools, and has also tested their suitability in a range of real world applications. Her recent applications focus on defining robust strategies for coping with challenging climate and socio-economic conditions, while reliably meeting environmental, energy and food demands in transboundary river basins.

Pradeep K. Murukannaiah (PhD, Computer Science, 2016) is an Assistant Professor in Interactive Intelligence (II/INSY/EEMCS). Pradeep's goal is to develop novel methods for engineering collaborative AI systems, involving autonomous principals (humans and organizations) and AI agents that represent principals. Pradeep has a strong AI background spanning multi-agent systems, automated negotiation and deliberation, and AI ethics. Pradeep is a participant in the OCW/NWO Gravitation project on Hybrid Intelligence. He collaborates with the Delft Design for Values institute and AiTech, whose goals align with those of the Hippo lab. Pradeep has been actively collaborating with TPM researchers (in the Participatory Value Evaluation project), gaining practical experience in AI use for policy deliberations.

2 Scientific Challenges

We will develop and test theory and tools to address the following major scientific challenges.

1. How can we integrate subjective notions of **ethics and fairness** into algorithmic design?
2. How can we create an **optimality taxonomy** for finding effective tradeoffs in policy search?
3. How can we develop **interpretable AI** methods to gain public confidence and provide unbiased support in real decision-making contexts?
4. How can we support public policy **deliberation** via collaborative AI methods?

We select heuristic methods as a starting point for our proposed developments due to their flexibility to (1) simultaneously find the Pareto trade-offs across multiple conflicting objectives and (2) their ability to enable complex problem abstractions that remain true to the nature of the policy problem. However, Pareto non-dominance (i.e., when one objective cannot improve without degrading performance in another objective), in isolation, is insufficient for highly consequential decision contexts.

1. We will develop novel hyper-heuristic methods that achieve fair efficiency, where solutions are evaluated not only on the basis of their non-dominance, but also on the basis of distributive justice.
2. We will explore hybrid evolutionary and memetic optimization techniques that balance global and local search, where the hyper-heuristic notion is used two-fold: (1) to enable the global search with self-adaptive operators inspired by natural evolution and biological processes; and (2) by the adaptive use of distributive justice indicators to guide the local search.
3. We will develop techniques to turn black-box heuristic methods into interpretable AI systems, which can (1) explain the processes underlying preference elicitation, policy search, and optimization, and (2) enable the exploration of a rule-base for fair and efficient policy design.
4. We will develop AI techniques to facilitate stakeholder deliberations and negotiations throughout the public policy development cycle. Our techniques will advance discourse analysis, a significant challenge in natural language processing (NLP). In particular, our techniques will elicit stakeholders' preferences to systematically formulate policy alternatives for the heuristic methods.

3 Societal Impact

We seek to develop and test our methods in the context of **climate change**, an urgent problem that presents an existential threat. The effects of climate change are expected to exacerbate existing inequalities, prompt social unrest, and lead to stark geopolitical tensions. To address the interests and well-being of all the affected parties and minimize conflict, distributive justice should be front and center in policy design. For instance, improving the situation of populations that are already better off than vulnerable communities that suffer the brunt of climate change may be Pareto optimal but is not just. The importance of considering distributive justice in climate policy, and other large societal challenges, motivates research in AI-based decision support to search for alternatives that are balanced across multiple sectors, regions, and generations, and counteract existing asymmetries in policy design. Particularly, the Hippo lab will advance AI to assure broader societal benefits in the following areas:

Equitable design The use of Pareto non-dominance has remained unchallenged since the advent of multi-objective optimization. However, in societal decision contexts involving stark asymmetries among stakeholders, Pareto non-dominance alone is insufficient to guide the exploration of alternatives. In contrast, we intend to use broader definitions of success, incorporating fairness to drive the search within our algorithmic design. Researchers recognize that existing (mis)measures of fairness may harm the very groups they seek to protect. We will systematically investigate the existing measures and develop new metrics that incorporate the broader idea of distributive justice (fairness being one aspect of it). In this endeavor, we will advance the nascent field of Fair (broadly, Ethical) AI.

Interpretability The lack of transparency hinders the use of heuristic methods in real decision-support contexts since results obtained from 'black box' heuristic methods are generally not trusted. We would not expect decision makers to leverage policy recommendations when we cannot fully explain how a certain policy advise was reached. The growing body of work on Explainable AI is focused on deep learning models. Complementing these efforts, we will focus on formalizing the patterns and pathways of bio-inspired operators when searching for efficient tradeoffs. Additionally, we will set foundations to assess interpretability, as currently there is ambiguity about what exactly constitutes interpretability in heuristic methods. Further, we will advance knowledge about the trade-offs between interpretability and predictive accuracy. In short, we are contributing a new research line on interpretable heuristic methods to the field of Explainable AI with the aim to improve accountability of AI-based decision support.

Hybrid intelligence Collective action by all stakeholders (including citizens) is necessary to tackle a global societal challenge such as climate change. Thus, we must engage the stakeholders in systematic deliberations and negotiations in each step of the policy development cycle, which can go on for years. Current (largely, manual) approaches for deliberation and negotiation are subject to significant pitfalls, including limited participation, information overload, balkanization, and a variety of cognitive biases. The hybrid methods we propose augment human intelligence with artificial intelligence in the deliberation and negotiation processes to reduce these pitfalls. In addition, engaging citizens in each step of the policy development cycle garners support for effective policy implementation.

4 Proposed PhD Projects

Figure 1 shows an overview of the four PhD projects in the Hippo lab. The applicants with complementary expertise will co-supervise each project.

PhD Thesis 1 (IN AI): Hyper-heuristic optimization for fair public policy design

- Daily supervisor: Jazmin Zatarain Salazar; Co-supervisor: Pradeep Murukannaiah
- Promotor: Alexander Verbraeck

This project will design a new hyper-heuristic algorithm for fair efficiency. The algorithmic architecture will be designed as a high-level search strategy that integrates a suite of operators that enhance its performance and flexibility as a general-purpose optimization tool for multiple domains. It will be rigorously assessed against state-of-the-art multi-objective optimization algorithms. For the local search, indicator-based strategies and participatory techniques will be explored and tested to integrate broader definitions of success, which are centered on finding fair policies.

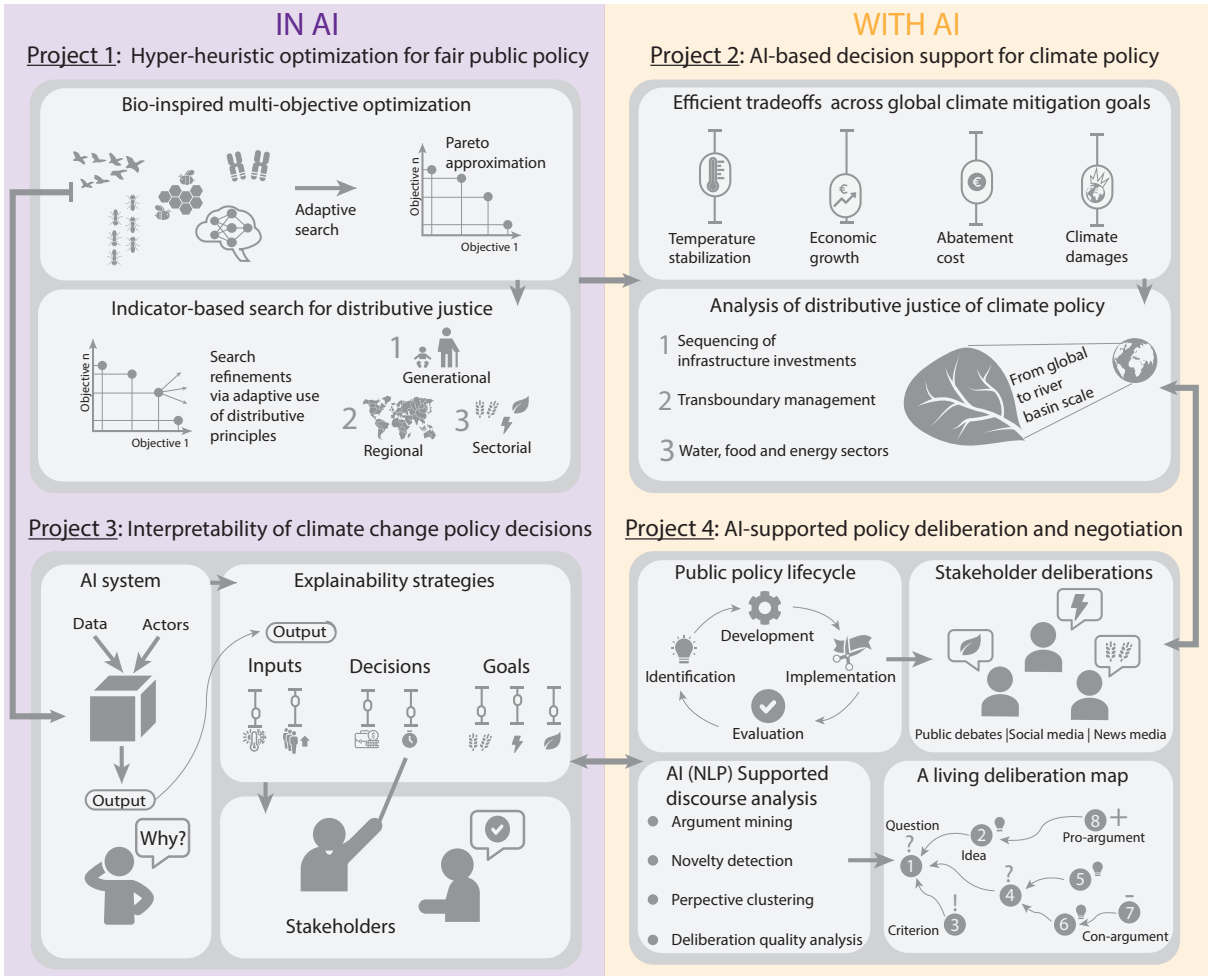

Figure 1: The proposed PhD projects and connections among them.

PhD Thesis 2 (WITH AI): AI-based decision support for climate policy design

- Daily supervisor: Jazmin Zatarain Salazar; Co-supervisor: Pradeep Murukannaiah
- Promotor: Jan Kwakkel

This project will test the ability of hyper-heuristic methods to find efficient tradeoffs across global climate mitigation policies. The consequences of these efficient strategies at the global level will then be evaluated at the transboundary river basin level to understand how they are received at a finer scale. Further, the search will be refined by testing novel definitions of distributive justice in their ability to balance regional, multi-sectorial and inter-generational conflicts. Specifically when:(1) upstream and downstream users are at odds, (2) stark asymmetries for access to water resources exist between rich and poor countries, (3) infrastructure investments need to balance inter-generational interests, and (4) competing uses need to be leveraged across the energy, food and environmental sectors. These findings can then serve as feedback to guide the search for global climate policy.

PhD Thesis 3 (IN AI): Interpretability of climate change policy decisions

- Daily supervisor: Pradeep Murukannaiah; Co-supervisor: Jazmin Zatarain Salazar
- Promotor: Catholijn Jonker

This project, in collaboration with the other three projects, will develop strategies to enhance the interpretability of the proposed methods for policy deliberation, search, and optimization. In particular, we will investigate integrated as well as post-hoc methods for interpretability. The integrated methods focus on making the preferences, processes, and metrics used in policy analysis transparent, and providing tools to study the influence of these parameters on the final policy. The post-hoc methods are used to

learn rules on top of black-box optimization techniques to explain optimal choices.

PhD Thesis 4 (WITH): Engaging stakeholders in public policy deliberations

- Daily supervisor: Pradeep Murukannaiah; Co-supervisor: Jazmin Zatarain Salazar
- Promotor: Catholijn Jonker

The search and optimization techniques in the above projects start from well-formulated (albeit conflicting) objectives. However, formulating such objectives requires extensive deliberation and negotiation among, e.g., political leaders, scientists, and citizens. This project will contribute the development of an online platform for public policy deliberations and negotiations. Our hybrid deliberation platform will support humans with AI (specifically, via natural language processing, and knowledge representation and reasoning techniques) to recognize the deliberative structure of discussion, cluster and visualize perspectives, and find novel objectives. Similar to the other WITH AI project, we will test our methods on the use case of global climate change policy on water management in transboundary river basins.

5 Education

The Hippo lab will exploit expertise from both TPM and EEMCS to open curriculum development opportunities that reduce an existing campus-wide educational gap (Figure 2) with the following courses.

Applied Hyper-Heuristic Optimization For many optimization problems, there is no feasible way to find optimal solutions. Heuristic optimization methods are able to effectively find trade-offs for complex systems that need to meet multiple goals, and have multi-modal, stochastic and non-convex search spaces. The goal of this course is to offer students an introduction to popular heuristic methods, as well as their guiding principles and applications in different domains, granting them hands-on experience and intuition about their strengths and weaknesses.

Multi-Agent Systems AI systems are often developed as centralized solutions. Although AI systems support distributed computing, often, they are conceptually centralized in that a single entity controls the system. This course will teach the design and development of decentralized AI systems, involving autonomous social principals (humans and organizations) as well as technical entities (agents) that represent the principals. The course will introduce computational modeling of challenges such as trust, values, autonomy, and accountability essential for AI adoption in societal applications.

Natural Language Processing An essential aspect of any sociotechnical system is the interaction among the actors. With human actors, the interaction often involves a discourse in natural language. We will develop a course to teach natural language processing (NLP) techniques for discourse analysis. Since EEMCS does not have an NLP course currently, we (in collaboration with other interested faculty) will develop a course on NLP foundations. In addition, we will develop project ideas for public discourse analysis building on the NLP foundations, which can potentially be extended as thesis projects.

These courses can operate as part of a joint minor on **AI for Policy Analysis** which capitalizes on existing courses in TPM and EEMCS. With this minor, TPM students will benefit by acquiring the necessary knowledge to design AI systems through the AI (or Algorithmics) track, and EEMCS students can pursue a Policy track with a range of application domains offered at TPM. In Figure 2, we select existing courses within the curricula from both faculties which offer complementary knowledge and that would enable students to design and apply AI-based decision support.

The Hippo lab will also benefit from recent TU Delft HPC investments required to test our proposed methods in hyper-heuristic optimization, which requires extended search to learn which operators yield acceptable performance; thus, benefiting from parallelization. As a result, we intend to provide guidance on parallel computing applications through pedagogical material.

6 Timeline

We are prepared to start the PhD student recruiting process immediately, if the proposal is accepted. We anticipate the recruitment process to take a maximum of three months. The suggested start date for all four PhD students will be within six months of the deliberation, i.e., within September 2021. We will

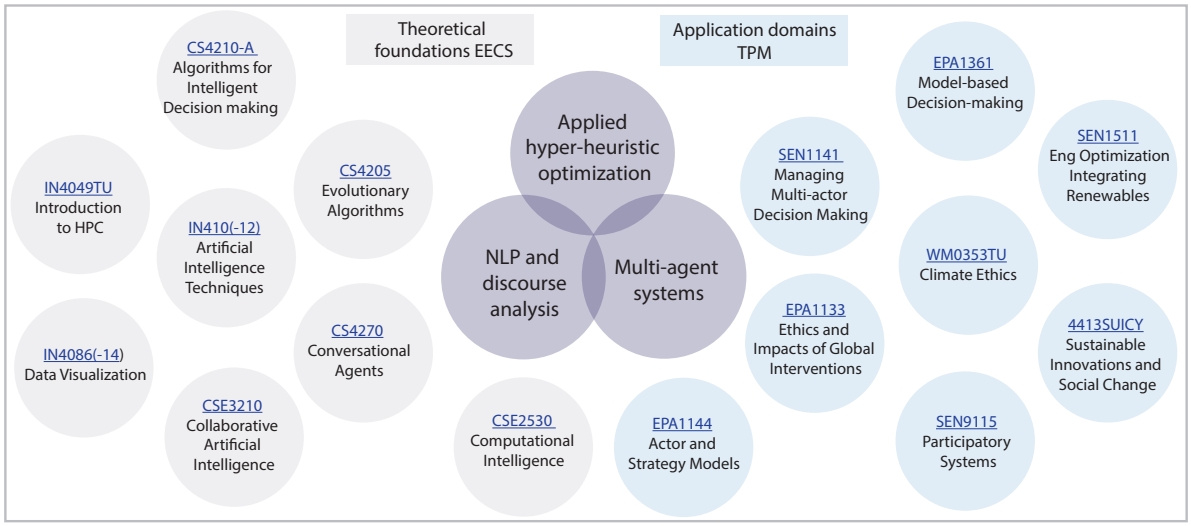

Figure 2: The coursework Hippo lab proposes to develop.

also start the proposed course design and preparation immediately. We expect to start teaching at least one of the proposed courses in the Academic Year 2021–2022.

7 Growth Plan

The Hippo lab's goals are important for both TPM and EEMCS. There are clear overlapping interests in the two faculties both in terms of the AI techniques as well as the applications. The lab establishes a partnership between both faculties to answer questions related to: How to make AI fair and useful for addressing societal challenges? How can AI support systematic multi-stakeholder deliberations on societal challenges? How to minimize unintended consequences of AI-based decision support? How to improve accountability in algorithmic design? In addition to collaboration between the Policy Analysis (TPM) and Interactive Intelligence (EEMCS) sections, we envision the Hippo lab to collaborate with:

- Algorithmics section Software Technology (EEMCS) for research on meta-heuristics design.
- Ethics/Philosophy of Technology section in Values, Technology and Management (TPM) and the recently established TPM AI Lab for research in ethics and accountable AI design.
- Climate Institute for research related to climate information and policy.
- Delft Design for Values and AiTech on incorporating human values in policy deliberations.
- Hybrid Intelligence Centre on the collaborative, responsible, and explainable hybrid intelligence.

To sustain and grow the Hippo lab beyond the DAI labs initiative, we plan (1) joint submissions to national calls such as the NWA Routes, Gravitation, and Perspectives, as well as to international calls such as the H2020 (and its sequels); (2) formulating collaborations with industry, non-governmental organizations, and governmental organizations in the form of five-year ICAI labs; and (3) preparing solid proposals when structural money for AI will become available, such as the Sector plans, Hoekstra funding, and NL AI Coalition. The lab will also tap into a growing network of researchers in Environmental Intelligence worldwide, with interest in artificial intelligence techniques targeted to solve global challenges, including scholars in Politecnico di Milano, Cornell University, Stanford University, UCDavis, Tufts University, University of West Virginia, National University of Singapore, and ETH Zurich with whom the applicants are collaborating. These collaborations are well suited to target joint calls such as the H2020 FETPROACT call on Environmental Intelligence. Other suitable funding streams for the lab are through the NWO call on *Open Competition domain science* and the NWO call on *Data and Intelligence: A safe society with the help of data and intelligence*. Further, we have growing access to a number of river basin case studies worldwide through our international network including the Mekong, Omo-Turkana, Zambezi, Red River, the Nile, and the Mexican Valley river basins, which provide great opportunities to test the scientific developments proposed within the Hippo lab.