# Preliminary Schedule (Subject to Change)<sup>1</sup>

## Monday, October 27th, 2025 -

09:00-09:15	Welcome by the Organizers by Aina Niemetz, Guido Tack, Johannes Fichte, Matti Järvisalo	Johannes Fichte
09:15-10:00	Introduction by the Participants by $-The\ Participants-$	(1 minute each)
10:00-10:45	Coffee and Tea	
10:45-12:00	Talks (Competition Surveys I) by • Sutcliffe Geoff (CASC) • Ashlin Iser (SAT Competition) • Matti Järvisalo (MaxSAT Evaluation) • Markus Hecher (#SAT Competition)	(10+5 mins) Aina Niemetz
12:15-14:00	Lunch	
14:00-15:00	Talks (Competition Surveys II) by  • Jason Nguyen (CP: MiniZinc Challenge)  • Olivier Roussel (Pseudo-Boolean Competition)  • Martin Jonáš (SMT-COMP)	(10+5 mins) Matti Järvisalo
15:00-16:00	Coffee and Cake	
16:00-17:20	Collecting Challenges Challenges, focus, schedule updates for the week  • Topic Suggestions: Infrastructure, Bias, Repeatability, Ranking, Interview	Johannes Fichte
17:20-18:00	Free Time / Discussions	
18:00-19:30	Dinner	

 $<sup>^1\</sup>mathrm{Suggestions}$  for possible discussion topics on the last page

## Tuesday, October 28th, 2025 -

09:00-10:00	Survey / Tutorial Talk (Benchmark Selection) by Kate Smith-Miles		Guido Tack
	Optimal selection of benchmarking datasets / Portfolios Algorithm Selection / Diversity and Robustness of Test Suites Anomaly Detection in High-Dimensional Data		
			(~50+10)
10:00-10:45	Coffee and Tea		
10:45-12:00	Talks (Competition Surveys III) by • Dirk Beyer (SV-Comp) • Mathias Preiner (HWMCC) • Florian Frohn (termComp) • Guillermo Pérez (SYNTCOMP)	(10+5 mins)	Matti Järvisalo
12:15-14:00	Lunch		
14:00-15:15	Talks (Competition Surveys IV and Formats)  (10+5 mins) Johannes Fichte by  Mario Alviano (ASP Competition; LP/CP Programming Contest)  Andreas Niskanden (ICCMA)  Malte Helmert / Scott Sanner (Classical IPC/ Probabilistic Tracks Planning Competition)  Scott Sanner  (The RDDL Language and Planning Competitions 15 Years of Trying to make Impact)		
15:15-16:00	Coffee and Cake		
16:00-17:20	Discussions and Summary of Competition Challenges — Common Problems, Goals, Directions —		Johannes Fichte
17:20-18:00	Free Time / Discussions (tbd) by —Decision Participants—		Groups
18:00-19:30	Dinner		

## Wednesday, October 29th, 2025 -

09:00-10:00	Tutorial Talk (Infrastructure and Evaluation)	Aina Niemetz
	BenchExec by Philipp Wendler	(~45+15)
10:00-10:45	Coffee and Tea	
10:45-11:15	Talks (Infrastructure and Evaluation) BenchCloud by Dirk Beyer https://doi.org/10.1145/3691620.3695358	Johannes Fichte $(\sim 20+5)$
11:15-12:00	Discussions / Panel. Topic tbd, e.g., Infrastructure Requirements	Johannes Fichte
12:00-12:15	Group Photo by — All Participants —	I
12:15-14:00	Lunch	
14:00-18:00	<b>Hike / Excursion</b> by — Interested Participants —	
18:00-19:30	Dinner	

## Thursday, October 30th, 2025 -

09:00-10:00		er
10:00-10:30	Coffee and Tea	
10:30-11:30	Talks (Sustainability / Reproducibility)  • Global Benchmark Database by Ashlin Iser  https://doi.org/10.4230/LIPIcs.SAT.2024.18  • Understanding 20 Years of SMT Benchmarks by Hans-Jörg Schurr  https://zenodo.org/records/16290040 /  https://schurr.io/pubs/smt2025-slides.pdf  (~20+5)	
11:00-12:00	<ul> <li>— Joint Session with Research Meeting 25444 —</li> <li>Better Benchmarking Setups for Optimisation: Design, Curation and Long-Term Evolution</li> <li>— PANEL —-</li> </ul>	
12:15-14:00	Lunch	ı
14:00-14:30	<b>Discussions</b> FLoC Olympics by Matti Järvisalo	
14:30-15:00	<ul> <li>Discussions</li> <li>Public infrastructure for rolling open competitions in NP-hard domains</li> <li>Post StarExec Options</li> <li>Define a competition infrastructure that is open-source and independent</li> <li>Public infrastructure for rolling open competitions</li> </ul>	1
15:00-16:00	Coffee and Cake	
16:00-17:20	Talks / Group Works/ Discussions (tbd) by —Decision Participants—	is
17:20-18:00	Free Time / Discussions (tbd)	
18:00-19:30	Dinner	

#### Friday, October 31th, 2025 -

09:00-10:00	Outcomes by Discussion with the Participants and Outcomes of the Group Works	Guido Tack / Johannes Fichte
10:00-10:30	Coffee and Tea	
10:30-11:00	<b>Talk on Benchmarking Libraries</b> Quantum Optimization Benchmarking Library – The Intractable Decat by <i>Thorsten Koch</i>	Johannes Fichte hlon
11:00-12:00	Discussions (TBD) by — Decision by Participants —	
12:00-12:15	Closing	Guido Tack
12:15-14:00	Lunch	

## Possible Topics for Discussion:

#### Benchmark selection

- How can we get "good" benchmarks (and what does "good" mean here)?
- How can we ensure a solid empirical methodology in the community?
- How can we select a "fair" subset from the available benchmarks (and what does "fair" mean here)?
- How can we avoid running on (more or less) the same benchmarks every year without introducing too much randomness, such that the results of the competition become meaningless?
- How can we construct representative sets beyond annual competitions?
- How can we reduce overfitting of algorithms and solvers with effective measures that are reasonable for researchers and the community?
- Can we provide inclusive access for researchers with limited computational infrastructure access?
- Select benchmark sets and how to report on an experimental evaluation

### Best practices

- What makes a competition to a stable, long-lasting community support?
- Role of competitions in setting standards for rigorous empirical research
- Standards for how to perform experimental evaluations
- "The SAT Practitioner's Manifesto": Pros and Cons
- Useful measures
- Scientific progress and contributions (Generative AI to produce solvers?)
- Challenge Problems for Scientific Computing Correctness

#### Evaluation

- Solver engineering, algorithm design, and application of the tools
- Empirical community is a robust and effective evaluation
- Is it beneficial to design shared tools and frameworks to access public (computation clusters) or private (AWS) infrastructure effectively?
- Is it reasonable and sustainable in the long-term to establish a central system (such as StarExec) for empirical research and/or competitions in AR and how can it be established?
- How can we support replication, artifact evaluation, and improving software quality in the field of AR in general?