

Monday, October 27th, 2025 –

09:00-09:15	<b>Welcome by the Organizers</b> by <i>Aina Niemetz, Guido Tack, Johannes Fichte, Matti Järvisalo</i>	Johannes Fichte
09:15-10:00	<b>Introduction by the Participants</b> by — <i>The Participants</i> —	(1 minute each)
10:00-10:45	<b>Coffee and Tea</b>	
10:45-12:00	<b>Talks (Competition Surveys I)</b> by <ul style="list-style-type: none"> <li>• <i>Geoff Sutcliffe</i> (CASC)</li> <li>• <i>Ashlin Iser</i> (SAT Competition)</li> <li>• <i>Matti Järvisalo</i> (MaxSAT Evaluation)</li> <li>• <i>Markus Hecher</i> (#SAT Competition)</li> </ul>	(10+5 mins) Aina Niemetz
12:15-14:00	<b>Lunch</b>	
14:00-15:30	<b>Talks (Competition Surveys II)</b> by <ul style="list-style-type: none"> <li>• <i>Jason Nguyen</i> (CP: MiniZinc Challenge)</li> <li>• <i>Olivier Roussel</i> (Pseudo-Boolean Competition)</li> <li>• <i>Martin Jonáš</i> (SMT-COMP)</li> </ul>	(10+5 mins) Matti Järvisalo
15:30-16:00	<b>Coffee and Cake</b>	
16:00-17:30	<b>Collecting Challenges</b> Challenges, focus, schedule updates for the week <ul style="list-style-type: none"> <li>• <i>Topic Suggestions: Infrastructure, Bias, Repeatability, Ranking, Interfaces</i></li> </ul>	Johannes Fichte
17:30-18:00	<b>Free Time / Discussions</b>	
18:00-19:30	<b>Dinner</b>	

Tuesday, October 28th, 2025 –

09:00-10:00	<b>Survey / Tutorial Talk (Benchmark Selection)</b> by <i>Kate Smith-Miles</i> Optimal selection of benchmarking datasets / Portfolios Algorithm Selection / Diversity and Robustness of Test Suites Anomaly Detection in High-Dimensional Data	Guido Tack
		(~50+10)
10:00-10:45	<b>Coffee and Tea</b>	
10:45-12:00	<b>Talks (Competition Surveys III)</b> by <ul style="list-style-type: none"> <li>• <i>Dirk Beyer</i> (SV-Comp)</li> <li>• <i>Mathias Preiner</i> (HWMCC)</li> <li>• <i>Florian Frohn</i> (termComp)</li> </ul>	(10+5 mins) Aina Niemetz
12:15-14:00	<b>Lunch</b>	
14:00-15:00	<b>Talks (Competition Surveys IV)</b> by <ul style="list-style-type: none"> <li>• <i>Guillermo Pérez</i> (SYNTCOMP)</li> <li>• <i>Mario Alviano</i> (ASP Competition; LP/CP Programming Contest)</li> <li>• <i>Andreas Niskanen</i> (ICCMA)</li> </ul>	(10+5 mins) Matti Järvisalo
15:00-16:00	<b>Coffee and Cake</b>	
16:00-17:20	<b>Talks (Competition Surveys V) and Discussions on Competition Challenges</b> <ul style="list-style-type: none"> <li>• <i>Lars Kotthoff</i> (Using the Shapley value to analyze competition results)</li> <li>• <i>Malte Helmert</i> (Classical IPC)</li> <li>• <i>Scott Sanner</i>              (Probabilistic Tracks Planning Competition +              The RDDDL Language and Planning Competitions 15 Years of Trying to make Impact)</li> <li>• <i>Johannes Fichte</i> (Discussion Competition Common Problems, Goals, Directions)</li> </ul>	Johannes Fichte
17:20-18:00	<b>Free Time / Discussions (tbd)</b> by — <i>Decision Participants</i> —	Groups
18:00-19:30	<b>Dinner</b>	

Wednesday, October 29th, 2025 –

09:00-10:00	<b>Tutorial Talk (Infrastructure and Evaluation)</b> BenchExec by <i>Philipp Wendler</i> <a href="https://doi.org/10.1007/s10009-017-0469-y">https://doi.org/10.1007/s10009-017-0469-y</a>	Aina Niemetz
		(~45+15)
10:00-10:45	<b>Coffee and Tea</b>	
10:45-11:30	<b>Talks (Infrastructure and Evaluation)</b> • <i>BenchCloud</i> by <i>Dirk Beyer</i> <a href="https://doi.org/10.1145/3691620.3695358">https://doi.org/10.1145/3691620.3695358</a>	Matti Järvisalo
11:30-12:00	<b>Discussions (FLoC Olympics)</b> • <i>Collaborations</i> by <i>Marie Anastacio</i> • <i>Joining Competitions</i> : <i>Laurent Simon</i> • <i>FLoC</i> by <i>Johannes Fichte</i> , <i>Matti Järvisalo</i> , <i>Guido Tack</i> , <i>Aina Niemetz</i> , <i>Armin Biere</i>	Guido Tack
12:00-12:15	<b>Group Photo</b> by — <i>All Participants</i> —	
12:15-14:00	<b>Lunch</b>	
14:00-18:00	<b>Hike</b> by — <i>Interested Participants</i> —	
15:00-16:00	<b>Coffee and Cake</b>	
18:00-19:30	<b>Dinner</b>	

Thursday, October 30th, 2025 –

09:00-10:00	<b>Talks (Evaluation) and Discussion</b> <ul style="list-style-type: none"> <li>• <i>Understanding 20 Years of SMT Benchmarks</i> by Hans-Jörg Schurr  <a href="https://zenodo.org/records/16290040/">https://zenodo.org/records/16290040/</a>  <a href="https://schurr.io/pubs/smt2025-slides.pdf">https://schurr.io/pubs/smt2025-slides.pdf</a></li> <li>• <i>StarExec in the Cloud</i> by Geoff Sutcliffe  <a href="https://doi.org/10.1007/s10009-017-0469-y">https://doi.org/10.1007/s10009-017-0469-y</a></li> <li>• <i>Discussions: Requirements on Competition Infrastructure</i> by Ciran McCreesh</li> </ul>	Aina Niemetz
10:00-10:45	<b>Coffee and Tea</b>	
10:45-12:00	— <b>Joint Session with Research Meeting 25444</b> — <ul style="list-style-type: none"> <li>• <i>Global Benchmark Database</i> by Ashlin Iser  <a href="https://doi.org/10.4230/LIPIcs.SAT.2024.18">https://doi.org/10.4230/LIPIcs.SAT.2024.18</a></li> <li>• <i>Lessons from Continuous Optimization</i> by Anna Kononova and Olaf Mersmann</li> </ul>	Guido Tack
12:15-14:00	<b>Lunch</b>	
14:00-15:00	<b>Panel on Benchmarks</b> —Panelists— Erika Ábrahám, Malte Helmert, Mathias Preiner, and Laurent Simon	Markus Hecher
15:00-16:00	<b>Coffee and Cake</b>	
16:00-17:00	<b>Discussions</b> by — <i>Decision Participants</i> — <ul style="list-style-type: none"> <li>• <i>Controversial Evaluation Statements and Benchmarking</i> by Ciran McCreesh</li> <li>• <i>Initiative on Contributing Benchmark Instances</i> by Daniel Le Berre  <a href="https://mysolvetimesout.org">https://mysolvetimesout.org</a></li> </ul>	Ciran McCreesh
17:20-18:00	<b>Free Time</b>	
18:00-19:30	<b>Dinner</b>	



Survey on competition details  
<https://forms.gle/4pX9pqReFech3fFv6>  
organized by Geoff Sutcliffe

Friday, October 31th, 2025 –

09:00-10:00	<b>Outcomes and Follow Ups</b> by — <i>all Participants</i> — <ul style="list-style-type: none"> <li>• <i>Distribution and Acknowledgement of Benchmarks</i> by Erika Abraham</li> <li>• <i>Journal Special Track on Competitions</i> Aina Niemetz</li> <li>• <i>ERC Cost Action</i> by Erika Abraham and Matti Järvisalo</li> <li>• <i>Future Workshops/Research Meetings</i> by Geoff Sutcliffe</li> <li>• <i>Possible FLoC Events</i></li> <li>• <i>Comprehensive Survey Paper on AR Competitions</i> by Guido Tack</li> </ul>	Johannes Fichte / Matti Järvisalo
10:00-10:30	<b>Coffee and Tea</b>	
10:30-11:15	<b>Talk on Benchmarking Libraries (Virtual)</b> Quantum Optimization Benchmarking Library – The Intractable Decathlon by <i>Thorsten Koch</i> <a href="https://arxiv.org/abs/2504.03832">https://arxiv.org/abs/2504.03832</a>	Johannes Fichte
11:15-12:00	<b>Final Discussions and Closing</b> by — <i>all Participants</i> — <ul style="list-style-type: none"> <li>• <i>Joint Website on AR Competitions</i> by Geoff Sutcliffe</li> <li>• <i>Closing</i></li> </ul>	Guido Tack
12:15-14:00	<b>Lunch</b>	

## Discussion Topics:

### 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?*