Dagstuhl Perspectives Workshop 26162

Autonomous AI Agents in Computer Security

April 12–17, 2026 | Schloss Dagstuhl, Wadern, Germany

Organizers: Alvaro Cárdenas (UC Santa Cruz) · Kathrin Grosse (Independent Researcher)

· Nicole Nichols (Palo Alto Networks) · Konrad Rieck (TU Berlin)

Tentative agenda subject to change.

Sunday, April 12 — Arrival

15:00–19:00	**Check-in**	Self-service check-in	
18:00	*Dinner*	Welcome dinner — randomized seating begins	
20:00–late	**Evening Social**	Wine bar, billiards, foosball, ping pong.	

Monday, April 13 — Discover & Commit

7-7:45am	**Running/biking club**	Optional activity every morning.	
7:30-8:45am	**Breakfast**		
09:00–09:30	**Workshop Opening**	Organizers present: workshop goals, the week's arc, connection to IEEE S&P special issue (deadline May 1), and the Dagstuhl Manifesto deliverable. Set expectations: every participant leaves with a group, a position, and a writing plan.	
09:30–10:30	**Participant Introductions**	Each participant: 2 min, one slide. Format: your name, one sentence on your expertise, and one bold claim or open question about AI agents in security.	***Lightning round***
10:30–10:45	*Coffee Break*		
10:45–12:15	**Participant Introductions**	Continued — remaining participants	***Lightning round***
12:15–14:00	*Lunch*	Randomized seating	
14:00–15:00	**Problem Mapping**	Each participant writes 2–3 research directions, challenges, topics of interest, on large cards (one per card). Cards go on the wall. Silent reading (10 min). Then collaborative clustering: participants physically move cards into thematic groups, discussing as they go.	***Post-it + affinity mapping***
15:00–15:30	*Coffee & Cake*	Served in dinner hall	
15:30–16:15	**Speed Discussions**	Top 6 challenge clusters assigned to 6 tables. Participants rotate through 3 tables of interest (15 min each). At each table: Is this the right framing? What's the core question? What would a paper argue?	***Speed dating / carousel***
16:15–16:45	**Priority Vote**	Each participant gets 3 dot-votes. Vote on the challenge clusters you most want to work on. Top 5–6 clusters survive. Results announced immediately.	***Dot-voting***
16:45–17:30	**Working Group Formation**	Surviving clusters become working groups. Self-select (organizers balance if needed). Each group picks a lead, a scribe, and a devil's advocate.	***Self-selection + roles***
17:30–18:00	**Group Kickoff**	First working session:	
18:00	*Dinner*	Randomized seating — cross-pollinate ideas	
20:00–late	**Evening Social**	Wine bar, billiards, foosball, ping pong.	

Tuesday, April 14 — Write & Stress-Test

7-7:45am	**Running/biking club**	Optional activity every morning.	
7:30-8:45am	**Breakfast**		
09:00–09:45	**Research Talks I (TBD)**	Somesh Jha: Formal Approaches to Securing Agents Monica Landoni: Collaborative Design	
09:45–10:30			
10:30–10:45	*Coffee Break*		
10:45–12:15	**Group Kickoff**	First working session: agree on the core claim your group will defend. What do you believe is true? Why does it matter?	
12:15–13:45	*Lunch*	Randomized seating	
13:45–14:00	**Group Picture**	Meet at the entrance gate and walk to chapel	
14:00–14:45	**Research Talks II (TBD)**	Wenke Lee: Constraint-Driven, Threat-Intelligence-Aware Self-Auditing AI Developers Andrew Paverd: Challenges and Opportunities for Vulnerability Response	
15:30–16:00	*Coffee & Cake*		
16:00–18:00	**Group Work**	Agree on the core claim your group will defend. What do you believe is true? Why does it matter?	
		Prepare key points to present on Wednesday	
18:00	*Dinner*		
20:00–late	**Evening Social**	Wine bar, billiards, foosball, ping pong.	

Wednesday, April 15 — Deepen & Recharge

7-7:45am	**Running/biking club**	Optional activity every morning.	
7:30-8:45am	**Breakfast**		
09:00–10:30	**Position Statement Presentations**	Each group presents their position (5 min) followed by constructive challenge from the room (5 min).	
10:30–10:45	*Coffee Break*		
10:45–12:15	**Breakout Sessions**	key arguments per section	***Paper skeleton***
12:15–14:00	*Lunch*		
14:00–20:00	**Social Event**	Traditional Dagstuhl Wednesday afternoon excursion. Hike, sightseeing, or activity.	***Dagstuhl tradition***
18:00	*Dinner*		
20:00–late	**Evening Social**	Wine bar, billiards, foosball, ping pong.	

Thursday, April 16 — Present & Plan

7-7:45am	**Running/biking club**	Optional activity every morning.	
7:30-8:45am	**Breakfast**		
09:00–10:30	**Group Presentations I**	Groups 1–3: 15 min presentation + 15 min plenary discussion each.	***Presentation + Q&A***

10:30–10:45	*Coffee Break*		
10:45–12:15	**Group Presentations II**	Groups 4–6: 15 min presentation + 15 min plenary discussion each.	***Presentation + Q&A***
12:15–14:00	*Lunch*	Randomized seating	
14:00–15:30	**Research Talks III (TBD)**	Mario Fritz: TBD Lorenzo Cavallaro: TBD	
15:30–16:00	*Coffee & Cake*		
16:00–17:00	**Group Discussion**	Finalize writeup and prepare plans for future if group wants to submit to the special issue of IEEE S&P	
17:00–17:45			
18:00	*Dinner*	Last group dinner	
20:00–late	**Farewell Social**	Wine bar, music room. Final night together.	

Friday, April 17 — Wrap-Up & Departure

7-7:45am	**Running/biking club**	Optional activity every morning.	
7:30-8:45am	**Breakfast**		
09:00–10:00	**Future Directions Discussion**	Open plenary: What did the workshop reveal? What collaboration opportunities continue beyond Dagstuhl? What research should happen next?	***Open plenary***
10:00–10:30	**Milestones**	Which groups are writing papers?	
10:30–10:45	*Coffee Break*		
10:45–11:30	**Feedback & Survey**	Open feedback round and Dagstuhl survey.	***Dagstuhl survey***
11:30–12:00	**Closing Remarks**	Summary of outcomes, thanks, and next steps. Appoint collector for Dagstuhl Reports abstracts.	
12:15–13:15	*Lunch & Departure*	Safe travels!	

Key Dates: IEEE S&P submission: May 1, 2026 | Publication: Nov/Dec 2026 | Dagstuhl Manifesto: ~8–12 weeks post-workshop
Meals: Breakfast 7:30–8:45 | Lunch 12:15 | Dinner 18:00 (seating randomized) | Coffee 10:00–11:00 | Cake 15:00–16:00

Alternative Schedule Options Suggested by Claude:

Three alternative approaches are presented below. Each preserves the same Sunday arrival, Friday wrap-up, Wednesday social, Thursday panel, and paper planning session.

At a Glance

	Option A: Groups First	Option B: Discussion-Driven	Option C: Inspire Then Build
Group formation	Mon afternoon	Tue afternoon	Tue afternoon
Total group work	~12 hours	~7 hours	~10 hours
Invited talk slots	3	3	5
Plenary discussion	~4 hours	~8 hours	~4 hours
Best for	Max output, concrete paper drafts	Community alignment on challenges	Diverse group needing shared grounding

Option A: Groups First

Form groups early, maximize dedicated working time. Talks are brief catalysts, not the main event.

Monday, April 13 — Meet, Map, Mobilize		
09:00–09:30	**Workshop Opening**	Goals, structure, IEEE S&P special issue overview, Manifesto process
09:30–10:30	**Participant Introductions**	2–3 min per person (one slide): your expertise and the challenge that drives your work
10:30–10:45	*Coffee Break*	
10:45–12:15	**Participant Introductions**	Continued
12:15–14:00	*Lunch / Siesta*	
14:00–14:45	**Invited Talk I (TBD)**	Scene-setting overview of the landscape
15:00–15:30	*Coffee & Cake*	
15:30–16:30	**Challenge Mapping**	Pitch open problems, cluster via affinity mapping on a shared board
16:30–17:00	**Priority Vote & Group Formation**	Dot-voting. Top 5–6 themes become working groups. Self-select.
17:00–18:00	**Working Group Kickoff**	Define scope, assign roles, sketch initial outline
18:00	*Dinner*	

Tuesday, April 14 — Deep Work I		
09:00–09:45	**Invited Talk II (TBD)**	Targeted talk to spark new angles
09:45–10:30	**Breakout Sessions**	Working group time
10:30–10:45	*Coffee Break*	
10:45–12:15	**Breakout Sessions**	Develop arguments, identify gaps

12:15–14:00	*Lunch / Siesta*	
14:00–15:30	**Breakout Sessions**	Cross-group ambassadors encouraged
15:30–16:00	*Coffee & Cake*	
16:00–17:00	**Mid-Week Check-In**	5-min progress flash per group. Plenary feedback.
17:00–18:00	**Open Working Time**	Groups continue or explore cross-group ideas
18:00	*Dinner*	

Wednesday, April 15 — Deep Work II & Social

09:00–10:30	**Breakout Sessions**	Refine findings, begin drafting presentations
10:30–10:45	*Coffee Break*	
10:45–12:15	**Breakout Sessions**	Prepare presentation outlines
12:15–13:15	*Lunch*	
13:15–14:00	**Invited Talk III (TBD)**	Industry or practitioner perspective
14:00–20:00	**Social Event**	Afternoon excursion organized by Schloss Dagstuhl

Thursday, April 16 — Present & Synthesize

09:00–10:30	**Group Presentations I**	Groups 1–3: 15 min + 15 min plenary discussion each
10:30–10:45	*Coffee Break*	
10:45–12:15	**Group Presentations II**	Groups 4–6: 15 min + 15 min plenary discussion each
12:15–14:00	*Lunch / Siesta*	
14:00–15:30	**Panel: AI Agents and Society**	Podium discussion (4 panelists, preferably industry)
15:30–16:00	*Coffee & Cake*	
16:00–17:00	**Paper Planning for IEEE S&P**	Crystallize paper ideas, co-author teams, milestones
17:00–17:45	**Manifesto Planning**	Key messages, writing leads, section owners
18:00	*Dinner*	

Friday, April 17 — Wrap-Up

09:00–10:00	**Future Directions Discussion**	Open plenary: emerging themes, collaboration opportunities
10:00–10:30	**Writing Milestones**	Lock in timelines for manifesto and IEEE S&P submissions
10:30–10:45	*Coffee Break*	
10:45–11:30	**Feedback & Survey**	Open feedback round and Dagstuhl survey
11:30–12:00	**Closing Remarks**	Summary, thanks, next steps
12:15–13:15	*Lunch & Departure*	

Option B: Discussion-Driven

Structured plenary discussions and debates throughout the week to surface diverse perspectives before focused group work.

Monday, April 13 — Landscape & Provocations		
09:00–09:30	**Workshop Opening**	Goals, structure, IEEE S&P special issue, Manifesto process
09:30–10:30	**Participant Introductions**	2–3 min per person: your expertise and one bold claim about AI agents in security
10:30–10:45	*Coffee Break*	
10:45–12:15	**Participant Introductions**	Continued
12:15–14:00	*Lunch / Siesta*	
14:00–14:45	**Invited Talk I (TBD)**	Setting the stage — state of the field
15:00–15:30	*Coffee & Cake*	
15:30–17:00	**Fishbowl: Offensive vs. Defensive AI Agents**	Structured fishbowl debate. Rotating seats, open mic.
17:00–17:45	**Challenge Harvesting**	Capture open questions and tensions from the day. Organize on shared wall.
18:00	*Dinner*	

Tuesday, April 14 — Explore & Converge		
09:00–09:45	**Invited Talk II (TBD)**	Deep dive on a specific challenge area
09:45–10:30	**Invited Talk III (TBD)**	Complementary perspective (e.g., industry vs. academia)
10:30–10:45	*Coffee Break*	
10:45–12:15	**World Café**	Rotating small-group discussions on 5–6 themes. 25 min per table, then rotate.
12:15–14:00	*Lunch / Siesta*	
14:00–15:30	**World Café Readouts & Plenary Debate**	Table hosts report back. Open debate to sharpen and prioritize.
15:30–16:00	*Coffee & Cake*	
16:00–16:45	**Working Group Formation**	Top themes become groups. Self-select. Define scope, leads, scribes.
16:45–18:00	**Working Group Kickoff**	Initial brainstorming and work plan
18:00	*Dinner*	

Wednesday, April 15 — Breakout & Social		
09:00–10:30	**Breakout Sessions**	Working group time — develop ideas, structure arguments
10:30–10:45	*Coffee Break*	
10:45–12:15	**Breakout Sessions**	Cross-group ambassadors encouraged
12:15–13:15	*Lunch*	
14:00–20:00	**Social Event**	Afternoon excursion organized by Schloss Dagstuhl

Thursday, April 16 — Present, Debate, Plan		
09:00–10:30	**Group Presentations I**	Groups 1–3: 15 min + 15 min plenary discussion each
10:30–10:45	*Coffee Break*	
10:45–12:15	**Group Presentations II**	Groups 4–6: 15 min + 15 min plenary discussion each
12:15–14:00	*Lunch / Siesta*	
14:00–15:00	**Cross-Cutting Discussion**	Identify connections, contradictions, and gaps across all groups
15:00–15:30	*Coffee & Cake*	

15:30–16:30	**Panel: AI Agents and Society**	Podium discussion (4 panelists, preferably industry)
16:30–17:15	**Paper Planning for IEEE S&P**	Co-author teams, paper outlines, milestones
17:15–17:45	**Manifesto Planning**	Key messages and writing responsibilities
18:00	*Dinner*	

Friday, April 17 — Wrap-Up

09:00–10:00	**Future Directions Discussion**	What did we miss? What should the community prioritize?
10:00–10:30	**Writing Milestones**	Confirm timelines for manifesto and IEEE S&P papers
10:30–10:45	*Coffee Break*	
10:45–11:30	**Feedback & Survey**	Open feedback round and Dagstuhl survey
11:30–12:00	**Closing Remarks**	Summary, thanks, next steps
12:15–13:15	*Lunch & Departure*	

Option C: Inspire Then Build

Front-load talks to build shared context, then pivot entirely to working group time. Two clear phases: absorb, then create.

Monday, April 13 — Context & Perspectives I

09:00–09:30	**Workshop Opening**	Goals, structure, IEEE S&P special issue, Manifesto process
09:30–10:30	**Participant Introductions**	2–3 min per person: background and current focus
10:30–10:45	*Coffee Break*	
10:45–12:15	**Participant Introductions**	Continued
12:15–14:00	*Lunch / Siesta*	
14:00–14:45	**Invited Talk I (TBD)**	30 min + 15 min discussion
14:45–15:15	**Short Talk (TBD)**	15 min + 15 min discussion
15:15–15:45	*Coffee & Cake*	
15:45–16:30	**Invited Talk II (TBD)**	30 min + 15 min discussion
16:30–17:00	**Short Talk (TBD)**	15 min + 15 min discussion
17:00–17:45	**Open Discussion**	Reactions, questions, and themes emerging from the day's talks
18:00	*Dinner*	

Tuesday, April 14 — Context & Perspectives II + Group Formation

09:00–09:45	**Invited Talk III (TBD)**	30 min + 15 min discussion
09:45–10:30	**Invited Talk IV (TBD)**	30 min + 15 min discussion
10:30–10:45	*Coffee Break*	
10:45–11:30	**Invited Talk V (TBD)**	30 min + 15 min discussion
11:30–12:15	**Challenge Collection**	Participants propose open problems inspired by the talks. All captured on shared board.
12:15–14:00	*Lunch / Siesta*	
14:00–15:00	**Challenge Prioritization**	Affinity mapping into clusters. Dot-voting on priorities.
15:00–15:30	*Coffee & Cake*	
15:30–16:15	**Working Group Formation**	Top 5–6 themes become groups. Self-selection, define leads and scribes.
16:15–18:00	**Working Group Kickoff**	Define scope, draft initial outline, assign tasks
18:00	*Dinner*	

Wednesday, April 15 — Build I & Social

09:00–10:30	**Breakout Sessions**	Develop core arguments and findings
10:30–10:45	*Coffee Break*	
10:45–12:15	**Breakout Sessions**	Cross-group visits encouraged
12:15–13:15	*Lunch*	
14:00–20:00	**Social Event**	Afternoon excursion organized by Schloss Dagstuhl

Thursday, April 16 — Build II, Present, & Plan

09:00–10:30	**Breakout Sessions**	Final group work — prepare presentations, refine conclusions
10:30–10:45	*Coffee Break*	
10:45–12:15	**Group Presentations I**	Groups 1–3: 15 min + 15 min plenary discussion each
12:15–14:00	*Lunch / Siesta*	

14:00–15:30	**Group Presentations II**	Groups 4–6: 15 min + 15 min plenary discussion each
15:30–16:00	*Coffee & Cake*	
16:00–17:00	**Panel: AI Agents and Society**	Podium discussion (4 panelists, preferably industry)
17:00–17:45	**Paper & Manifesto Planning**	IEEE S&P paper teams, milestones (May 1 deadline). Manifesto section assignments.
18:00	*Dinner*	

Friday, April 17 — Wrap-Up

09:00–10:00	**Future Directions Discussion**	Research priorities, collaboration plans, next steps
10:00–10:30	**Writing Milestones**	Finalize timelines for all deliverables
10:30–10:45	*Coffee Break*	
10:45–11:30	**Feedback & Survey**	Open feedback round and Dagstuhl survey
11:30–12:00	**Closing Remarks**	Summary, thanks, next steps
12:15–13:15	*Lunch & Departure*	