First International Workshop on AI and Serverless (WoAIS) 2026

Part of 20th ACM International Conference on Distributed and Event-based Systems (DEBS 2026), Lisbon, Portugal (June 23rd–26th 2026) (https://2026.debs.org/)

WoAIS will be hybrid with both virtual and on-location formats. Please note that while hybrid formats will be supported for workshops, the DEBS 2026 steering committee wants the main conference to be held in in-person only. Prospective attendees of the workshop should keep this in mind if they plan to attend both WoAIS and DEBS 2026.

The rise of new AI agents—which involve multiple LLM calls, dynamic plans, and a mix of code and AI models—is creating uniquely complex workloads for serverless platforms. This new reality has already resurfaced classic serverless headaches, such as cold starts, managing state, and figuring out resource allocation, but within the demanding context of modern AI applications. Furthermore, new complications are emerging, including mixed GPU/CPU needs, unpredictable execution plans, long-running yet highly bursty processes, and the critical need for robust agent-to-agent communication.

Looking beyond the cloud, the scope of serverless is expanding. We aim to look ahead at future architectures involving AI, hybrid clouds, and especially edge/IoT devices, which current serverless platforms are not well-equipped

to support. This naturally leads to a discussion on the role of LLMs in Serverless, where we will explore how hybrid serverless platforms can be leveraged for the entire lifecycle of Large Language Models (LLMs) and Foundation Models (FMs), from fine-tuning to serving.

This workshop brings together researchers and practitioners to explore aspects such as use cases, resource allocations, optimizations, and using AI to improve serverless experience and to discuss their experiences and ideas for future directions in serverless research.

As this year the workshop is hybrid and we are looking not only for research papers, experience papers, demonstrations, or position papers but also for live presentations of ongoing work, demonstrations, and anything else that may be interesting to the workshop audience.

The latest version of this CFP is available at http://serverlesscomputing.org/woais1/ (http://serverlesscomputing.org/woais1/)

Topics

This workshop solicits papers from both academia and industry on the state of practice and state of the art in AI and serverless computing. Topics of interest include but are not limited to:

- AI Operating System: research on system-level infrastructure for scalable intelligence, viewing world models as "operating systems" for future AI agents
- Infrastructure for Scaling AI: Research is sought for distributed training and evaluation systems, compilers, and data engines that support continual updates
- Infrastructure and network optimizations for AI and serverless applications
- Cloud Services for AI: scalable AI cloud services for tasks such as the inspection of large-scale infrastructures
- Operating AI at Scale: extending monitoring and adaptation to large-scale production environments with heterogeneous data streams and real-time costs
- Cost per token, energy usage, cost to operate AI agents and models, cost models, pricing models, and economics of AI and serverless

- Using serverless for AI pipelines
- Using serverless for inference serving of AI models and agents
- Autonomous Lifecycle Management for AI: topics such as transitioning AI to operations and managing the infrastructure challenges for AI adoption across its entire lifecycle
- Orchestrating multiple AI agents running as serverless functions, addressing issues such as how to manage the synchronization if they are working on sections of the same task
- Memory Layers for AI: implementing memory architectures for LLM-based agents, including episodic vs. semantic memory, retrieval mechanisms, and consolidation of interaction-driven experiences
- Cloud AI Security: AI-driven cybersecurity risk analysis and secure machine learning for the vulnerability assessment of AI and related technologies using serverless and other scalable techniques
- Elastic AI platforms and pay-as-you-go for GPUs with different cost metrics. Using AI assist and generative LLMs such as ChatGPT for building, running, and maintaining serverless-like applications.
- Supporting AI agents and with serverless approaches in agentic platforms
- Multi-cloud and hybrid cloud for AI and serverless computing in Edge, Fog, IoT, etc.
- Supporting Model Context Protocol (MCP) and using it in serverless AI applications
- Supporting customization and running user provided AI models any place: Cloud, Edge,Fog, IoT, etc.
- Evaluation, Benchmarking, and Reproducibility: early work and proposals for standardized benchmarks and
- evaluations, in particular about reproducible AI research results
- Mixed GPU, accelerators, and CPU workloads for AI agents
- Running AI and serverless applications with stochastic plans (unpredictable execution paths), bursty long-running processes, and inter-process communications
- Developer experience as we transition from “traditional” serverless and FaaS to AI agentic programming
- Developer productivity to build AI serverless code: from local source to observability and maintenance in production

- Serverless data management for AI, Retrieval-Augmented Generation (RAGs), vector and graph databases applied to serverless experience,
- AI and serverless for next-gen computing in Industry such as Platform Engineering and Internal Developer Platforms and other areas
- Low-code and no-code - new programming abstractions for AI and serverless
- Debugging AI serverless applications
- Use cases, experiences with AI and serverless
- DevOps for AI and serverless
- Confidential computing
- Sustainable computing
- Granular computing
- Super-lightweight containers Web Assembly
- Swarm intelligence
- Other topics related to AI and serverless computing

Important Dates

Paper Submission: **April 15, 2026** (AOE)
Notification of Acceptance: **April 28, 2026**
Final Camera-Ready Manuscript (Hard Deadline): **May 15, 2026**
Non-paper submissions (demos and other proposals): **June 1, 2026**
Author registration deadline: TBD
Workshop: June 23, 2026 (Tuesday)
Conference: June 23–26, 2026

Papers and Submissions

Papers submissions

Authors are invited to submit original, unpublished research/application papers that are not being considered in another forum.

Submitted manuscripts should be structured as technical papers and may not exceed six (6) single-spaced double-column pages using ACM proceeding style two-column “sigconf” which can be found on the ACM template page https://www.acm.org/publications/proceedings-template (https://www.acm.org/publications/proceedings-template), font size set to 10 pt. The page limit contains all the content, including bibliography, appendix, etc.

Shorter formats like extended abstracts (5 pages or less), posters, or opinions are also accepted.

For information about Auxiliary Materials, please check here. https://2026.debs.org/acm-auxiliary-materials/ (https://2026.debs.org/acm-auxiliary-materials/)

Note that submissions must be doubly anonymous - authors' names must not appear on the manuscript, and authors must make a good-faith attempt to anonymize their submissions. This means that submitted papers must be anonymous (not revealing author names). References to authors' previous work should be done in the third person to not reveal their identities. There should be no acknowledgments of people or projects. Supplementary material (e.g., GitHub or GitLab repository) should not reveal the authors' identities; to this end, anonymized repositories can be used (e.g., https://anonymous.4open.science (https://anonymous.4open.science)). See below for more details on the anonymity requirements for doubly-anonymous reviewing.

This year's edition will place ACM Artifacts Available badges on papers that make their artifacts available according to ACM's rules.

The DEBS conference organizers will provide companion proceedings including all workshop papers, which will be available in the ACM Digital Library. This is subject to the availability of their camera-ready papers by the deadline for camera ready papers.

Authors should submit the manuscript in PDF format. All manuscripts will be reviewed and will be judged on correctness, originality, technical strength, rigour in analysis, quality of results, quality of presentation, and interest and relevance to the conference attendees. Papers conforming to the above guidelines can be submitted to "Workshop on AI and Serverless Computing" track through the paper submission system powered by Microsoft CMT https://cmt3.research.microsoft.com/DEBS2026/ (https://cmt3.research.microsoft.com/DEBS2026/).

All submitted manuscripts (following DEBS conference requirements on formatting and page limits) will be peer-reviewed by at least 3 program committee members. Accepted papers with confirmed presentation will appear in the conference proceedings as well as in the ACM Digital Library.

The authors of accepted papers will be given a choice between different copyright agreements, according to the recent changes in the ACM policy. The options will include opportunities for open access as well as the traditional ACM copyright agreement.

Note that at least one author of each accepted workshop paper must hold a full pre-conference registration.

ACM Policies

By submitting your article to an ACM Publication, you are hereby acknowledging that you and your co-authors are subject to all ACM Publications Policies (https://www.acm.org/publications/policies), including ACM's new Publications Policy on Research Involving Human Participants and Subjects (https://www.acm.org/publications/policies/research-involving-human-participants-and-subjects). Alleged violations of this policy or any ACM Publications Policy will be investigated by ACM and may result in a full retraction of your paper, in addition to other potential penalties, as per ACM Publications Policy.

Please ensure that you and your co-authors obtain an ORCID ID (https://orcid.org/register), so you can complete the publishing process for your accepted paper. ACM has been involved in ORCID from the start and we have recently made a commitment to collect ORCID IDs from all of our published authors (https://authors.acm.org/author-resources/orcid-faqs). We are committed to improve author discoverability, ensure proper attribution and contribute to ongoing community efforts around name normalization; your ORCID ID will help in these efforts.

Important update on ACMs new open access publishing model for 2026 ACM Conferences!

Starting January 1, 2026, ACM will fully transition to Open Access. All ACM publications, including those from ACM-sponsored conferences, will be 100% Open Access. Authors will have two primary options for publishing

Open Access articles with ACM: the ACM Open institutional model or by paying Article Processing Charges (APCs). With over 2,600 institutions already part of ACM Open, the majority of ACM-sponsored conference papers will not require APCs from authors or conferences (currently, around 76%).

Authors from institutions not participating in ACM Open will need to pay an APC to publish their papers, unless they qualify for a financial waiver. To find out whether an APC applies to your article, please consult the list of participating institutions in ACM Open and review the https://www.acm.org/publications/policies/policy-on-discretionary-open-access-apc-waivers (https://www.acm.org/publications/policies/policy-on-discretionary-open-access-apc-waivers"). Keep in mind that waivers are rare and are granted based on specific criteria set by ACM.

Understanding that this change could present financial challenges, ACM has approved a temporary subsidy for 2026 to ease the transition and allow more time for institutions to join ACM Open. The subsidy will offer:

$250 APC for ACM/SIG members

$350 for non-members

This represents a 65% discount (https://www.acm.org/publications/openaccess), funded directly by ACM. Authors are encouraged to help advocate for their institutions to join ACM Open during this transition period.

Anonymity Requirements for Doubly-Anonymous Reviewing

Every research paper submitted to DEBS 2026 will undergo a "doubly-anonymous" reviewing process: in addition to maintaining the anonymity of the reviewers of the papers, the PC members and reviewers will not know the identity of the authors. To ensure the anonymity of authorship, authors must at least do the following:

1. Authors' names and affiliations must not appear on the title page or elsewhere in the paper.

2. Funding sources must not be acknowledged anywhere in the paper under review; these can be added to accepted papers upon submission of the camera-ready manuscript.
3. Non-anonymized links to the authors' online content must be removed.
4. Research group members, or other colleagues or collaborators, must not be acknowledged anywhere in the paper.
5. The paper's file name must not identify the authors of the paper.

Authors should also use care in referring to related past work. The solution is to reference past work in the third person (in the same way that one would reference work by anyone else). This allows you to set the context for your submission while at the same time preserving anonymity.

Despite the anonymity requirements, authors should still include all relevant work, including their own; omitting them could reveal the author's identity by negation. However, self-references should be limited to the essential ones, and extended versions of the submitted paper (e.g., technical reports or URLs for downloadable versions) must not be referenced. The goal is to preserve anonymity while allowing the reader to grasp the context of the submitted paper fully. It is the responsibility of authors to do their very best to preserve anonymity. Papers that do not follow the guidelines or potentially reveal the author's identity are subject to immediate rejection.

Other submissions

Authors are invited to submit proposals for demos and other presentations that are not papers.

Proposals must be submitted as short abstracts (not longer than one page) in PDF format using the paper submission system selecting "Other" as submission type.

Accepted presentations will not be part of the conference proceedings but will be part of the workshop agenda with dedicated time for live presentation (with video backup), questions etc.

Workshop co-chairs

Paul Castro, IBM Research
Pedro García López, University Rovira i Virgili
Vatche Ishakian, IBM Research
Vinod Muthusamy, IBM Research
Aleksander Slominski, IBM Research

Steering Committee

Geoffrey Fox, Indiana University
Dennis Gannon, Indiana University & Formerly Microsoft Research
Arno Jacobsen, MSRG (Middleware Systems Research Group)

Program Committee (tentative)

- Alexandru Iosup, Vrije Universiteit Amsterdam
- Ali Kanso, Microsoft
- Amine Barrak Assistant Professor, Oakland University
- Azer Bestavros, Boston University
- Cristina Abad, Escuela Superior Politécnica del Litoral (Ecuador)
- Dennis Gannon, Indiana University & Formerly Microsoft Research
- Eric Rozner, University of Colorado Boulder
- Etienne Rivière, UCLouvain
- Geoffrey Fox, Indiana University
- Gul Agha, University of Illinois at Urbana-Champaign
- Hans-Arno Jacobsen, MSRG (Middleware Systems Research Group)
- Ian Foster, University of Chicago and Argonne National Laboratory
- Josef Spillner, Zurich University of Applied Sciences
- Kyungyong Lee, Kookmin University
- Lucas Nussbaum, LORIA, France
- Maciej Malawski, AGH University of Science and Technology, Poland
- Maciej Pawlik, Academic Computer Centre CYFRONET of the University of Science and Technology in Cracow
- Marc Sánchez Artigas, Universitat Rovira i Virgili
- Per Persson, Ericsson Research
- Peter Pietzuch, Imperial College
- Rich Wolski, University of California, Santa Barbara
- Rodric Rabbah, Nimbella and Apache OpenWhisk

- Rodrigo Fonseca, Microsoft Research
- Samuel Kounev, University of Wuerzburg
- Tyler R. Caraza-Harter, University of Wisconsin-Madison
- Volker Hilt, Bell Labs (Nokia)
- Wes Lloyd, University of Washington Tacoma