ICLR 2026 Proposal for the 3rd Workshop on Test-Time Updates (TTU)

1. Workshop summary

The common paradigm of deep learning distinguishes the training stage, where model parameters are learnt on massive datasets, and deployment, during which the frozen models are tested on unseen data. In case the test-time data distribution changes, or the model needs to satisfy new requirements, a new training round is needed. **Test-time updates** (TTU), including test-time adaptation (TTA), post-training editing, in-context learning, and online continual learning, offer a complementary path to re-training: adapt *when and where* data shift occurs. Test-time updates are relevant across model size: they can be used to edit the knowledge in large foundation models for which re-training has prohibitive costs, as well as to adapt models on edge devices. Moreover, test-time adaption finds applications on a variety of tasks, from vision to natural language tasks or time series analysis, each presenting its specific challenges and methods. Finally, the goals of test-time approaches are multiple, spanning robustness, customization, and computational efficiency.

In this workshop we want to bring together these different facets of test-time updates, connecting researchers focusing on topics typically treated as independent problems. We believe that this will offer a unique opportunity for cross-area collaborations. Sharing domain-specific challenges and solutions will bridge diverse communities, providing beneficial contamination. In fact, we will welcome works on methods, theory, systems, and evaluations for TTU/TTA across modalities (vision, language, audio, etc.), scales (from edge to cloud), and openness (open/closed models, black-/white-box scenarios). We will highlight principled objectives, safe/robust updates, practical parameterizations (inputs, features, adapters, heads), and cost-aware/green practices that respect latency, energy, and monetary budgets.

Topics of Interest. We welcome work on test-time and post-training updates, including but not limited to

- **Foundations & Objectives:** Unsupervised/self-supervised losses at test time; implicit/explicit regularization; stability–plasticity trade-offs; theory of adaptation and generalization under shift.
- **Parameterizations & Interfaces:** Input-space updates (learnable augmentations, prompts), feature-space adapters (BN/affine, LoRA, adapters), head-level edits, retrieval-augmented updates, black-box query strategies for closed foundation models.
- **Shift, Attacks, & Tasks:** Coping with domain and style shift, distribution drift, adversarial perturbations, label shift, online continual learning and task switches, model availability attacks
- **Adaptation of Foundational Models (FM):** Adapting LLMs/VLMs and domain FMs to specialized/personalized settings via in-context learning, adapters/LoRA, TTU-RL, and model editing and unlearning.
- **Safety, Reliability, & Alignment:** Uncertainty, conformal prediction at test time, fallback/abstention, guardrails and risk monitors, privacy-preserving updates, auditability, and roll-back.

- **Dynamic Architectures:** Recurrent depth models, looped transformers, dynamically allocating compute (early-exit networks, mixture-of-depth), and iterative test-time optimization (deep equilibrium networks, implicit computation).
- **Metrics, Datasets, & Benchmarks:** End-to-end metrics that couple *utility* (accuracy, calibration) with *costs* (compute, memory, wall-clock, energy); realistic streams and recurrences; reproducible TTU pipelines.
- **Cost-Aware & Green TTU:** Methods and evaluations under compute/energy budgets, latency/throughput targets, edge constraints, carbon accounting, and cost–quality frontiers; *any improvement must justify its operational footprint.*

Questions. We hope our workshop will drive progress on the topics above, and furthermore answer these particular questions for the 3rd edition of the workshop that are needed now in the communities on these allied topics:

1. **Are test-time updates ready for deployment?** What supporting toolkit in model optimization, evaluation, versioning, and serving is needed for the reliable, efficient, and even profitable use of test-time updates?
2. **How can test-time updates balance adaptation and energy?** What factors limit test-time updates going from standard GPUs to more diverse edge devices and other platforms for applications and industrial adoption?
3. **What are the risks of test-time updates?** Does test-time adaptation make the models more or less vulnerable to (a) benign but corner-case distribution shifts and (b) attacks by malicious actors? Are dynamic architectures more or less prone to be hijacked for unsafe purposes?

2. Submissions

Main Track Submission. We will welcome submission of **short papers** (maximum 4 pages of content without counting references, optional Appendix with an unlimited number of pages) featuring works not previously published at peer-reviewed venues. Each accepted paper will be presented as a **poster and lightning talk** (= 1 slide in 1 minute). Moreover, we will select a number of outstanding works for **oral presentation**. The workshop will be without proceedings.

Tiny Papers Submission. We will welcome submission of **tiny papers** (1-2 pages of content without counting references, no Appendix) to enable the sharing of smaller projects, the promotion of new versions of methods/benchmarks/analyses, and the presentation of works-in-progress for feedback from the community. Each tiny paper will be reviewed by the same program committee but to a more accessible standard to help participation in these topics and community. Because explicit is better than implicit, we will provide guidance to authors and reviewers about this track to focus on *complete work of a smaller scope*, such as a few fair and correct experimental results, a short theoretical derivation or lemmas, or position papers and calls for discussions at the workshop (including our roundtable discussions during breaks and at the end of the workshop schedule).

Bonus Track Submission. Following the positive experience of last edition, we will again invite content for a "bonus track" of our workshop to better broadcast projects, progress, and announcements from and for the community. These will be limited to one-slide contributions, not accompanied by a paper hosted on the workshop webpage and OpenReview page, and undergo a lightweight review by the organizers. The bonus slides will be auto-played during the poster sessions and breaks and are shared online. The Bonus Track wants to allow last-minute contributions about on-going projects and development which are not suitable (yet) to be featured in a short paper. Moreover, we welcome the workshop attendees to use this track to communicate research ideas and insights which can sparkle discussion and collaborations.

Reviewing. We will use OpenReview, as we have done before, and we now have a pool of 100+ reviewers and the organizers are recruiting more. We will further open a public call for reviewers after the workshop acceptance. Reviewing will be double-blind. We will adopt the ICLR'26 Code of Conduct and Code of Ethics.

Timeline.

- Main Track + Tiny Papers Track Submission deadline: Feb 6, 2026
- Main Track + Tiny Papers Track Reviewing starts: Feb 6, 2026
- Main Track + Tiny Papers Track Reviewing ends: Feb 23, 2026
- Meta-Reviewing and Decisions: Feb. 24–27, 2026
- Decision to authors: Mar 1, 2026
- Bonus Track Submission deadline: ~1 week before the workshop (note: no paper submission and reviewing needed, but submitted bonus track slides will be accepted/rejected by the organizers)

Awards. We will select a best paper with a prize given during the event.

Conflicts of Interest. As in past editions of the workshop, all organizers will serve as meta-reviewers, and three organizers across separate institutions will serve as program chairs (Evan Shelhamer, Francesco Croce, Shuiacheng Niu) for making decisions, so that any one organizer can and will abstain from conflicting roles on submissions, sponsorship, and workshop participation. No organizer will be involved in the review process for a workshop contribution with which they are conflicted according to the NeurIPS rules as administered by the workshop program chairs through OpenReview. No organizer will give an invited talk at the workshop, and the in-person organizers will simply host and moderate.

3. Tentative Schedule

Our workshop collects different sessions

- **Lightning rounds:** each accepted paper is presented as a 1 minute lightning talk. The number and duration of lightning rounds will be adjusted depending on the exact number of accepted papers. We have included this type of session since the 1st edition and found it a powerful tool for platforming more diverse and early work because <u>every</u>

accepted paper has a lightning talk. By highlighting these before the poster sessions, the authors have a chance to attract attention to their posters.

- **Oral talks:** papers selected for oral presentation get a 12+3 minutes slot for talk and Q&A with in-person questions and virtual sharing of the materials for online and later conversations following the workshop.
- **Poster sessions:** each accepted paper can be presented in both morning and afternoon poster sessions for more opportunities to share and discuss.
- **Invited talks:** 25 minutes talks by our invited speakers, plus 5 minutes of QA
- **Roundtable discussions:** moments for open discussion and networking among invited speakers, authors and attendees, in which a topic is led by a speaker (+ organizer as moderator) in different parts of the space, such as a cluster of chairs.

9:00	9:15	Opening remarks
9:15	9:30	🎤 Invited Talk 1
9:30	9:45	⚡ Lightning 1
9:45	10:15	🎤 Invited Talk 2
10:15	10:30	⚡ Lightning 2
10:30	10:45	🎙️ Oral Talk 1
10:45	11:00	🎙️ Oral Talk 2
11:00	12:15	🖼️ Poster Session 1
12:15	12:45	☕ Lunch Break
12:45	13:15	🎤 Invited Talk 3
13:15	13:30	⚡ Lightning 3
13:30	14:00	🎤 Invited Talk 4
14:00	14:30	🎤 Invited Talk 5
14:30	14:45	☕ Break
14:45	15:00	⚡ Lightning 4
15:00	15:15	🎙️ Oral Talk 3
15:15	15:30	🏆 Oral Talk 4 - Best Paper
15:30	16:45	🖼️ Poster Session 2
16:45	17:15	💬 Roundtable Discussions
17:15	17:30	Closing remarks

4. Invited Speakers

Speakers. Our workshop will feature five invited talks, delivered from speakers with diverse backgrounds and expertise. All invited speakers have confirmed their intention of attending the workshop in case of acceptance.

- Jacob Andreas (MIT) is an Associate Professor at MIT EECS and CSAIL. His research focuses on the computational foundations of language learning and communication. His work aims to develop general-purpose intelligent systems that can understand, communicate, and learn effectively from human language and guidance. He has recently worked on how test-time training and updates can improve few-shot reasoning and other usage by large language modeling.

- Katerina Fragkiadaki (CMU) is a JPMorgan Chase Associate Professor of Computer Science in the Machine Learning Department at Carnegie Mellon University. She works in artificial intelligence at the intersection of computer vision, machine learning, language understanding, and robotics. Her most recent works also focus on test-time adaptation for reasoning and chain-of-thought.

- Jonas Geiping (ELLIS/MPI Tuebingen) is the Research Group Leader of ELLIS Institute & Max-Planck Institute for Intelligent Systems, Tübingen AI Center, Germany. His research focuses on the safety, security, and privacy of machine learning systems, as well as understanding and implementing reasoning in intelligent systems and developing efficient methods for language modeling.

- Guillaume Vray (EPFL) is a final-year PhD student at EPFL. He works on test-time adaptation with different modalities and applications across various domains, such as medical imaging. His research has been published at top-tier conferences like ICLR and CVPR.

- Eleni Trintafilou (GDM) is a Senior Research Scientist at Google DeepMind. Her main research interest is around creating methods that allow efficient and effective adaptation of deep neural networks to cope with distribution shifts, rapidly learning new tasks, and supporting efficient unlearning of data points. Her research falls in the areas of few-shot learning, meta-learning, domain adaptation and machine unlearning.

Roundtable Discussions. All speakers will be also invited to participate in the roundtable discussion together with the authors of accepted papers and attendees. The roundtable discussion will be done in parallel as "breakout sessions" in which attendees choose a speaker and topic and join them in a small group for open discussion. By declaring these topics at the beginning of the workshop, attendees will have a chance to think and talk about these during the lunch and at the end of the workshop in our roundtable session.

5. Organizers and biographies

Our organizing team of six members demonstrates strong diversity in gender (four male and two female), seniority, nationality, affiliation, and research perspective. To ensure both continuity and renewal, three organizers have led or participated in previous editions of our workshop, while three new members with strong expertise in test-time updates bring fresh insights. Collectively, the team has organized six workshops across vision and machine learning from 2017 to 2024 and published over twenty papers on test-time adaptation across diverse tasks, settings, and models. This balanced composition ensures a wide range of viewpoints, inclusive collaboration, and strong alignment with the workshop's theme.

- Evan Shelhamer (lead organizer) is an assistant professor at UBC in Vancouver, member of the Vector Institute, and CIFAR AI Chair. His research is on visual recognition, adaptation, and adaptive computation to update models on new and different data. He received his PhD from UC Berkeley in 2019 advised by Prof. Trevor Darrell. He was the lead developer of the Caffe deep learning framework from version 0.1 to 1.0. His research and service have received awards incl. the best paper honorable mention at CVPR'15, Mark Everingham award at ICCV'17, and the test-of-time awards at MM'24 and CVPR'25. He was the lead organizer for the 1st workshop on TTA at CVPR'24 and 2nd workshop on TTA at ICML'25 and he co-organized the 3rd workshop on Machine Learning for Remote Sensing at ICLR'25 and the 6th workshop on Continual Learning at ICCV'25. He is co-organizing another workshop at ICLR 2026.
  Email: shelhamer@cs.ubc.ca

- Francesco Croce is an assistant professor at Aalto University. His research mainly focuses on multimodal foundation models, in particular their adversarial robustness (jailbreaks, backdoors, etc.) for safe AI systems, and the ability of multimodal models to capture different aspects of human perception, e.g., visual and semantic similarity. He was co-organizer for the robustness workshops "A Blessing in Disguise: The Prospects and Perils of Adversarial Machine Learning" (ICML 21) and "The Art of Robustness: Devil and Angel in Adversarial Machine Learning" (CVPR 22), as well as for the 1st and 2nd Workshops on "Test-Time Adaptation" (CVPR 24 and ICML 25). He is proposing another workshop for ICLR 2026.
  Email: francesco.croce@aalto.fi

- Teresa Yeo is a Postdoctoral Researcher at the Singapore-MIT Alliance for Research and Technology, where she develops adaptive systems using neuro-symbolic methods across domains ranging from mathematics to embodied AI. Her research generally focuses on steering multi-modal foundation models to improve their robustness and adaptability under distributional and task shifts. She was also a co-organizer of the 2nd Test-time Adaptation Workshop at ICML 2025.
  Email: teresa.yeo@smart.mit.edu

- Shuaicheng Niu is a research fellow at Nanyang Technological University (NTU), Singapore. His research primarily focuses on out-of-distribution generalization and test-time learning for discriminative or generative models. He aims for general and efficient adaptation tools to enhance the real-world applicability of machine learning systems, by developing stable, on-device, forward-only, black-box adaptation methods, etc. His recent

projects on test-time adaptation have been recognized as Oral presentations at ICLR and ICML.
Email: niushuaicheng@gmail.com

- Behzad Bozorgtabar is a lecturer and head of the Computer Vision Team at École Polytechnique Fédérale de Lausanne (EPFL), the Signal Processing Lab (LTS5). He is a member of ELLIS and EPFL's ELLIS Unit. Formerly a postdoctoral researcher at IBM Research–Australia, his research focuses on test-time adaptation and multimodal foundation models, now extending to agentic AI for robust, calibrated, and budget-aware deployment under shift.
  Email: behzad.bozorgtabar@epfl.ch

- Xiaoxiao Li is currently an assistant professor in the Department of Electrical and Computer Engineering at the University of British Columbia, a faculty member at Vector Institute, and a visiting research scholar at Google. Dr. Li is recognized as a Canada Research Chair (Tier II) in responsible AI and a CIFAR AI Chair. Dr. Li's research interests primarily lie at the intersection of AI and healthcare, theory and techniques for artificial general intelligence (AGI), and AI trustworthiness. Dr. Li aims to develop the next-generation responsible AI algorithms and systems.
  Email: xiaoxiao.li@ece.ubc.ca

6. Anticipated audience size

We expect to receive around 100 submissions, which is a larger number than what we had for the 2nd edition of the workshop (72), in line with growth from the 1st edition. Therefore, we expect around 100-150 attendees at the in-person event based on attendance at the 2nd edition at ICML'25.

7. Plan to get an audience for a workshop (advertising, reaching out, etc.)

Following past experience, we will create a workshop website, and advertise the event via social media (e.g., dedicated X and Bluesky accounts) as well as organizers' personal networks including the CIFAR institutes of Canada, the ELLIS network of Europe, and universities and companies around the world including Asia and Brazil. Thanks to the popularity of the first two editions, the workshop has already established a solid community which will likely engage in the third edition.

8. Diversity commitment

We respect and encourage diversity in our workshop's content and attendance.

Technical diversity. The main goal of our workshop is to integrate different threads of research on test-time updates and their respective communities on technically-intersecting but socially and collaboratively distinct topics. We diversify across modalities (e.g. images/text, language/videos/tabular data, etc.), architectures (CNNs, transformers, shallow architectures) and applications, as exemplified by the diversity of the expertise of the invited speakers.

Speakers. From a topic perspective, we have drawn on and connected different sub-areas of machine learning, computer vision, and natural language processing (meaning different tasks, measures of robustness, and different kinds of data) to investigate adaptation post-training and during testing. Moreover, the invited speakers show diversity in gender (two female and three male), seniority (from senior PhD student to established researchers), nationality, affiliation (both academia and industry are represented), and research perspective.

Organizers. Our organizing team of six members demonstrates strong diversity in gender (four male and two female), seniority, nationality, affiliation, and research perspective. To ensure both continuity and renewal, three organizers have led or participated in previous editions of our workshop, while three new members with strong expertise in test-time updates bring fresh insights.

Participation diversity. As a topic, test-time and post-training model adaptation admits broader participation by its emphasis on deployment and computational efficiency. This makes it more affordable for larger and smaller groups.

9. Virtual access to workshop materials and outcome

Our workshop will be in-person, but will provide virtual content by streaming talks and hosting materials on our site for broader access. We will not hold a virtual poster session, but will instead highlight virtual contributions at the event and on the site. Virtual contributions will be added to our bonus track shown during breaks and lunch.

10. Previous related workshops

This will be the third edition of a series of workshops about test-time adaptation.

- The 1st Workshop on Test-Time Adaptation at CVPR'24 covered test-time updates at the top venue for computer vision, but did not cover broader modalities and post-training updates
- The 2nd Workshop on Test-Time Adaptation at ICML'25 expanded the scope of the workshop to more ML domains, modalities, and tasks, e.g. including works on large language models and time series.

The 3rd installment of the series will continue on this path, focusing on modern applications of test-time adaptation such as test-time RL and machine unlearning which are rapidly becoming the object of mainstream research.

Other related workshops include:

- Continual and Compatible Foundation Model Updates (NeurIPS'25) covers cost-effective methods for frequent updates and adaptation to keep foundation models up-to-date, but does not focus on test-time adaptation, updating during deployment, or dynamic architectures and smaller scales of modeling that may be more accessible.
- Adaptive Foundation Models (NeurIPS'24) covered adaptation and personalization, but only includes foundation models instead of scales from small to large, and lacks the focus on test-time and unsupervised learning.
- Continual Learning for Lifelong Foundation Models (NeurIPS'24) covered large-scale and supervised learning while we cover all scales and unsupervised learning.
- DistShift (NeurIPS'23) covered robustness and shift but not test-time updates.
- Shift Happens (ICML'22) covered shifts and benchmarking but not test-time updates.