



CONFERENCE

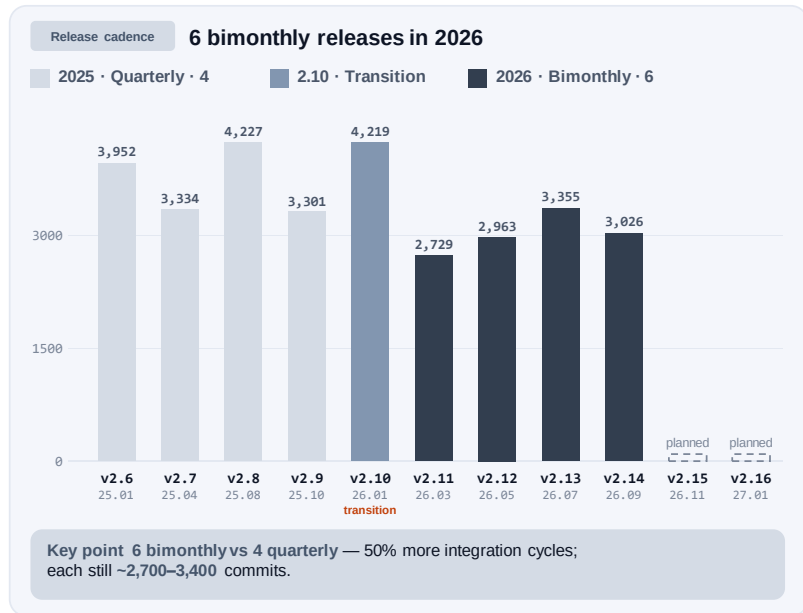
—NORTH AMERICA 2026—

Agile Readiness: Leveraging CI Relay for PyTorch Workflows

Zesheng Zong, Jiawei Li (AI Framework Engineer, Huawei)

Can New Accelerators Keep UP?

PyTorch is accelerating from **quarterly** to **bimonthly** releases — community CI / tests only cover **in-tree** devices.



Today The new accelerator's blind spot

- **In-tree:** full community CI + test coverage.
- **New (out-of-tree):** no coverage — invisible to upstream changes.

Challenges Four hard spots

1 Compressed window

Quarterly ≈3 mo → bimonthly 2 mo

2 Denser change impact

~2,700–3,400 commits / release

3 Quality-signal gap

Community CI doesn't cover new devices

4 Non-reusable tests

Hardware-coupled, re-port/rebuild every release

Next up Two solutions (RELEASE CADENCE × RELEASE QUALITY)

CRCR Cross-Repo CI Relay

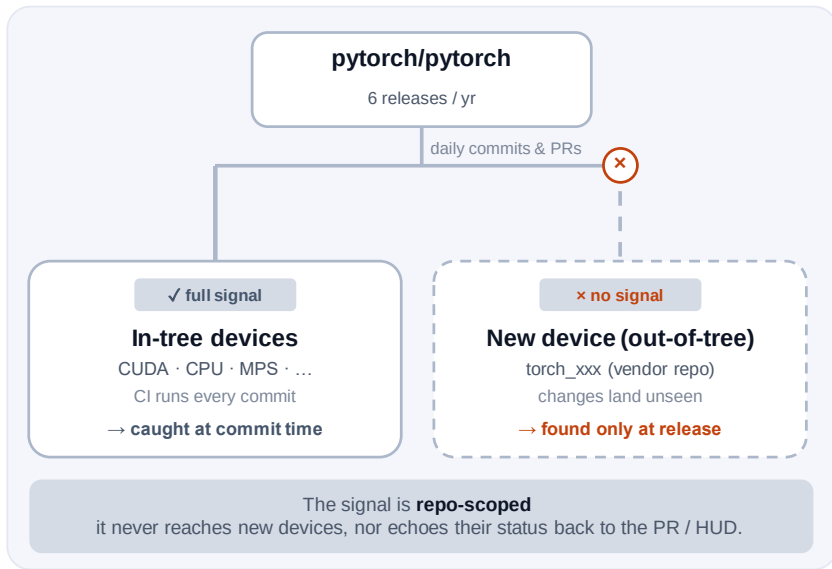
Relays upstream PR events to the new accelerator's repo — incompatible changes caught/adapted **in the daily flow**.

Tests Hardware-agnostic

Decouple community tests from hardware — new devices **reuse them directly**.

New Accelerators Can't See Upstream Changes

Upstream changes constantly — but the signal stops at the repo boundary.
Community CI covers only **in-tree** accelerators; a new accelerators in its own repo never sees it.



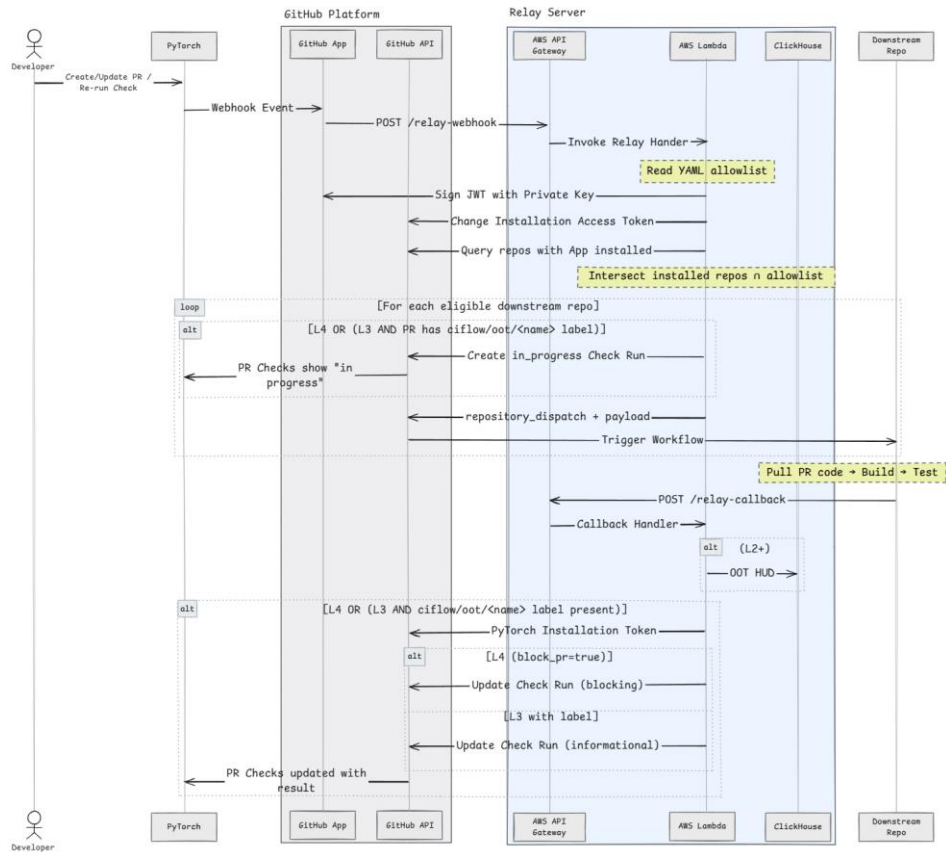
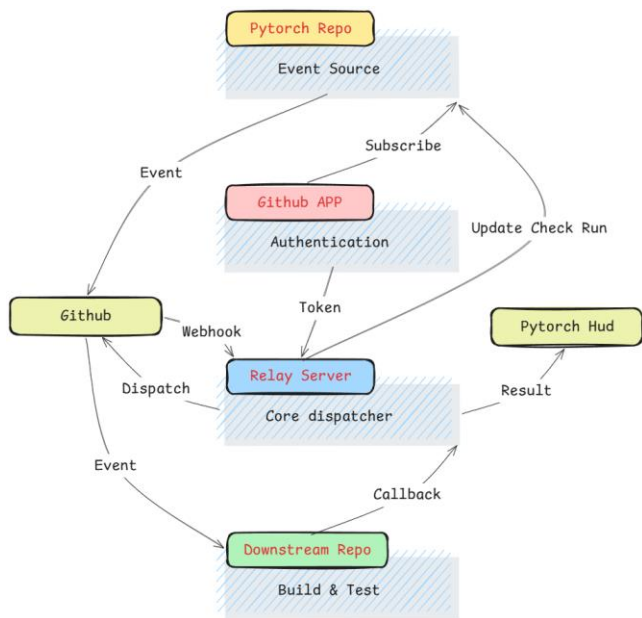
Without CRCR 4 gaps

01 New-feature detection New ops & APIs land daily — found only by manual diff each release.	02 Compatibility breakage Silent refactorors break the backend with no warning — found only at sync.
03 Batch, not stream No daily feed → work piles into a release-time report .	04 No PR / HUD echo A PR's impact on the new device never shows on the PR / PyTorch HUD.

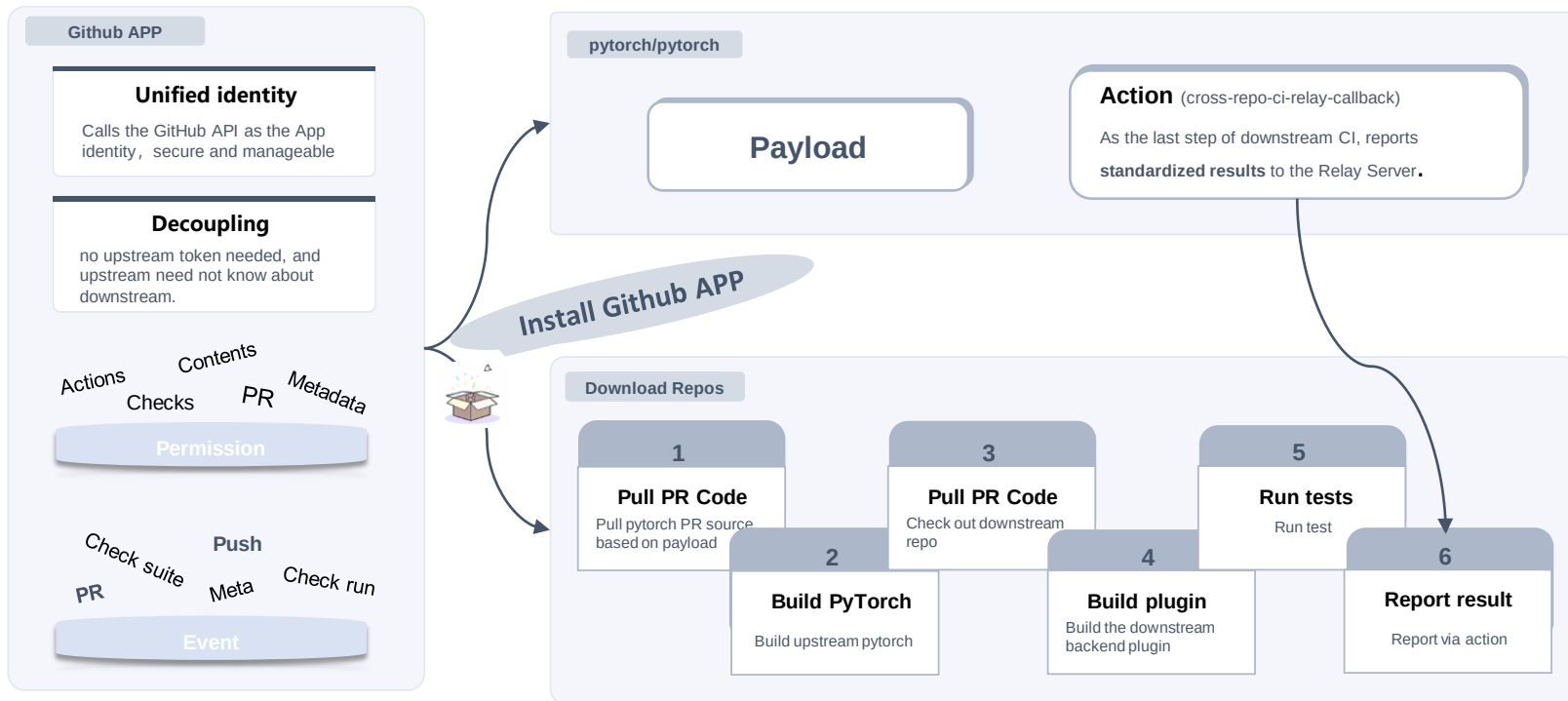
CRCR

- Relays upstream PR events across the repo boundary.
- Echoes the accelerator's results back to the PR / PyTorch HUD.

Architecture Design



GitHub APP & Repos



Relay Server

Core dispatcher

PR events from PyTorch

Webhook push -> trigger downstream CI

Callback from downstream

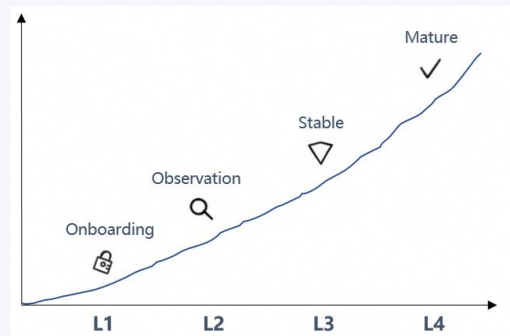
HTTP callback -> relay results back to the upstream PR or pytorch hud.

Allowlist Policy

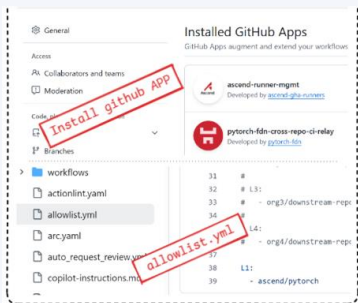
The allowlist is the core control mechanism of the relay server, control each downstream **repo's participation depth**.

	forward_event	display_on_hud	display_on_pr	block_pr
☐ L1	☒	☐	☐	☐
☐ L2	☒	☒	☐	☐
☐ L3	☒	☒	Label only	☐
☐ L4	☒	☒	☒	☒

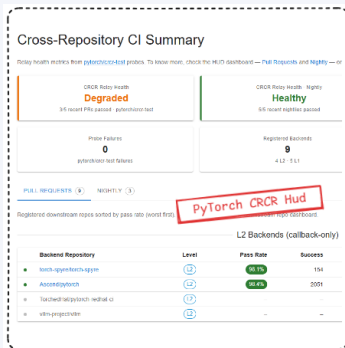
Evolution Path



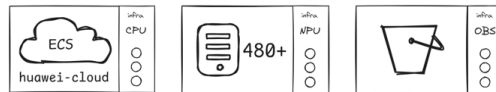
CRCR Journey



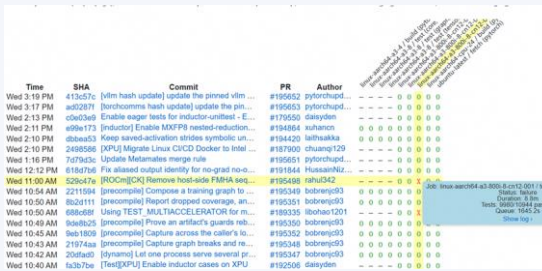
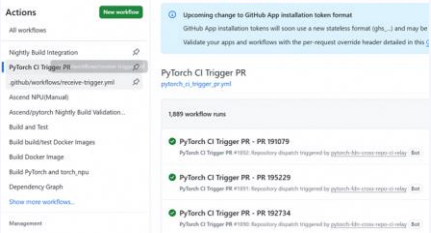
2026.4.21(L1)



2026.6.26(L2)



2026.9.x(L3)



Next Steps

CRCR

- Allowlist L4 entry standards
- Enrich Hud page for users and maintainers

Join us



PyTorch Slack Channel

[#tac-accelerator-integration-wg](#)



pytorch-fdn/accelerator-integration-wg

PyTorch Accelerator Integration

This repository is dedicated to improving and streamlining the integration of diverse AI hardware accelerators with PyTorch. It supports the efforts of the Accelerator Integration Working Group to reduce integration complexity, enhance the integration mechanism, and establish reliable CI infrastructure for out-of-tree accelerators. The goal is to build a scalable and inclusive PyTorch hardware ecosystem.

Working Group Members

All names below are arranged in alphabetical order; sequence carries no priority or seniority.

Co-Chairs

- Guangye Yu ([@guangyey](#))
- Zesheng Zong ([@zeshengzong](#))

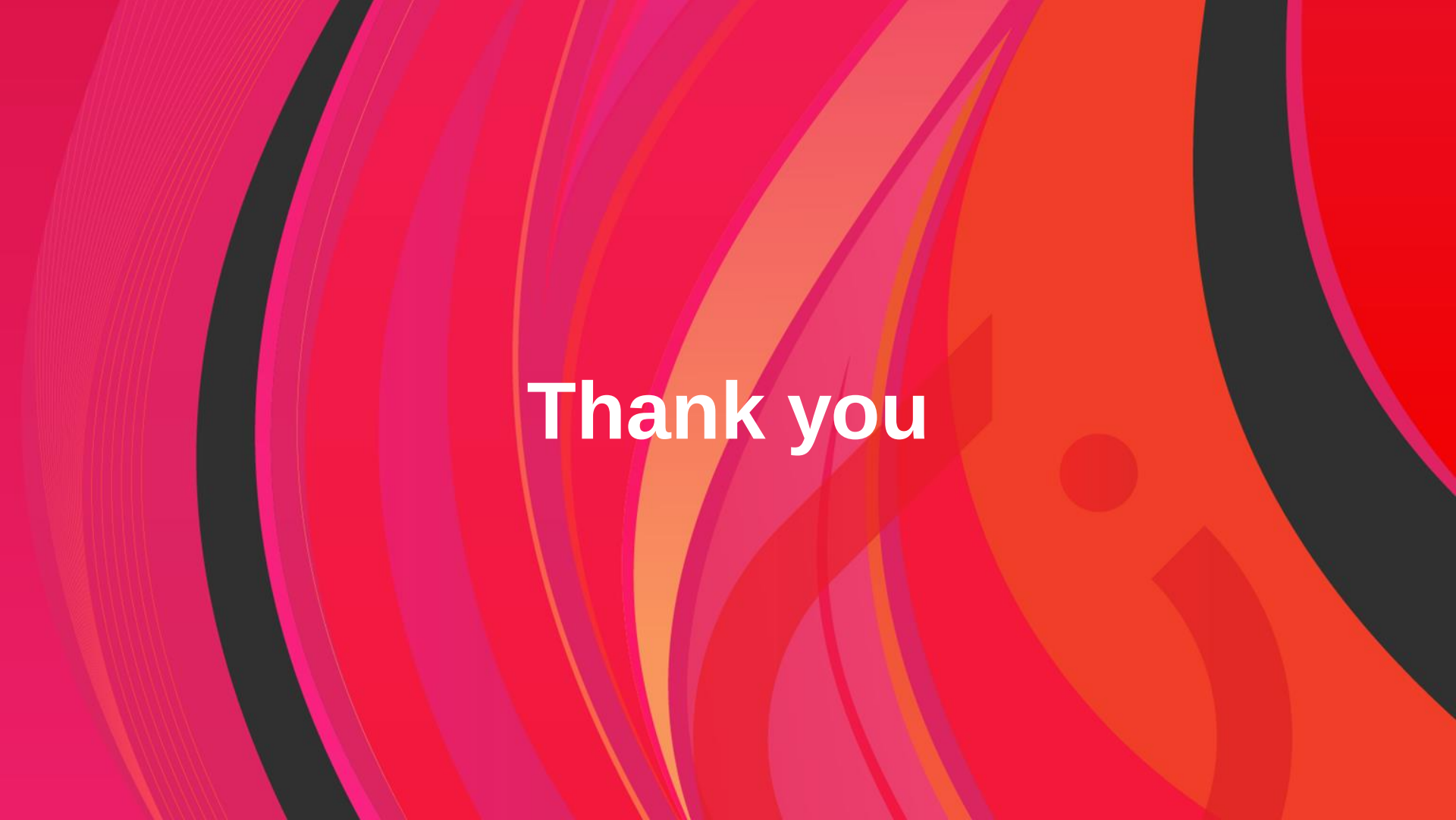
Voting Members

- Alban Desmaison ([@alband](#))
- Claudio Basile ([@cbsile-g](#))
- Guangye Yu ([@guangyey](#))
- Zesheng Zong ([@zeshengzong](#))

Accelerator Integration Guide

New hardware accelerators can integrate with PyTorch using the shared dispatch key [PrivateUse1](#). We're currently working on a comprehensive guide to help developers connect their custom hardware to PyTorch. This guide will include step-by-step instructions and practical examples based on the [OpenReg](#) reference implementation. For more details, refer to the related [RFC](#).

This effort is actively in progress. You can follow related discussions and open issues labeled [OpenReg](#). Community contributions are highly encouraged and welcome.



Thank you