Developing Data Science Skills Through Collaborative Hackathons: A Case Study of Hack4dev's Data Science Hackathon Programme

Dr Charles Takalana
charles@astro4dev.org

science, technology & innovation
Department:
Science, Technology and Innovation
REPUBLIC OF SOUTH AFRICA

NRF National Research Foundation | SAAO South African Astronomical Observatory

IAU

OAD
INTERNATIONAL ASTRONOMICAL UNION | OFFICE OF ASTRONOMY FOR DEVELOPMENT

BRICS ASTRONOMY

AfAS African Astronomical Society

IDIA

DARA DEVELOPMENT IN AFRICA WITH RADIO ASTRONOMY

What is Hack4dev?

Hackathons are intensive collaborative events where participants work in teams over a **limited period to develop innovative, data-driven solutions to real-world challenges** while building technical skills and professional networks.

Vision:

Hackathons for a more resilient world.

Objectives:

1. **Skills Development:** Building data science, problem-solving, and technical capacity across communities.
2. **Innovation:** Encouraging creative, data-driven solutions that address local and global development challenges.

What is Hack4dev?

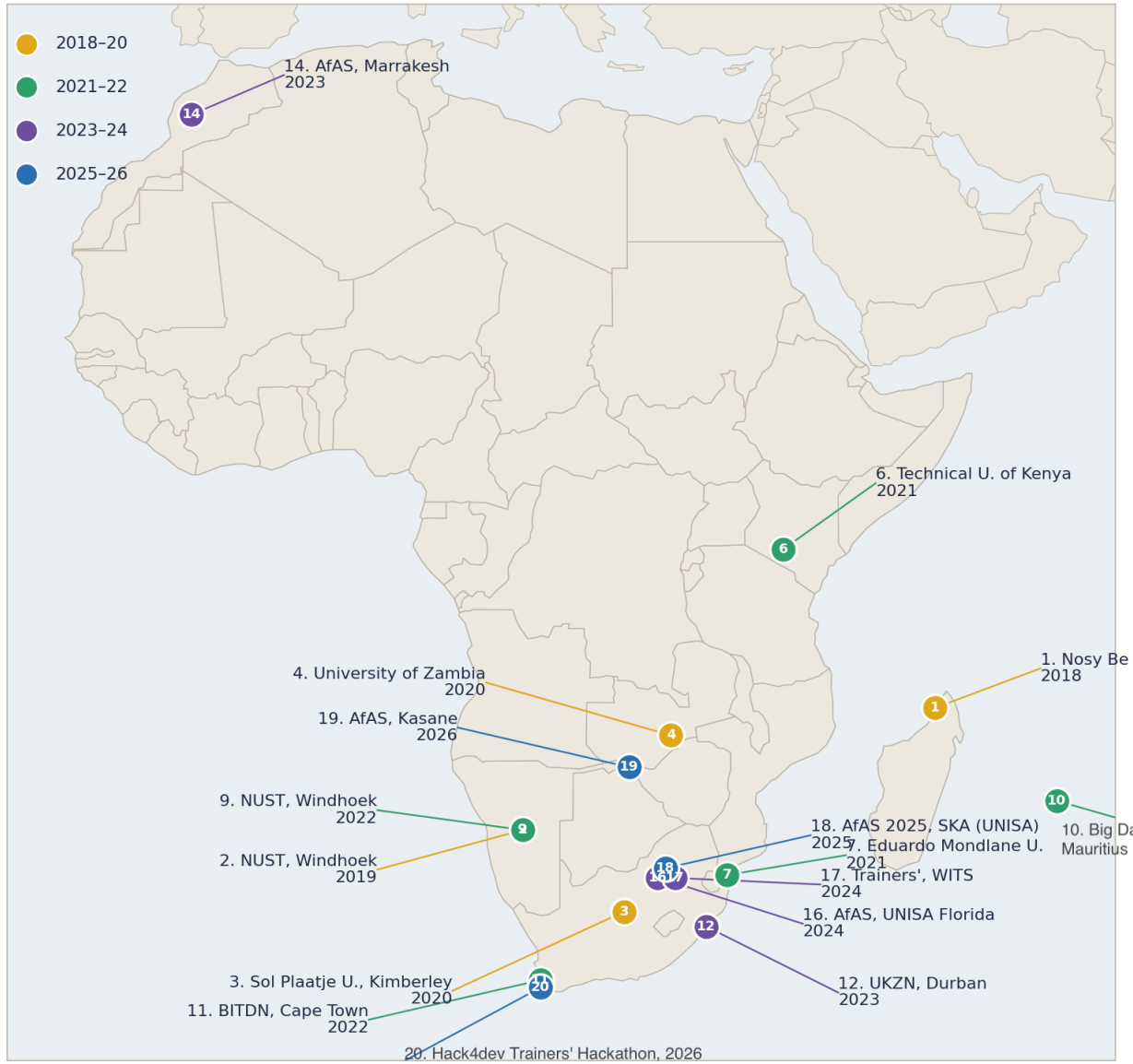
- Founded and launched in May 2022 by Prof. Carolina Ödman (IDIA-UWC), Hack4dev has grown into a global initiative that connects research and development through hackathons.
- It is a shared backbone for hackathons – offering models, guidance, and support, and welcoming tools, resources, and infrastructure from around the world.
- Founding partners: IDIA, OAD, BRICS (BITDN), DARA, and AfAS.

Our Hackathon Structure

1. Presentations - from industry and academia
2. Ice-breaker
3. Detailed tutorials in the form of Jupyter notebooks
4. Tutors trained on the hackathon challenge
5. Team presentations
6. Judging panel
7. Feedback survey
8. Prize giving
9. Social event (optional)

Hack4dev - Data Science Hackathons

2018–2026 | physical host sites, virtual events, and 2024/25 regional hackathons



Pins show host city/institution where supplied. Virtual events and events without a supplied location are listed in the key only.

Key

- Numbered physical host site
- 2024/25 regional hackathon country

Past Hackathons

1. Nosy Be, Madagascar — 2018
2. NUST, Windhoek, Namibia — 2019
3. Sol Plaatje University, Kimberley — 2020
4. University of Zambia, Lusaka — 2020
5. SGAC — virtual, 2020
6. Technical University of Kenya, Nairobi…
7. Eduardo Mondlane University, Maputo —…
8. Africa Women in Data Science — virtual,…
9. NUST, Windhoek, Namibia — 2022
10. Big Data Mauritius — 2022
11. BITDN, Cape Town — 2022
12. UKZN, Durban — 2023
13. Andean-OAD — virtual Andean region, 2023
14. AfAS Data Science Hackathon, Marrakesh…
15. BAWG Hackathon — 2023*
16. AfAS Hackathon, UNISA Florida — 2024
17. Hack4dev Trainers' Hackathon, WITS —…
18. AfAS Hackathon 2025, SKA Data Challenge…
19. AfAS Hackathon, Kasane, Botswana — 2026
20. Hack4dev Trainers' Hackathon, SAAO Cape…

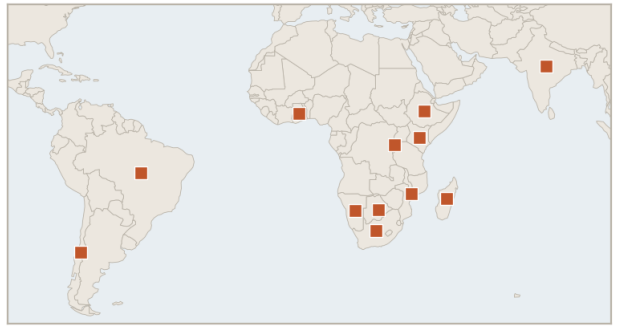
Virtual

5. Space Generation Advisory Council — virtual, 2020
8. Africa Women in Data Science — virtual, 2022
13. Andean-OAD — virtual Andean region, 2023
15. BAWG Hackathon — 2023, location not specified

2024/25 regional countries

Kenya, Ghana, Ethiopia, Botswana, South Africa, Madagascar, Rwanda, Namibia, Mozambique, India, Brazil and Chile.

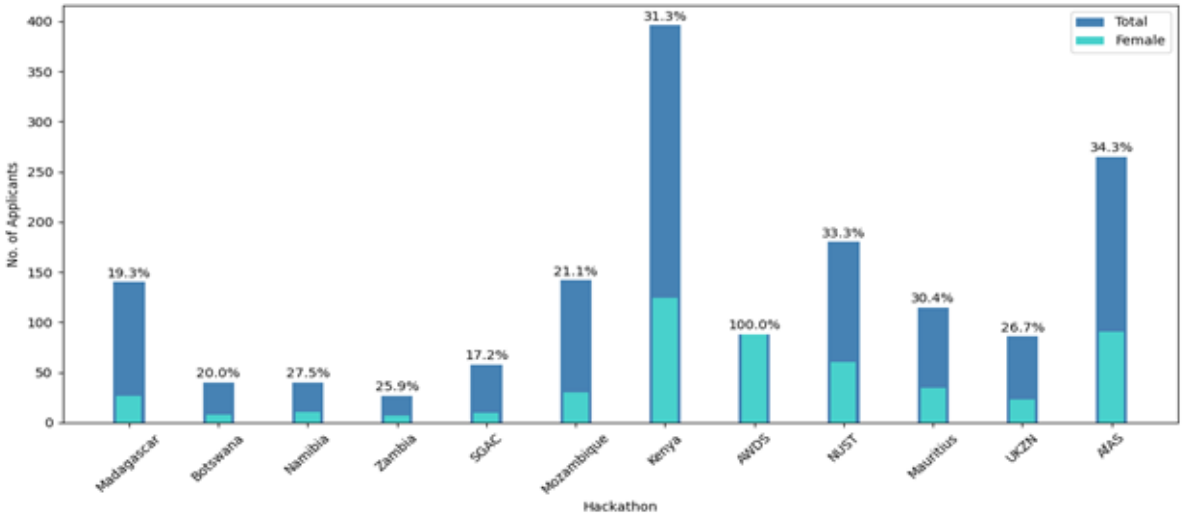
2024/25 regional hackathons — 13 events, 12 countries

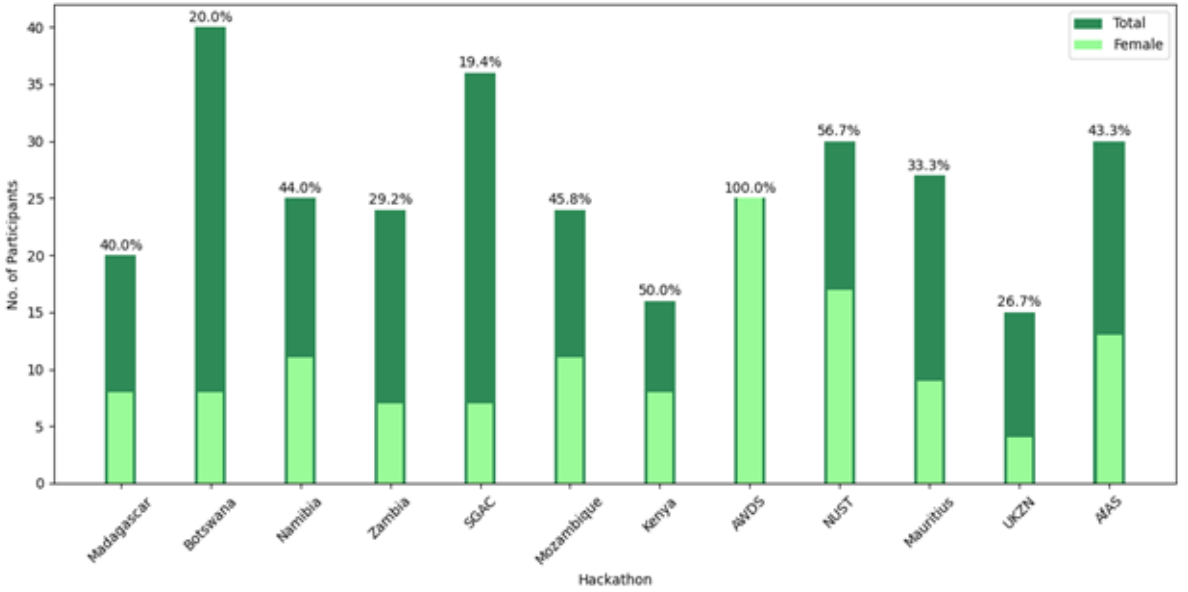


Participants

- **About 2000 applications** received across the hackathons, with **over 350 participants** selected and trained, demonstrating sustained demand for data science capacity-building.
- **Overall female participation reached 41% from just below 20%,** reflecting a strong emphasis on improving gender diversity across the programme.
- **Highly competitive selection**, with only ~20% of applicants ultimately participating, ensuring high-quality cohorts while highlighting the need for additional training opportunities.
- **Steady growth in scale and reach**, expanding from individual hackathons to a coordinated, multi-country programme spanning Africa and beyond.

Plots: 2018 - 2023





The Challenge

The gap:

- Uneven access to hands-on training
- Limited local mentorship
- Few applied learning opportunities

- AI, machine learning, and data science are now vital across research, industry, and development.

- But access to practical training, mentorship, and applied learning remains uneven – especially in emerging innovation ecosystems.

- The Data Science Hackathon Programme (DSHP) was built to **close this gap through a scalable, decentralised approach.**

The Data Science Hackathon Programme

Designed for:

- Scale – reaching many regions at once
- Local ownership – organisers lead
- Quality – a shared challenge framework

- A scalable, decentralised approach to technical skills development – expanding access to high-quality data science training.

- The design fosters local ownership and innovation, so training takes root in each community rather than being delivered from the outside.

- Use and develop data science skills to tackle global challenges.

A Two-Stage Cascade Model

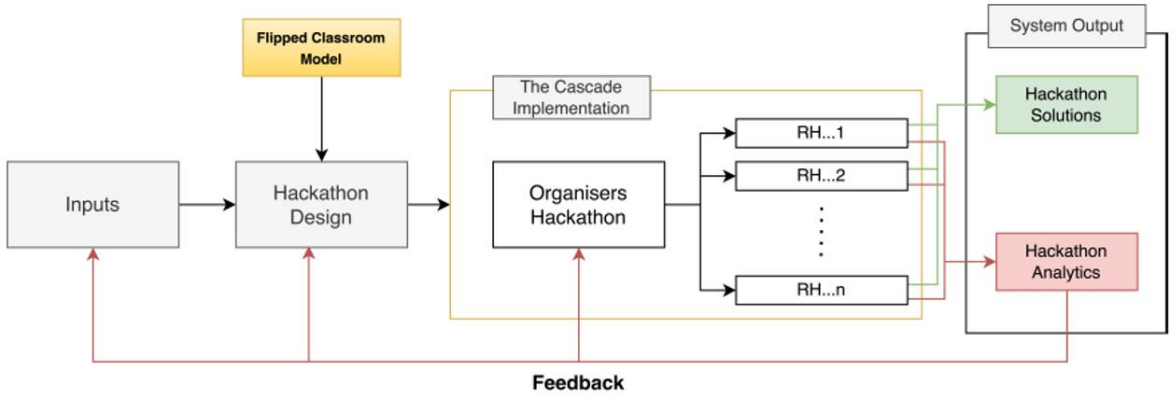
The programme expands access by training organisers, who then run their own hackathons – multiplying reach while keeping ownership local.

Stage 1 – Trainers Hackathon: selected organisers learn challenge facilitation and community engagement.

Stage 2 – Regional Hackathons: organisers run events in their own institutions and communities.

Local ownership: locally driven learning environments, tailored to each context.

Shared framework: a common global challenge connects every regional event.



First Cycle: The Results

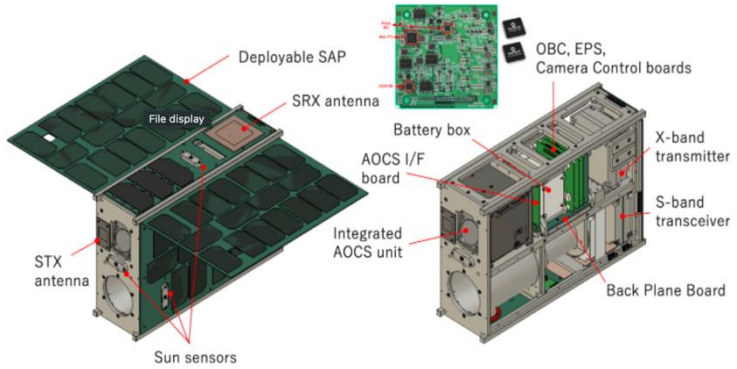
First cycle:

- 50+ organisers trained
- 150+ participants reached
- 48 solutions submitted

- The first cycle shows the cascade model works: a small group of trainers reached many participants across multiple regions.

- Short, challenge-driven hackathons produced real, benchmarked machine learning solutions in just three days.

- Evidence that decentralised training can scale without losing quality.



50+ organisers trained

150+ participants reached

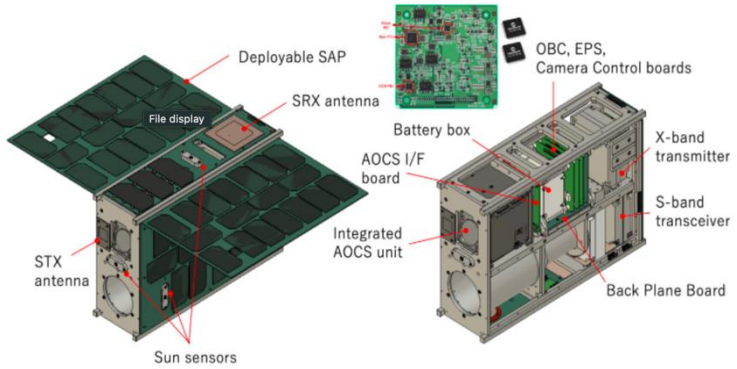
48 ML & DS solutions in 3 days

20 beat the benchmark

165× faster than baseline

Performance Highlights

- Across a three-day window, participants submitted 48 machine learning and data science solutions.
- Twenty submissions outperformed the benchmark model provided by the organisers.
- The highest-performing solution ran 165× faster than the baseline.



50+ organisers trained

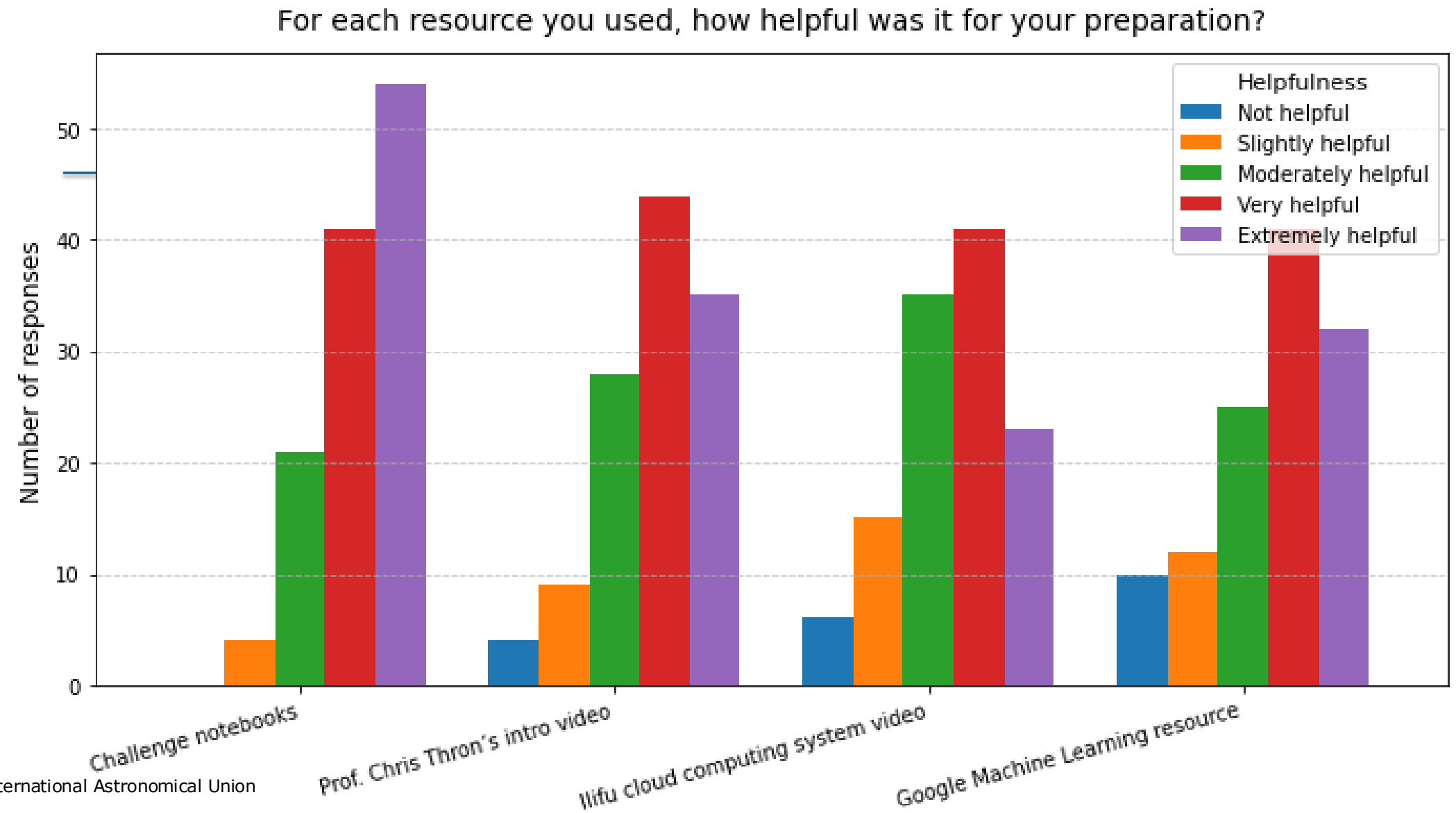
150+ participants reached

48 ML & DS solutions in 3 days

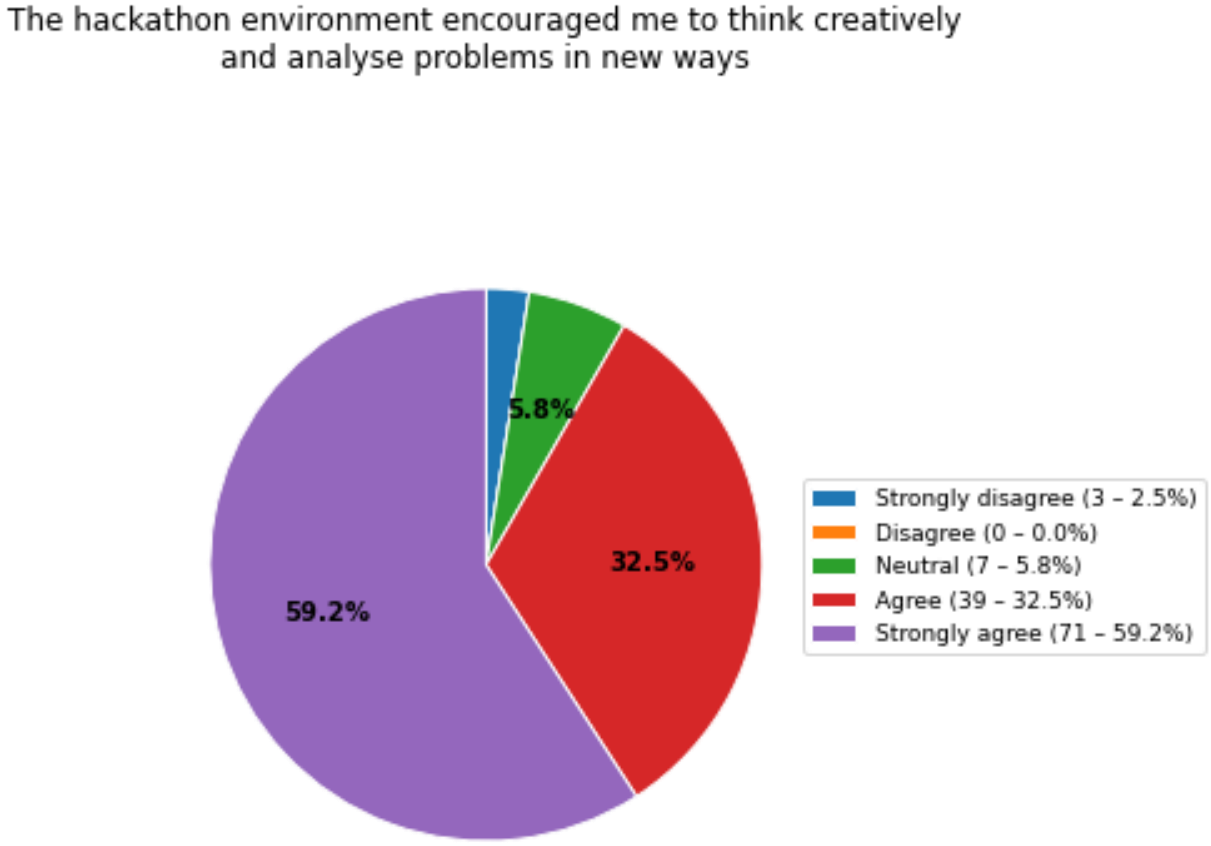
20 beat the benchmark

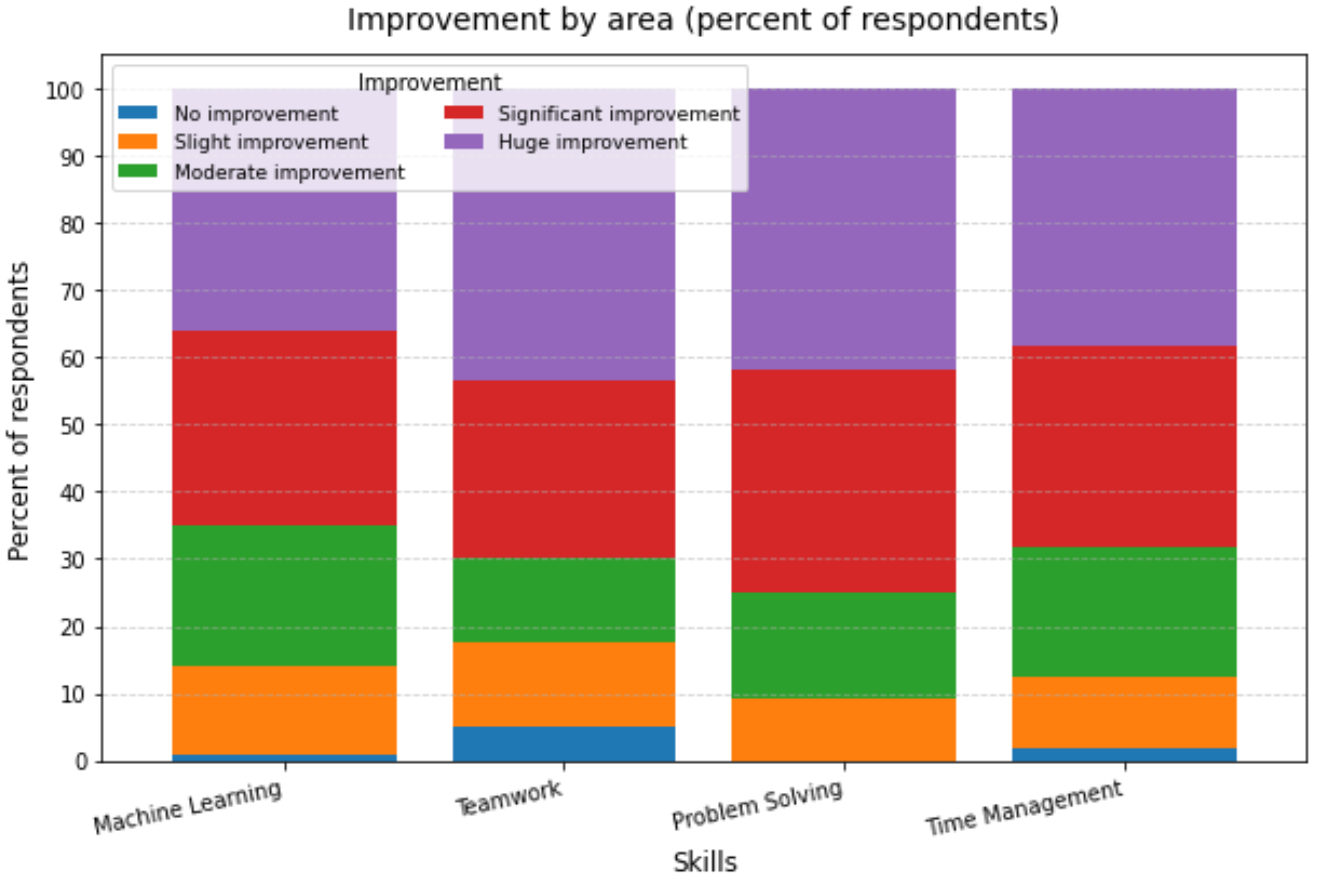
165× faster than baseline

Feedback from previous cycle

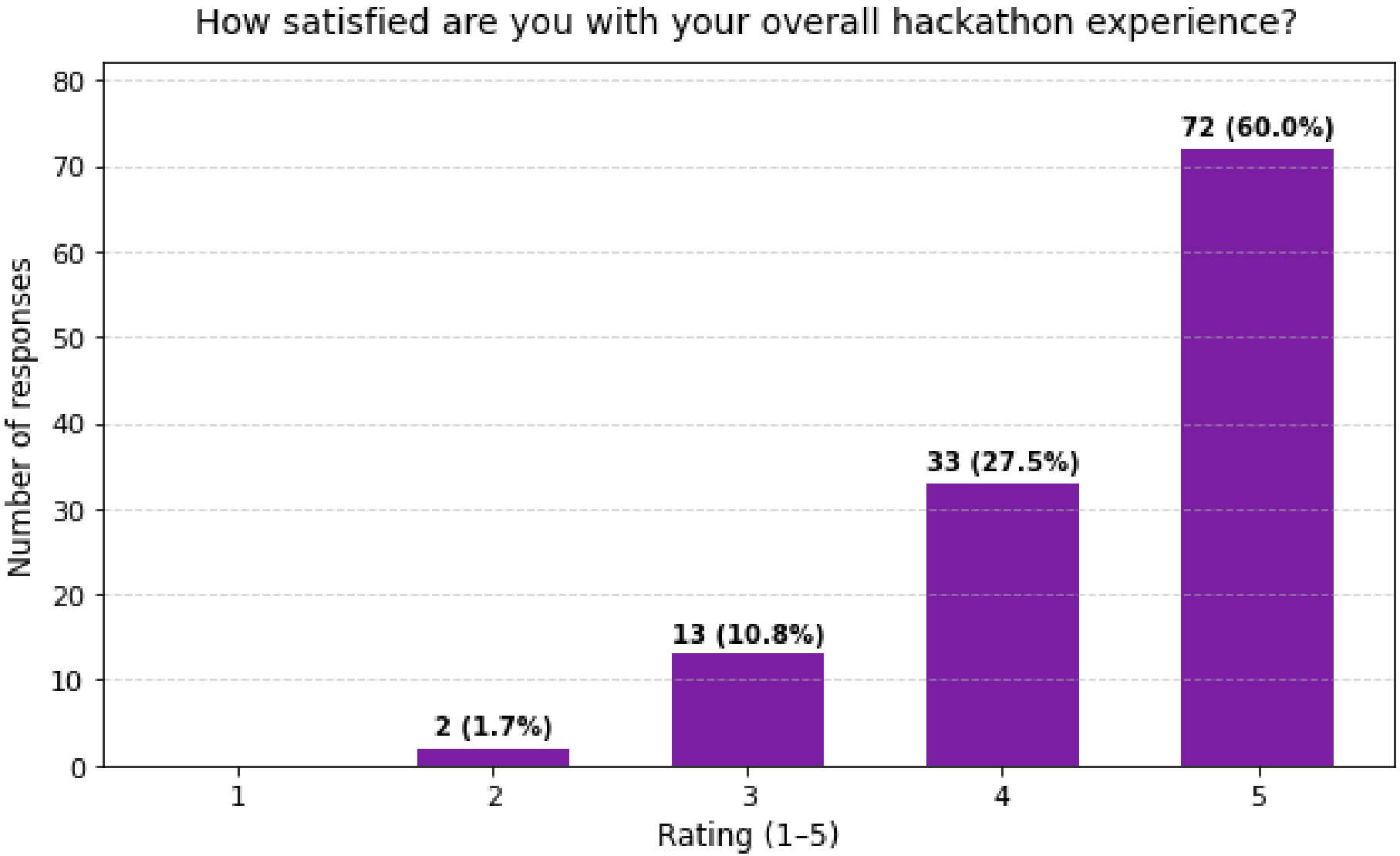


Feedback from previous cycle





Feedback from previous cycle



What Participants Gain

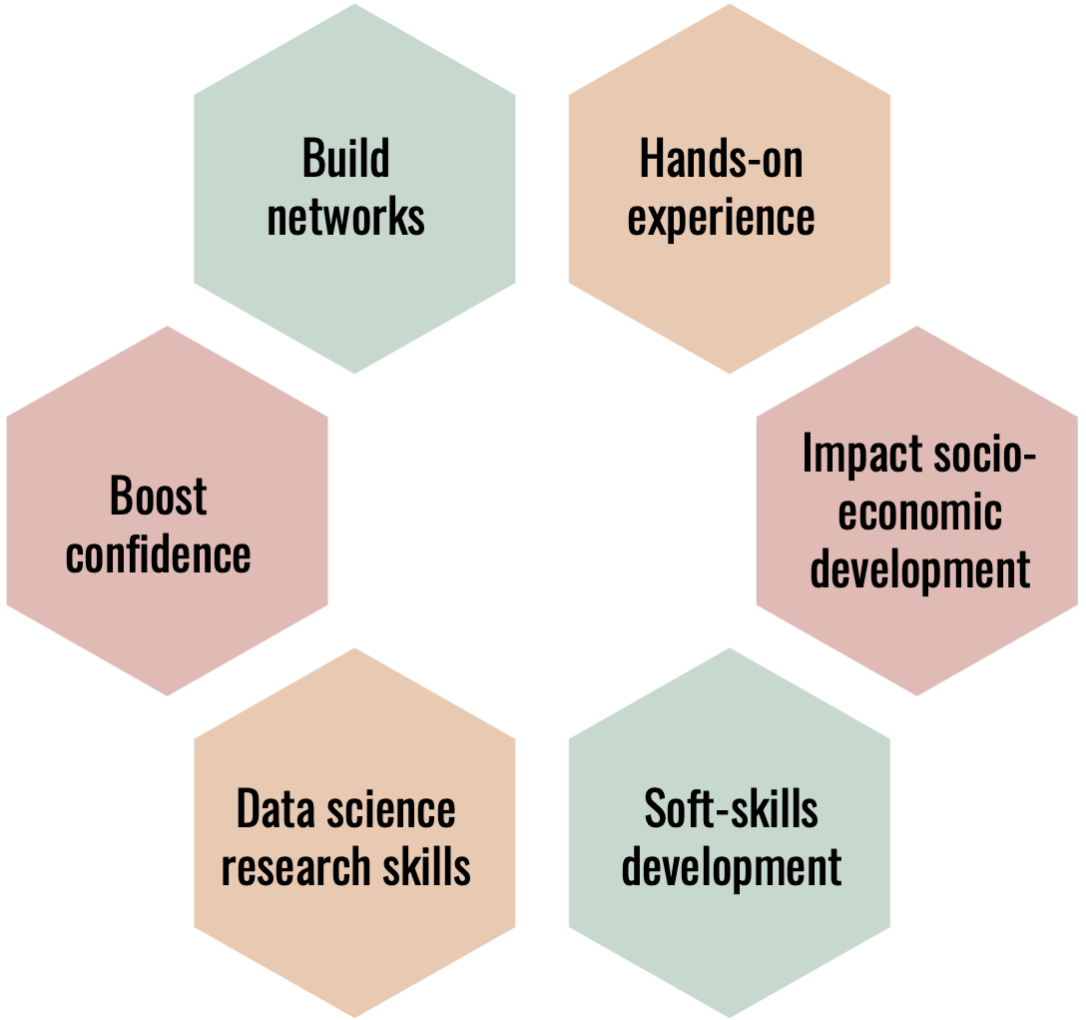
Short, challenge-driven hackathons build technical competence and encourage peer-to-peer learning at the same time.

Confidence: real submissions under real deadlines build self-belief.

Technical skills: hands-on machine learning and data science practice.

Networks: peer, mentor, and cross-institution connections.

Soft skills: teamwork, communication, and presenting results.



Scaling Up: 2025–26 Cycle

The jump:

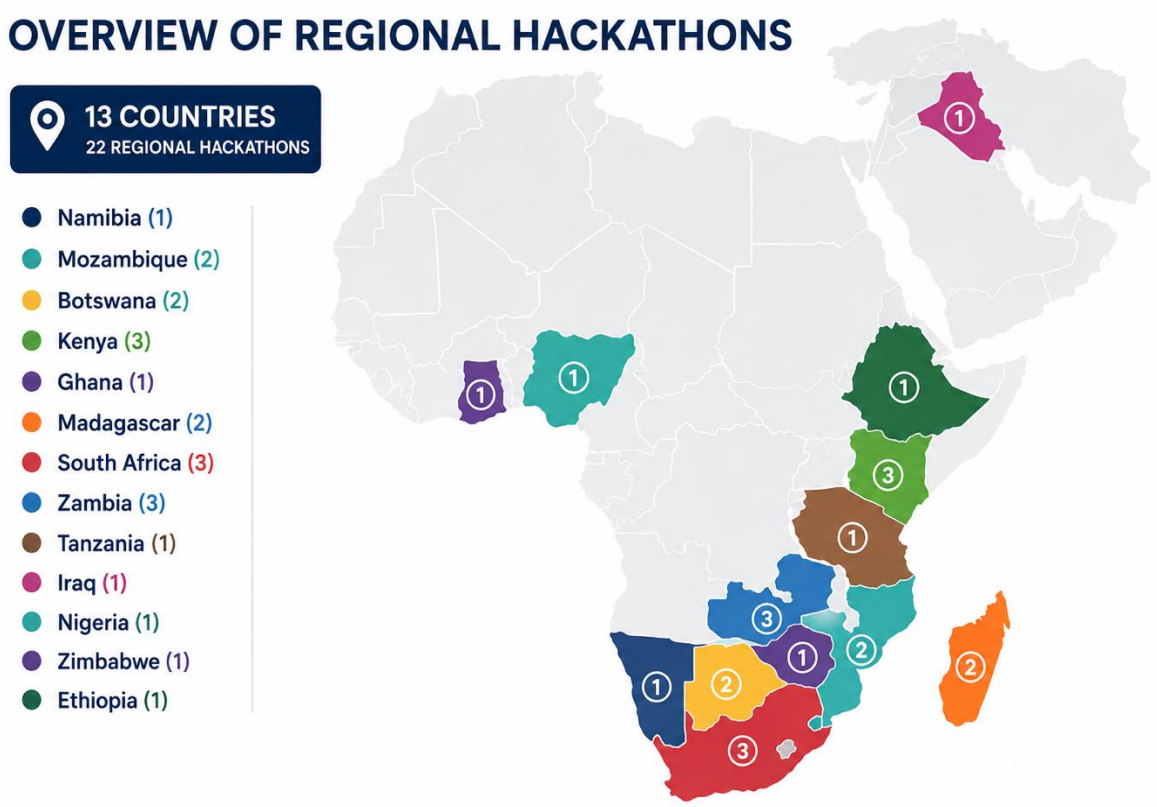
- 22 hackathons
- 13 countries
- Diverse institutions

- The next cycle expands to 22 hackathons hosted across 13 countries.

- It becomes a live test of how well AI and data science capacity-building scales across diverse educational and institutional contexts.

- Funding partners – DARA, OAD, AfAS, and RIIS – support hackathons across the network.

- A shared challenge framework keeps every regional event connected.



Looking Ahead

Next:

- New organisers and institutions
- More mentorship and training
- Stronger regional ecosystems

- A forthcoming third cycle will recruit new organisers and institutions, widening Hack4dev's reach further.

- More hands-on ML training, mentorship, and community-led innovation – strengthening regional data science ecosystems.

- Encourage use of AI tools, partnerships with providers

Lessons Learnt

Practical insights from the first cycle that shape how each new hackathon is run.

Facilitation: tutors guide without giving answers; prepare early; share notebooks ahead.

Selection: clean the data first; recommendation letters matter; allow time for scoring.

Teams: balance ML skill, Python, and gender; every team presents a result.

Pace & care: set clear deadlines; teach concepts over syntax; uphold the code of conduct.

Partners & Funders

Core Partners:

Industry Partners:

Exploration phase:

BREAKTHROUGH LISTEN

Thank You!

CONTACT US

charles@astro4dev.org
hack4dev.org

Credit: NASA, ESA, CSA, and STScI