Yandex SCHOOL OF DATA ANALYSIS

Towards gig science and augmented intelligence

July, 2019, DESY

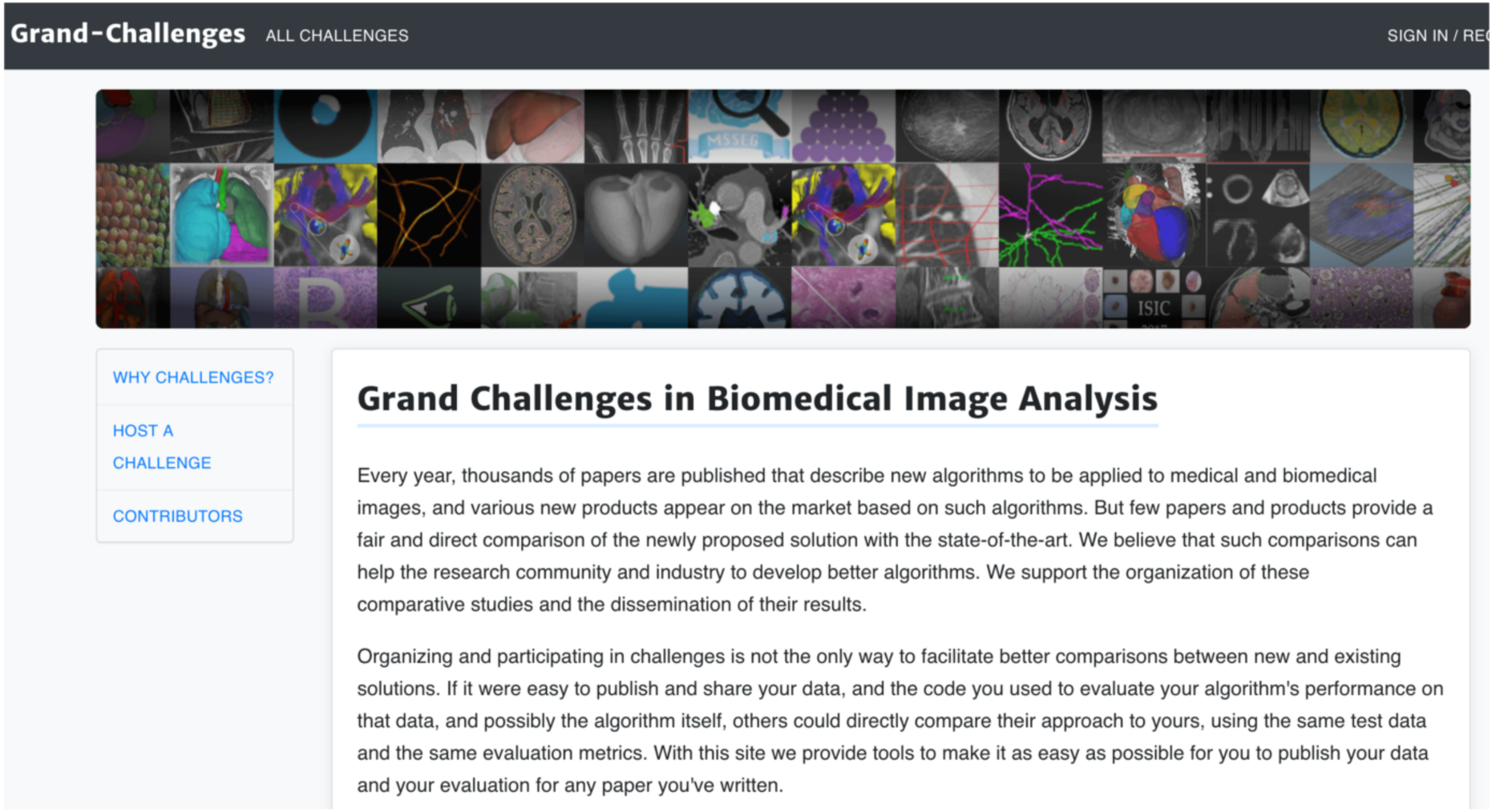
Andrey Ustyuzhanin

NRU HSE

YSDA

ICL

Machine Learning + Science examples



Machine Learning + Science examples


WE NEED TO GO
DEEPER
memegenerator.net

Is there a space for gig-science?

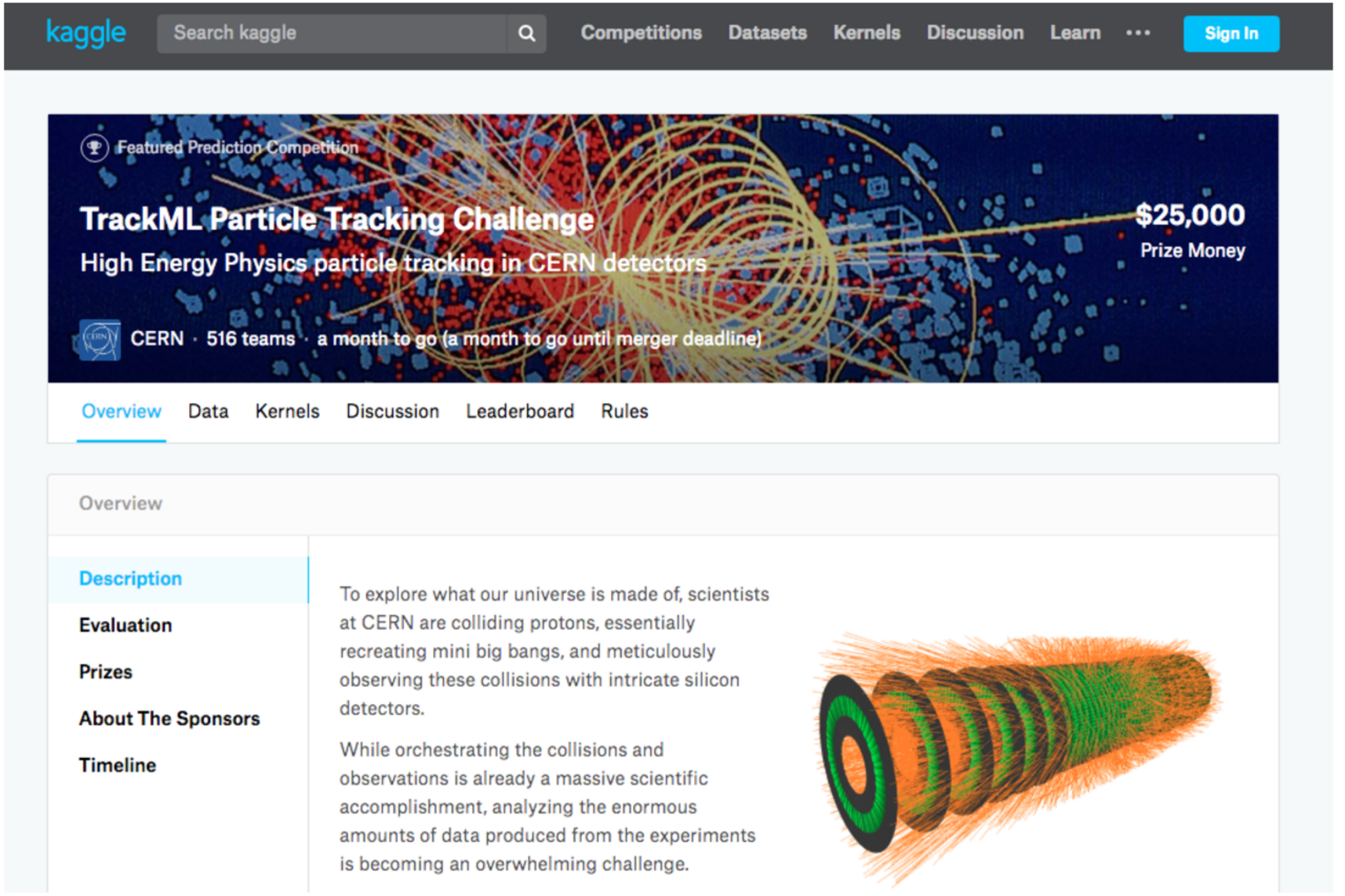
Science as a free-lance:

- Kolabtree.com
- http://labmate.us/
- nature.com/nature/journal/v550/n7676/full/nj7676-419a.html

Maybe Kaggle.com could help?

- Strong community
- Numerous challenges

Machine Learning + Science examples



Kaggle shortcomings

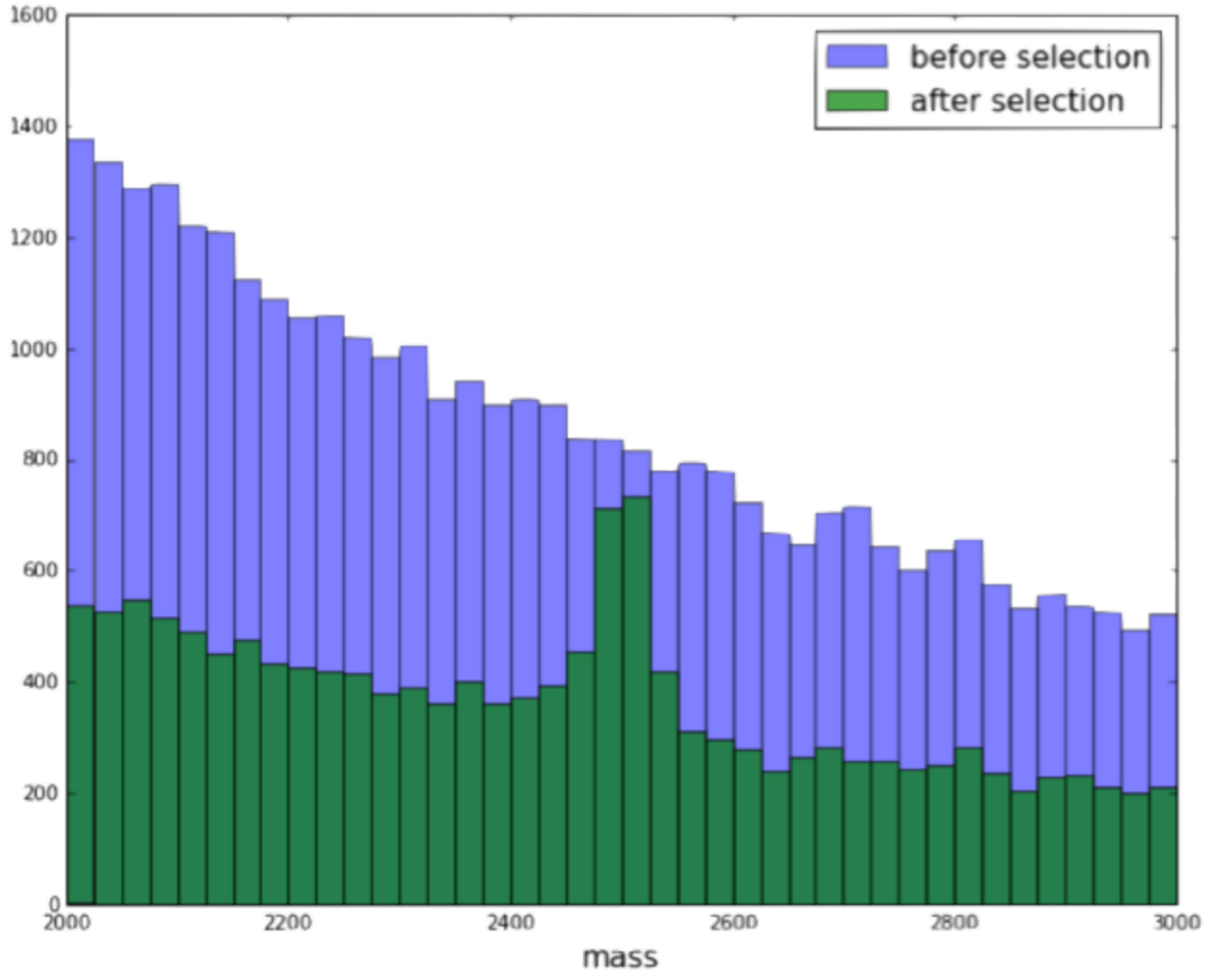
Domain specifics:

- Non-trivial metrics
- Constraints

AI beyond iid assumption (test set is not the same as training set)

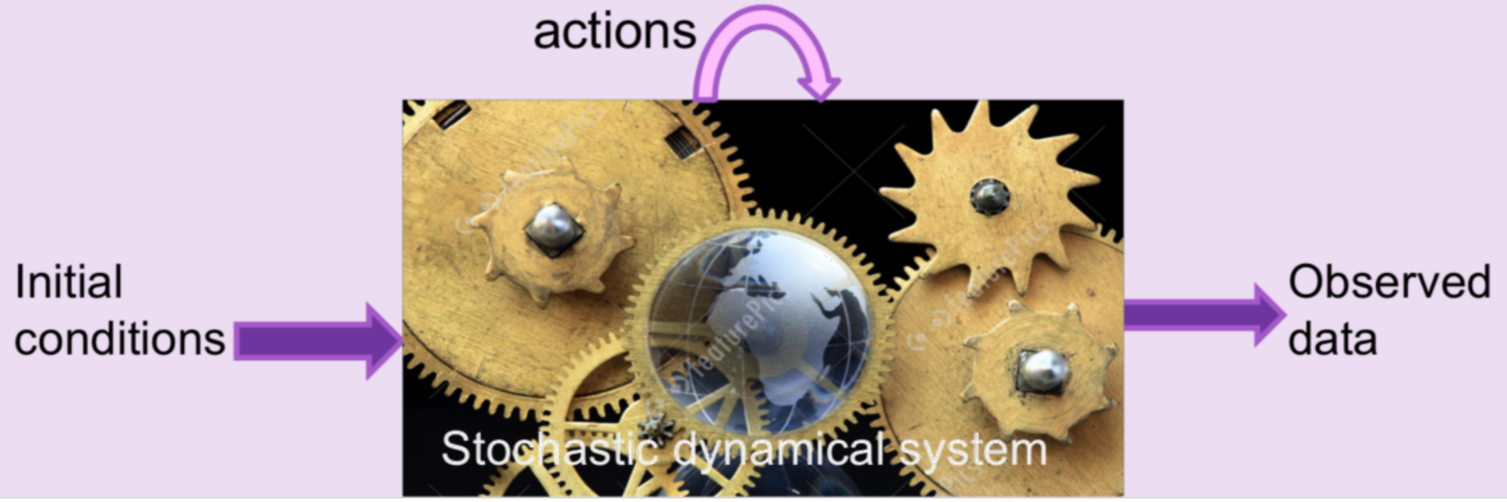
Participation Incentive

2. Domain specific example



Beyond iid assumption: causal mechanisms

- The assumption that the test data is from the same distribution as the training data is too strong, and it is often violated in practice, leading to poor out-of-distribution generalization.
- Consider relaxed assumptions: the test data was generated under the same ***causal dynamics***, but from different initial conditions (which may be unlikely under the training distribution) and agents' actions.



Mila

32

Beyond iid assumption

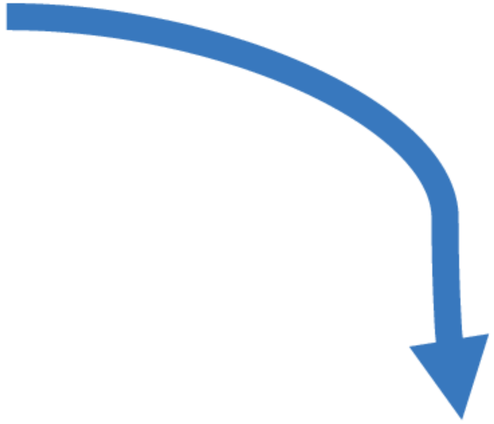
RoboCup
OpenAI gym
NeurIPS CIML workshop highlights

- AI Driving Olympics: Duckie Town
- Adversarial Attacks
- Chat bots
- TrackML

BabyAI

3. Kaggle incentive: winner takes it all

Can it be more fair?



Micro-rewards

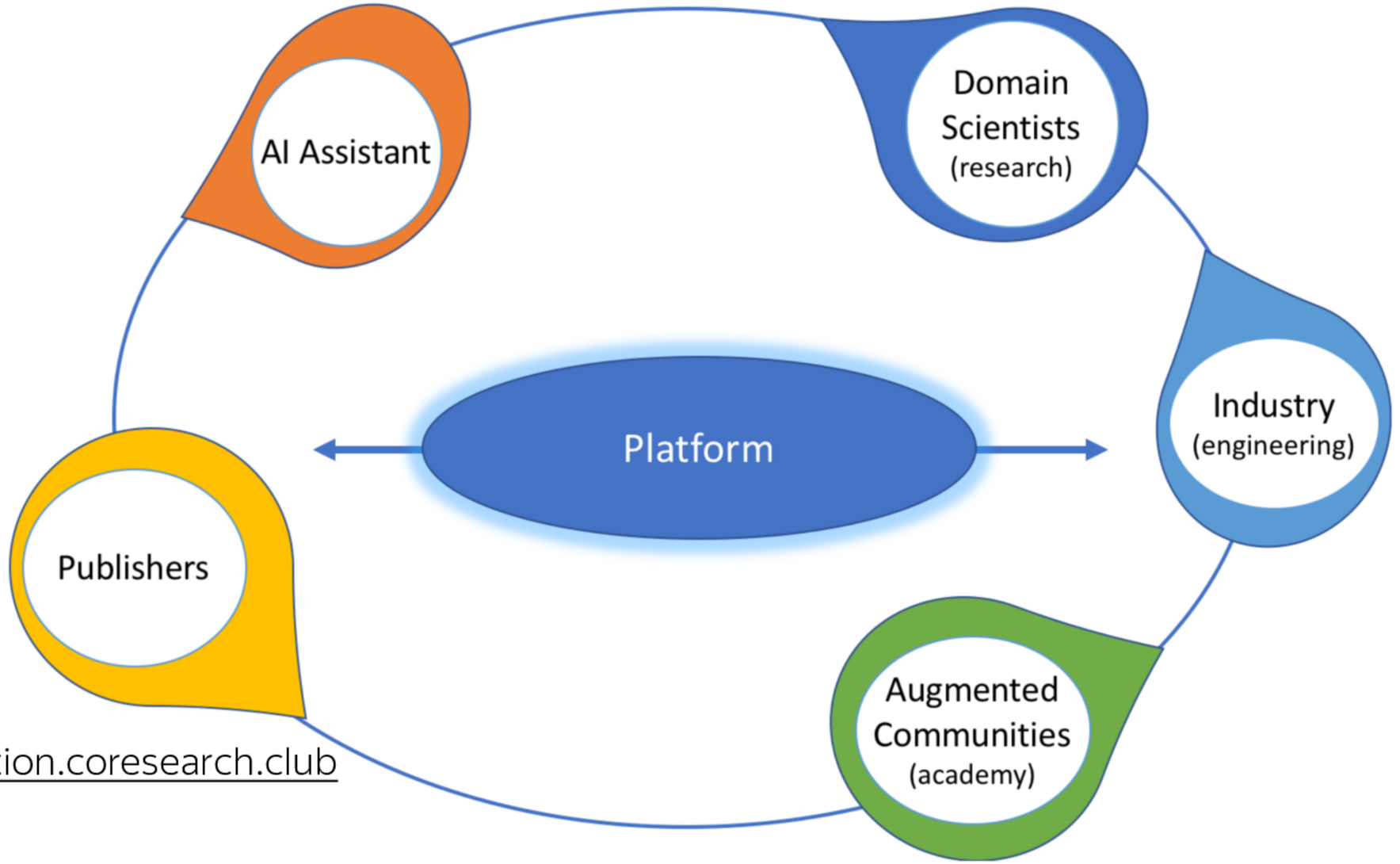
Metric-based:

- top X of the leaderboard,
- time at the top,
- persistence

Source code-based:

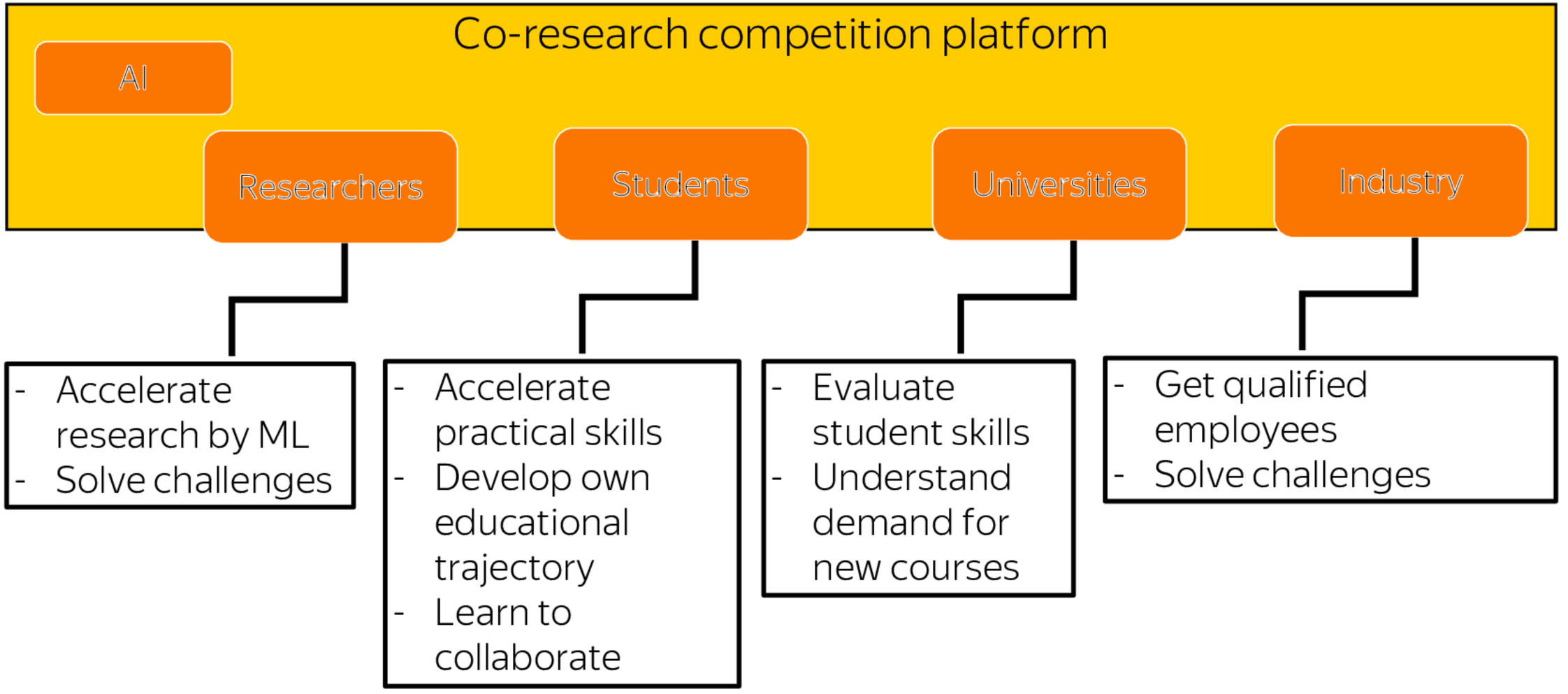
- Distinctive solution
- Source of other's inspirations
- Your code is re-used by others

Co-research platform

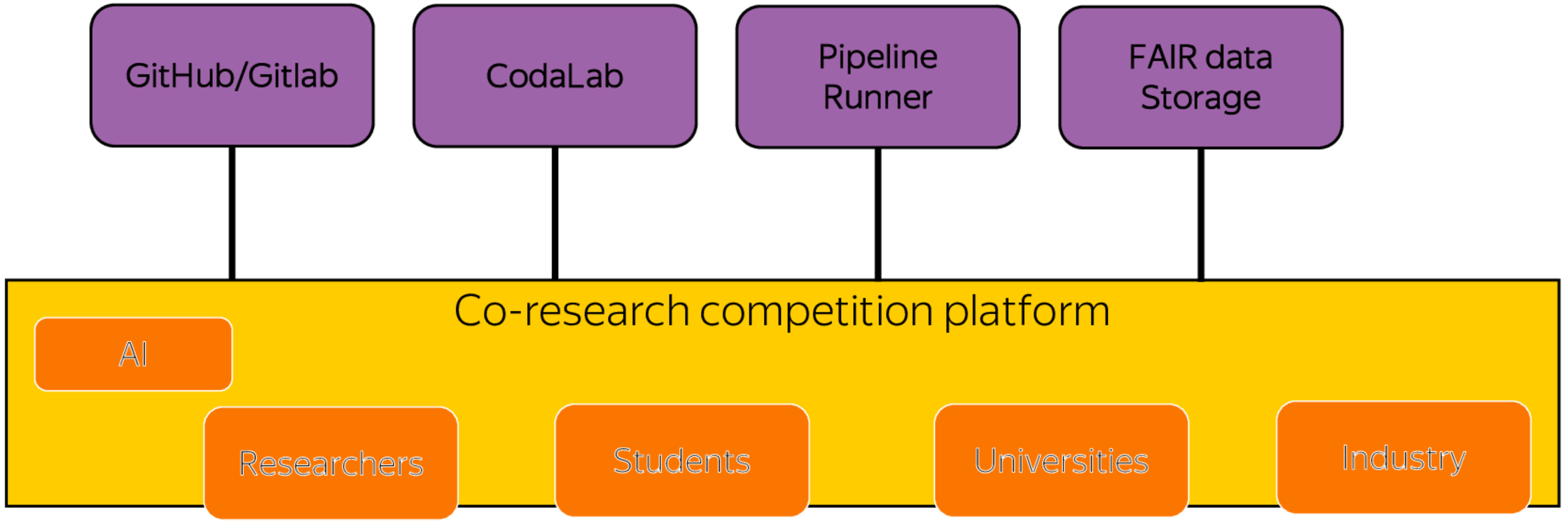


https://coopetition.coresearch.club

Co-Research Platform overview



Co-Research Platform overview



Regular Participant's Workflow

Register at the platform, chose a co-research project

- Read materials, research map, get initial advise to try

Fork baseline or other's participant github repository

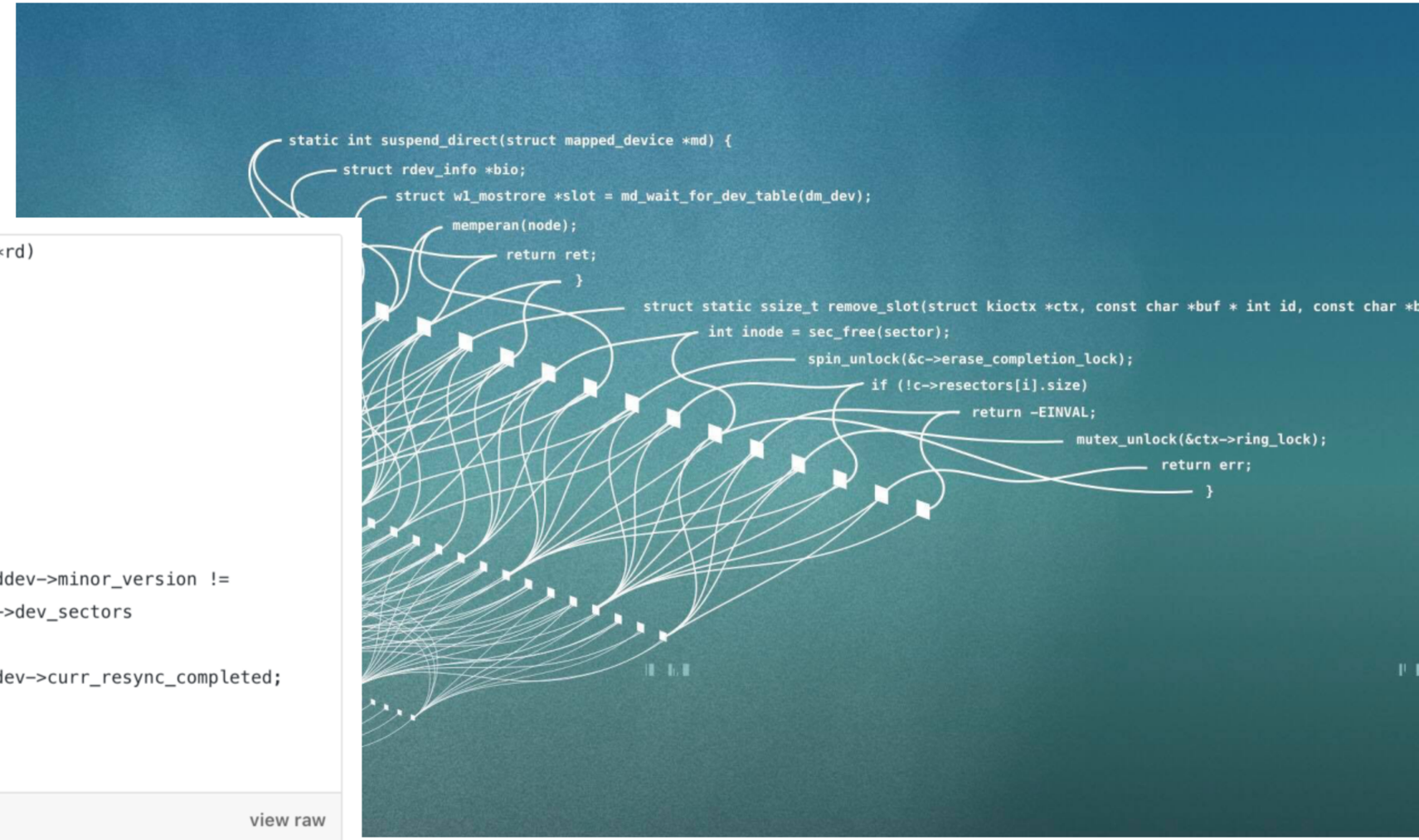
Iterate

- Read other's submissions / explore / communicate
- Realize an idea to try
- Write the code, commit to own repository
- Get evaluation, recommendations

AutoML opportunity

```
1  static int super_fold(struct mddev *mddev, void __user * *rd)
2  {
3          struct md_rdev *rdev;
4
5          if (!tryet &  gcov_ntreef(*stint)) {
6                  if (gc_th->max_sectors)
7                          if (task)
8                          goto next_start;
9                  if (!list_empty(&mddev->disks) {
10                         if (mddev->dev_sectors == 0 ||
11                             mddev->chunk_sectors == 0 && mddev->minor_version !=
12                                  mddev->max_disks && mddev->dev_sectors
13                                rdev2->rescan_recovnr != 0)
14                                 rdev->recovery_offset = mddev->curr_resync_completed;
15                 }
16         }
17 }
```

deep_generation_c_code_8.c hosted with ❤ by GitHub view raw



Conclusion

Coopetition is a platform for multidisciplinary data-driven research

- Motivation for at personal and scientific growth
- Interaction between domain-experts and machine-learning experts
- Source code-based submissions
- AutoML for AI research assistant

https://coopetition.coresearch.club
anaderiRu@twitter
austyuzhanin@hse.ru

Backup