Факультет компьютерных наук НИУ ВШЭ

XIX конференция «Свободное программное обеспечение в высшей школе»

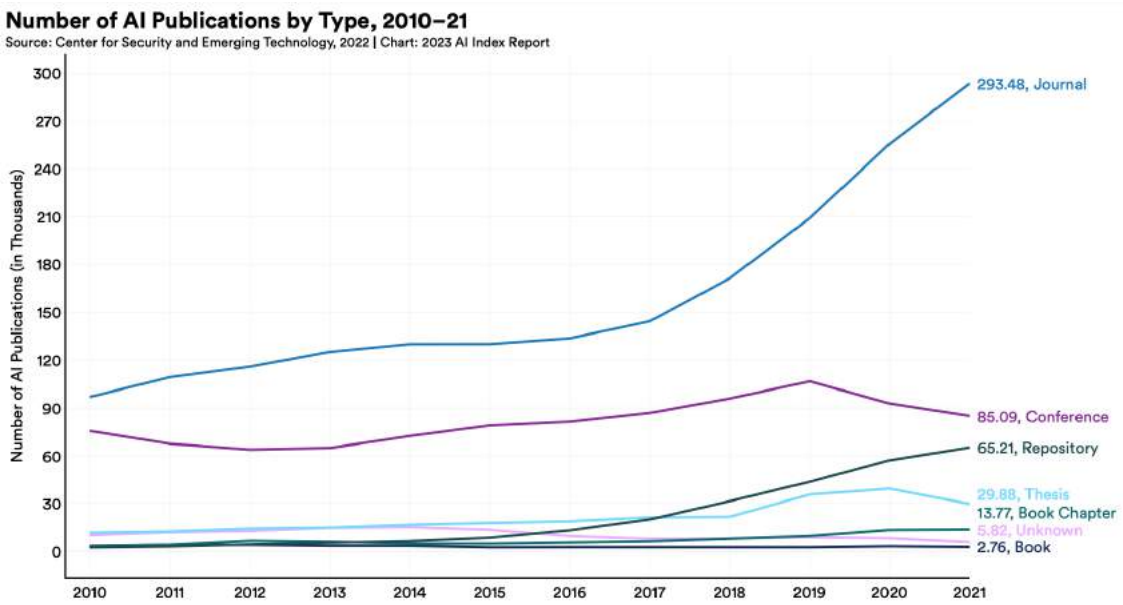
28 – 30 июня 2024
г. Переславль-Залесский

Проект открытого кода научных исследований ФКН

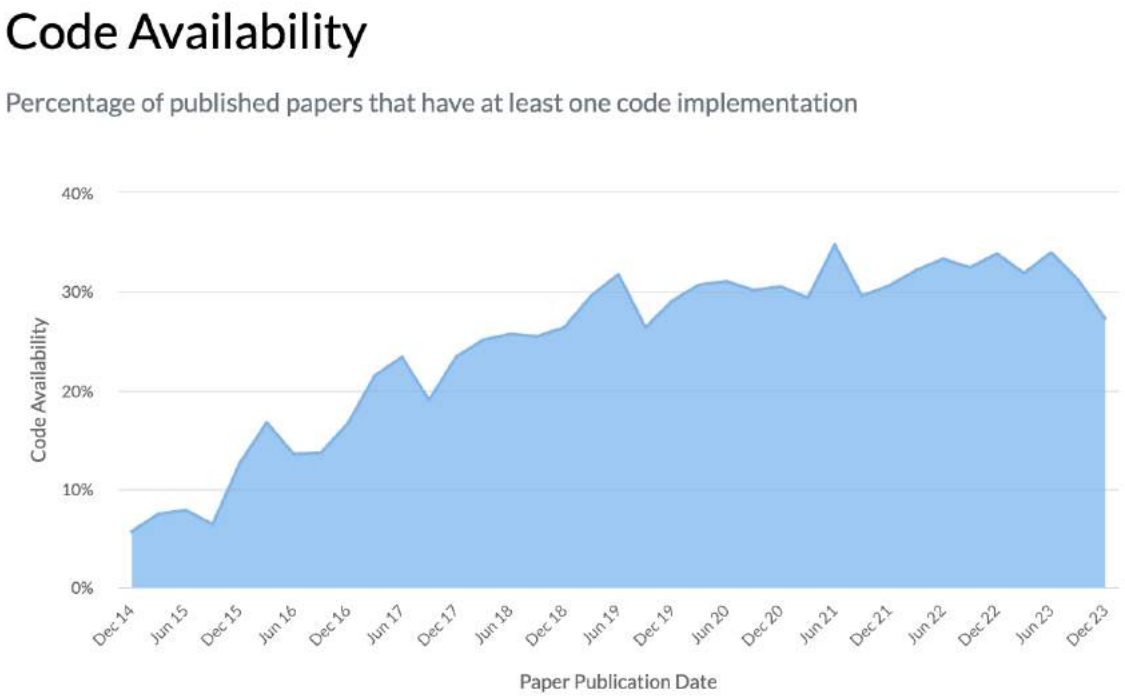
Гущин Михаил, mhushchyn@hse.ru

ФКН НИУ ВШЭ

Публикации и открытый код по ИИ

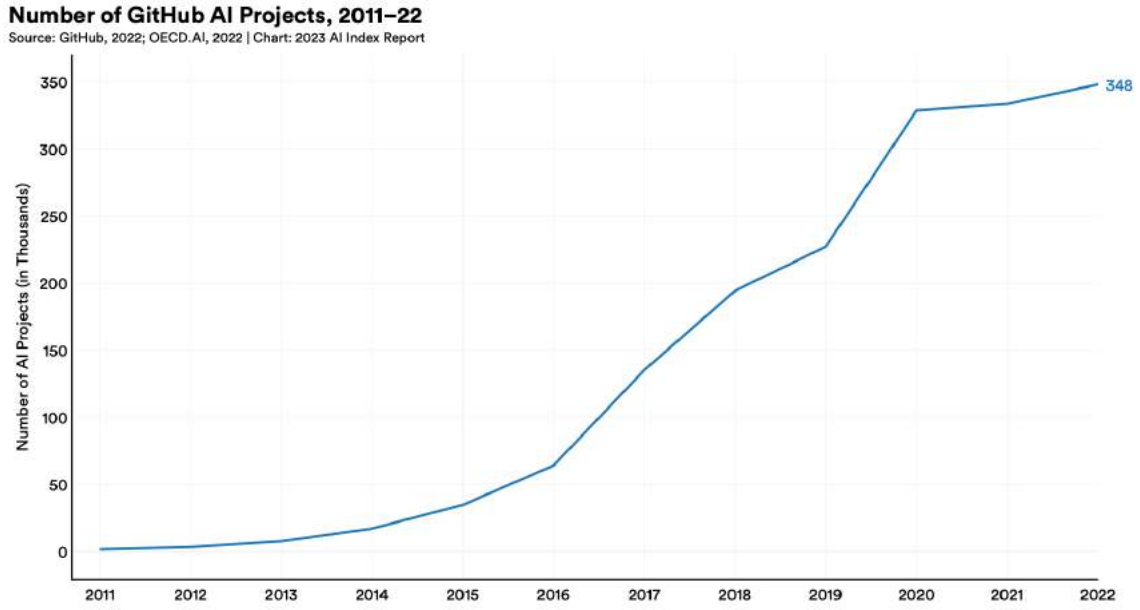


Более **290 000** публикаций в научных журналах в области ИИ за 2021 год согласно 2023 AI Index Report (Stanford)

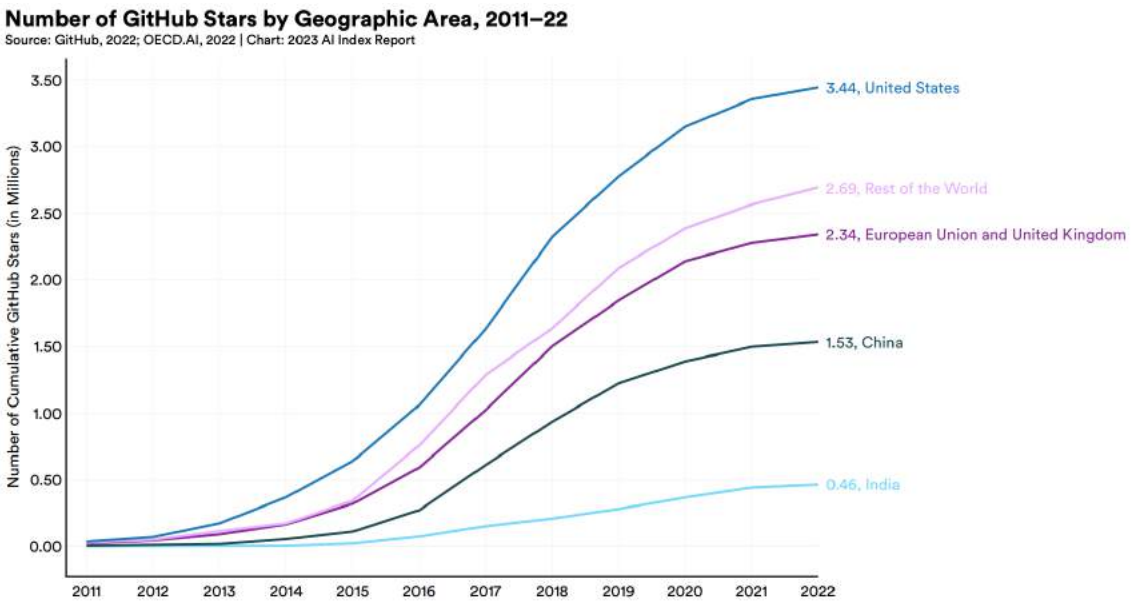


Более 30% публикаций имеют хотя бы одну открытую реализацию в коде, согласно paperswithcode.com/trends

Проекты по ИИ с открытым кодом на GitHub



Более 330 000 проектов с открытым кодом в области ИИ за 2021 год согласно 2023 AI Index Report (Stanford) (против 290 000 публикаций в журналах)



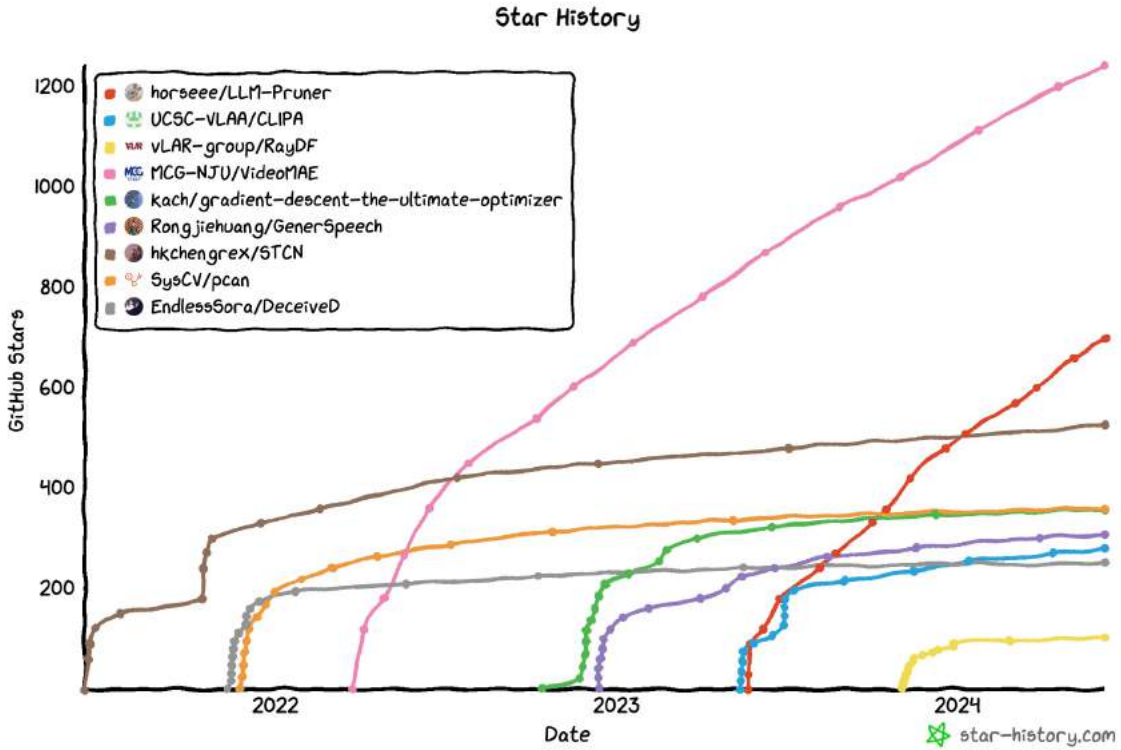
Более 8 000 000 звезд в проектах в области ИИ к 2022 году согласно 2023 AI Index Report (Stanford)

Код статей NeurIPS на GitHub

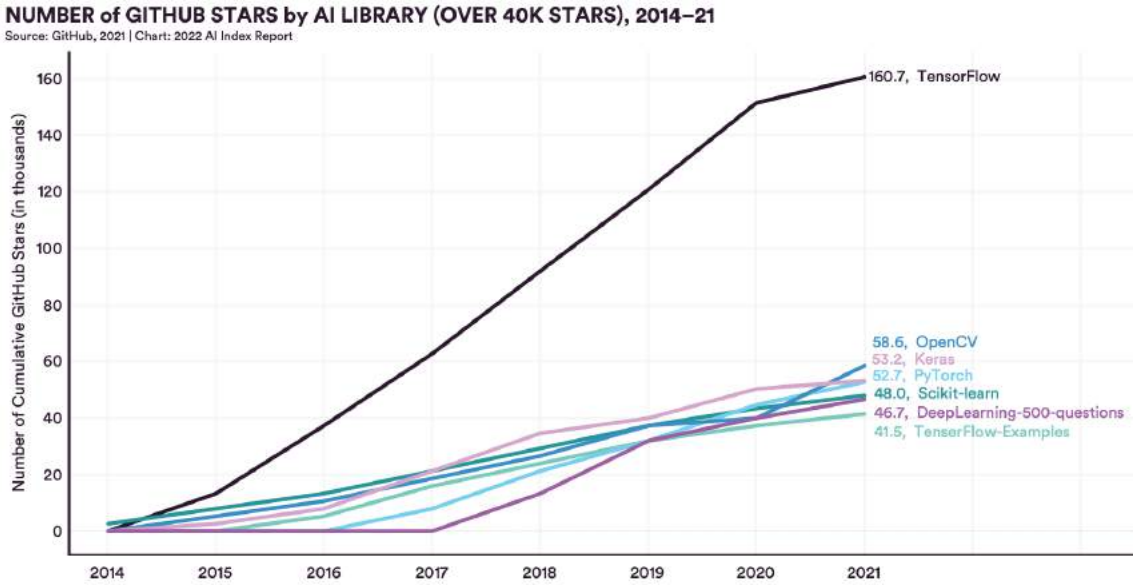
Год	Статья	Код	# цитирований на Scholar	# форков на GitHub	# звезд на GitHub
2023	LLM-Pruner: On the Structural Pruning of Large Language Models	https://github.com/horseee/LLM-Pruner	148	74	698
	An Inverse Scaling Law for CLIP Training	https://github.com/UCSC-VLAA/CLIPA	19	10	280
	RayDF: Neural Ray-surface Distance Fields with Multi-view Consistency	https://github.com/vLAR-group/RayDF	2	4	102
2022	VideoMAE: Masked Autoencoders are Data-Efficient Learners for Self-Supervised Video Pre-Training	https://github.com/MCG-NJU/VideoMAE	648	123	1241
	Gradient Descent: The Ultimate Optimizer	https://github.com/kach/gradient-descent-the-ultimate-optimizer	34	25	358
	GenerSpeech: Towards Style Transfer for Generalizable Out-Of-Domain Text-to-Speech	https://github.com/Rongjiehuang/GenerSpeech	62	44	307
2021	Rethinking Space-Time Networks with Improved Memory Coverage for Efficient Video Object Segmentation	https://github.com/hkchengrex/STCN	272	71	525
	Prototypical Cross-Attention Networks for Multiple Object Tracking and Segmentation	https://github.com/SysCV/pcan	82	50	360
	Deceive D: Adaptive Pseudo Augmentation for GAN Training with Limited Data	https://github.com/EndlessSora/DeceiveD	99	24	251

Источник: https://github.com/topics/neurips-2021?o=desc&s=stars

Популярность проектов на GitHub



Открытый код на GitHub создает быстрый рост популярности результатов исследований



Библиотеки – самые популярные проекты по ИИ согласно 2022 AI Index Report (Stanford)

Организации на GitHub (# звезд самого популярного проекта)

Мировые университеты:

- https://github.com/stanfordmlgroup (1550*)
- https://github.com/stanfordnlp (9343*)
- https://github.com/stanford-crfm (1499*)
- https://github.com/mit-han-lab (5894*)
- https://github.com/mit-acl (821*)
- https://github.com/MIT-LCP (2219*)
- https://github.com/thu-coai (1639*)
- https://github.com/AllenInstitute (316*)
- https://github.com/flatironinstitute (581*)

Организации:

- https://github.com/openai (56k*)
- https://github.com/google-research (32k*)
- https://github.com/facebook (219k*)
- https://github.com/apple (65k*)

Организации РФ:

- https://github.com/yandexdataschool (9570)
- https://github.com/Yandex (8197*)

Университеты РФ:

- https://github.com/SciProgCentre (637*)
- https://github.com/aimclub (618*)
- https://github.com/SkoltechRobotics (185*)
- https://github.com/AIRI-Institute (339*)
- https://github.com/cig-skoltech (94*)

НИУ ВШЭ:

- https://github.com/bayesgroup (1046*)
- https://github.com/HSE-LAMBDA (100*)

Открытый код ФКН

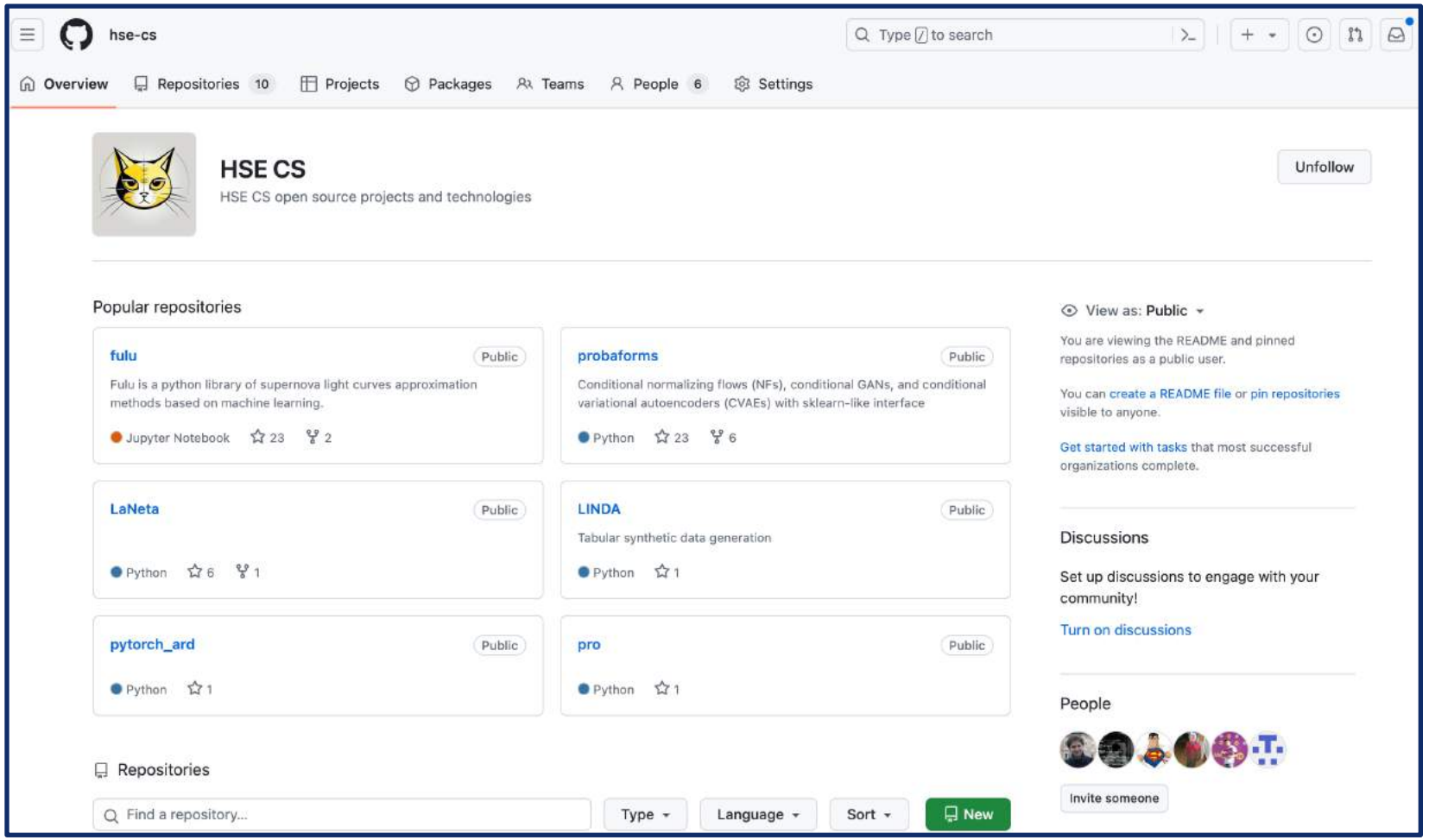
Цели проекта по открытому коду ФКН:

- Увеличить видимость и узнаваемость кода ФКН
- Увеличить вовлеченность студентов в проектах ФКН
- Привлечь сторонних разработчиков в проекты ФКН

Задачи проекта:

- Объединить код проектов факультета
- Привлекать студентов к проектам ФКН в рамках КР, ВКР и летних практик
- Организация докладов о лучших практиках разработки открытого кода
- Подготовка рекомендаций по разработке и поддержке проектов

ФКН на GitHub



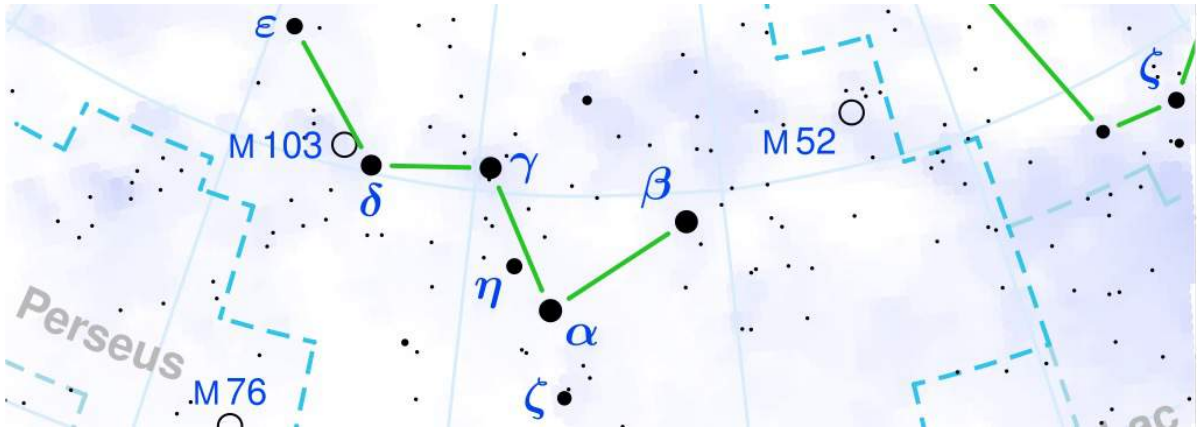
https://github.com/hse-cs

Наши коллеги ☺

Организация	Описание	Самые популярные репозитории
ITMO AIM.CLUB	Объединенный репозиторий AI/ML фреймворков ИТМО	FEDOT, BAMT, FEDOT.Indsustrial
ВШЭ	Репозитории AI/ML фреймворков ВШЭ	hsemotion, roerich, probaforms
МФТИ, SPC	Подборка проектов МФТИ	DeepPavlov, kmath
Сколтех	Новые официальные форки бывшего репозитория Сколтеха	ttpy, h2tools

Fulu

- https://github.com/hse-cs/fulu
- Библиотека методов аппроксимации кривых блеска астрономических объектов с использованием нейронных сетей.
- Названа в честь звезды Дзета Кассиопеи в 590 световых годах от нас, которая имеет официальное название Fulu
- Разработана в результате совместного научного исследования 7 организаций из 3 стран

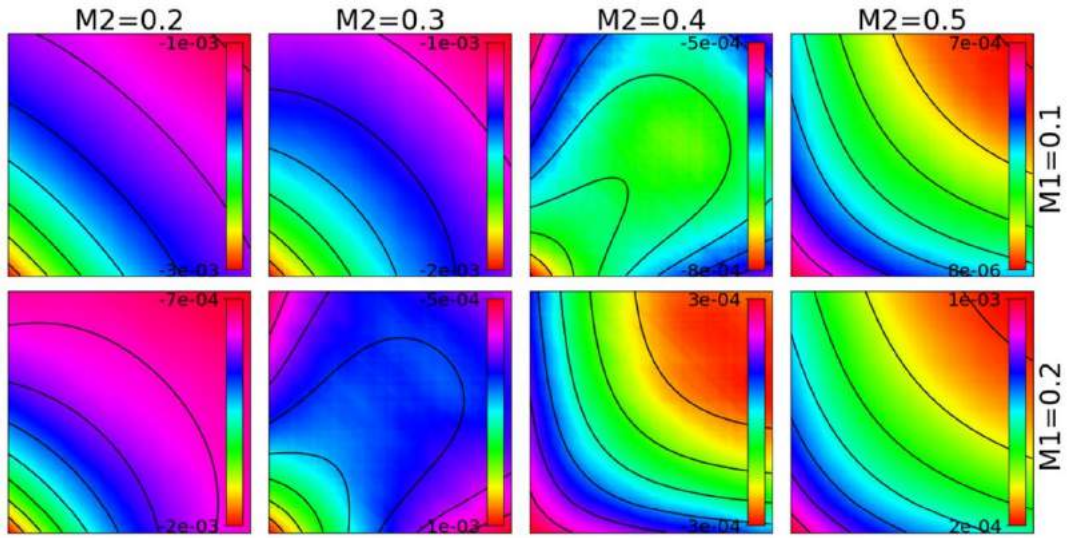


LaNeta

- https://github.com/hse-cs/LaNeta

- Библиотека для оценки времен примешивания между двумя популяциями при двух пульсах миграции.

- Позволяет точно исследовать недавнюю (в пределах нескольких десятков поколений) историю примешивания популяций в сложных сценариях, для которых существовавшие ранее методы были неприменимы или неточны.

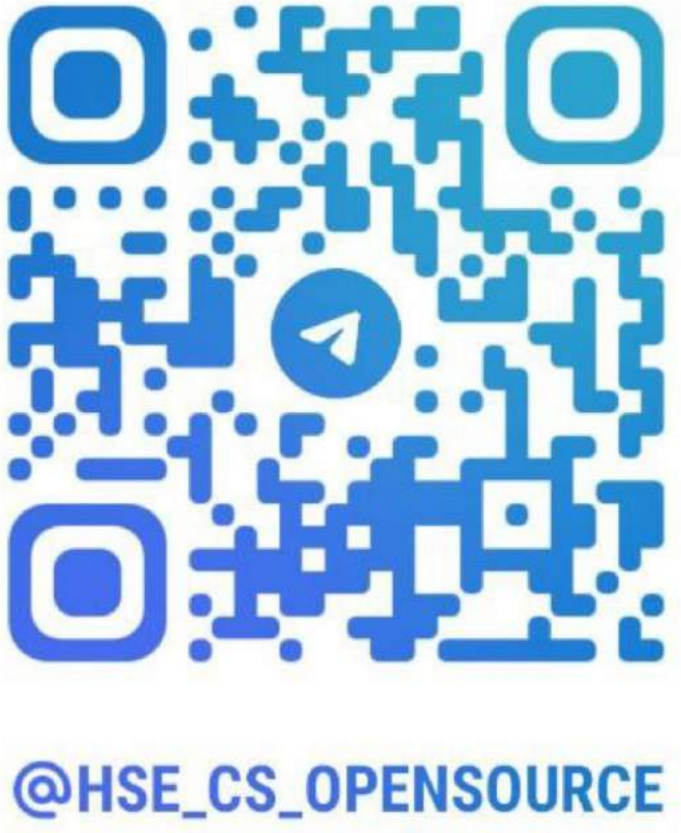


Заключение

Сайт: https://cs.hse.ru/opensource

Github: https://github.com/hse-cs

Канал: https://t.me/hse_cs_opensource



Контакты:

Гущин Михаил

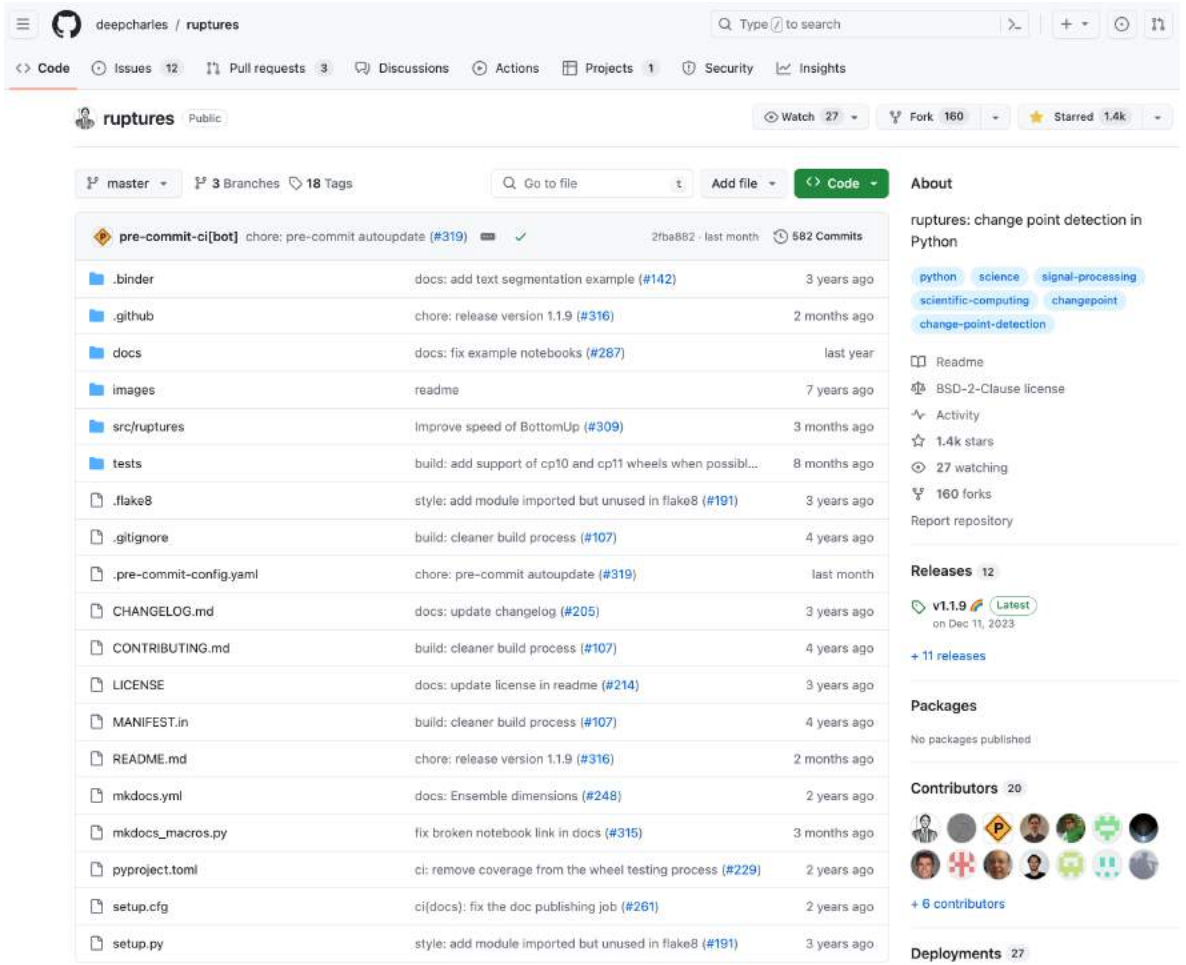
mhushchyn@hse.ru, @mikhail_h91


ВЫСШАЯ • ШКОЛА • ЭКОНОМИКИ •

Ruptures

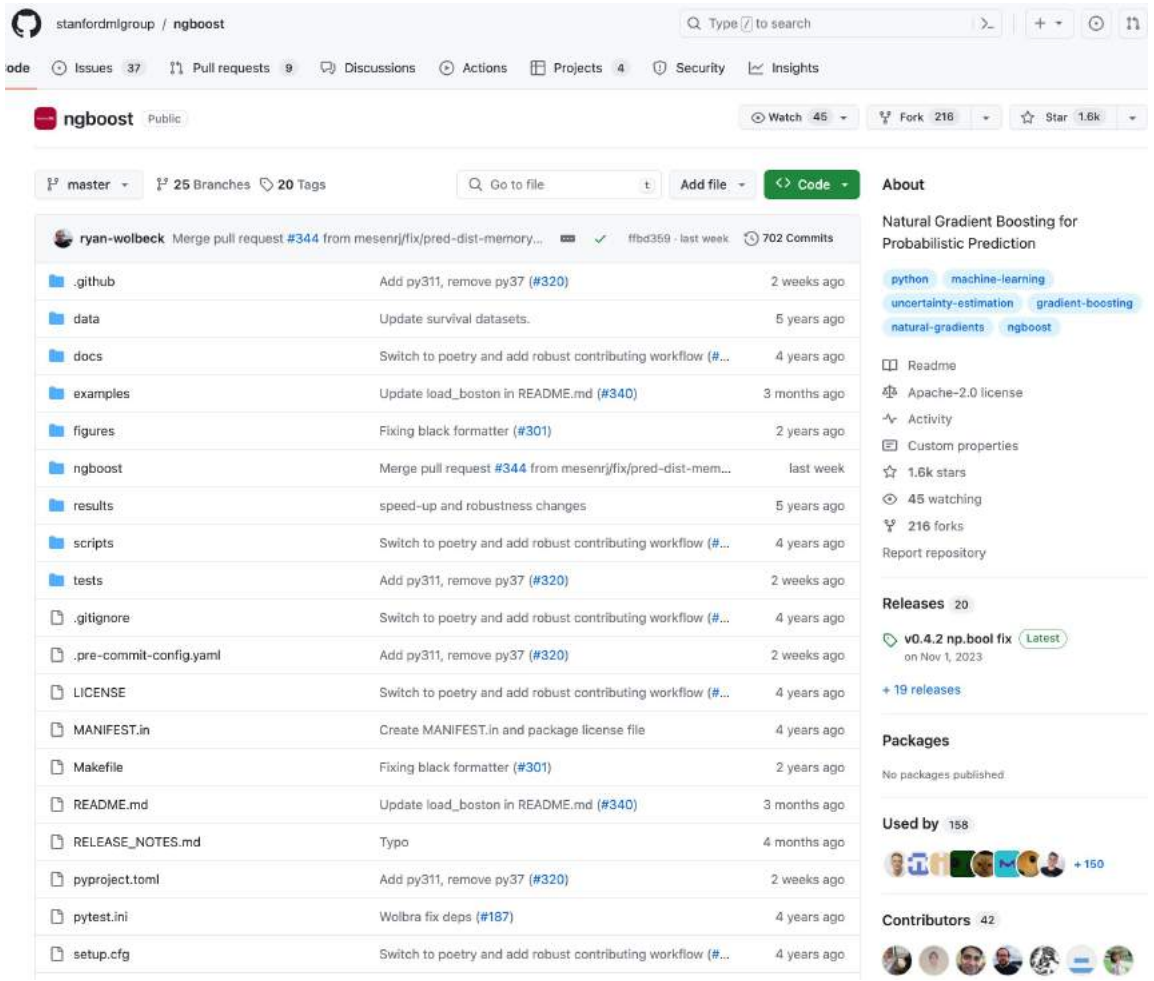
- C. Truong, L. Oudre, N. Vayatis. Selective review of offline change point detection methods. Signal Processing, 167:107299, 2020
  - 988 цитирований

- Код по статье: https://github.com/deepcharles/ruptures
  - 1400+ звезд
  - 11M скачиваний кода



NGBoost

- Tony Duan, Anand Avati, Daisy Yi Ding, Khanh K. Thai, Sanjay Basu, Andrew Ng, and Alejandro Schuler. 2020. NGBoost: natural gradient boosting for probabilistic prediction. In Proceedings of the 37th International Conference on Machine Learning (ICML'20), Vol. 119. JMLR.org, Article 252, 2690–2700.
  - 291 цитирование
- Код по статье: https://github.com/stanfordmlgroup/ngboost
  - 1600+ звезд
  - 11к скачиваний кода в месяц
- На странице Stanford ML Group: https://github.com/stanfordmlgroup



Streaming LLM

- Efficient Streaming Language Models with Attention Sinks, ICLR 2024
  - 48 цитирований
- Код по статье: https://github.com/mit-han-lab/streaming-llm
  - 5894 звезды
- На странице MIT HAN LAB: https://github.com/mit-han-lab