

The logo for ITMO University, consisting of the letters 'ITMO' in a bold, white, sans-serif font. The background is a dark purple grid with white wavy lines.

# Зачем математику репозитории и коммиты?

Александр Хватов, зав. лаб. (Лаборатория композитного ИИ)

[mailto:alex\\_hvatov@itmo.ru](mailto:alex_hvatov@itmo.ru)

ITMO OS Meetup, 16.12.2022

# Содержание

- Постановка задачи, начальные и граничные условия
- Вывод уравнения современного научного эксперимента
- Численный эксперимент
- Обсуждение и выводы

# Первый этап — «Отрицание и изоляция» ИТМО

Типичный учёный:

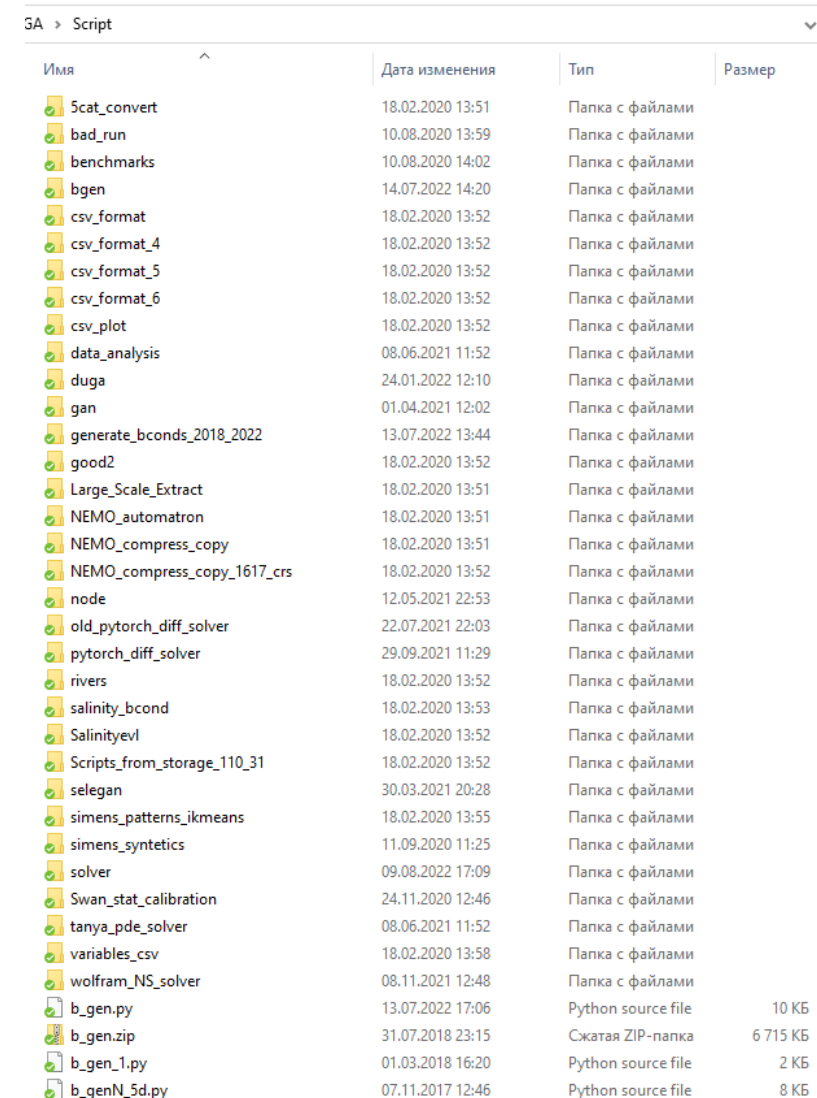
- Пишет код каждую новую статью
- Старый код хранится в папке scripts (или на почте)
- “Какой *гений* написал этот код?»
- Переиспользовать код == создать папку (скрипт) с новым именем и изменить всё внутри кода

В такой постановке:

- Конкретная задача решается
- Код доступен и даже «переиспользуется»

Граничные условия:

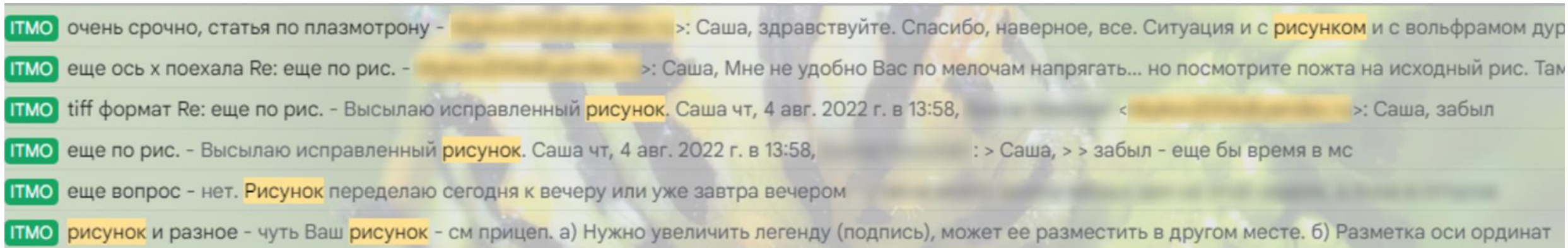
- Совместная работа невозможна
- Через несколько лет код «теряется»
- Серьёзные изменения == новый код



Имя	Дата изменения	Тип	Размер
5cat_convert	18.02.2020 13:51	Папка с файлами	
bad_run	10.08.2020 13:59	Папка с файлами	
benchmarks	10.08.2020 14:02	Папка с файлами	
bgen	14.07.2022 14:20	Папка с файлами	
csv_format	18.02.2020 13:52	Папка с файлами	
csv_format_4	18.02.2020 13:52	Папка с файлами	
csv_format_5	18.02.2020 13:52	Папка с файлами	
csv_format_6	18.02.2020 13:52	Папка с файлами	
csv_plot	18.02.2020 13:52	Папка с файлами	
data_analysis	08.06.2021 11:52	Папка с файлами	
duga	24.01.2022 12:10	Папка с файлами	
gan	01.04.2021 12:02	Папка с файлами	
generate_bconds_2018_2022	13.07.2022 13:44	Папка с файлами	
good2	18.02.2020 13:52	Папка с файлами	
Large_Scale_Extract	18.02.2020 13:51	Папка с файлами	
NEMO_automatron	18.02.2020 13:51	Папка с файлами	
NEMO_compress_copy	18.02.2020 13:51	Папка с файлами	
NEMO_compress_copy_1617_crs	18.02.2020 13:52	Папка с файлами	
node	12.05.2021 22:53	Папка с файлами	
old_pytorch_diff_solver	22.07.2021 22:03	Папка с файлами	
pytorch_diff_solver	29.09.2021 11:29	Папка с файлами	
rivers	18.02.2020 13:52	Папка с файлами	
salinity_bcond	18.02.2020 13:53	Папка с файлами	
Salinityevl	18.02.2020 13:52	Папка с файлами	
Scripts_from_storage_110_31	18.02.2020 13:52	Папка с файлами	
selegan	30.03.2021 20:28	Папка с файлами	
simens_patterns_ikmeans	18.02.2020 13:55	Папка с файлами	
simens_synetics	11.09.2020 11:25	Папка с файлами	
solver	09.08.2022 17:09	Папка с файлами	
Swan_stat_calibration	24.11.2020 12:46	Папка с файлами	
tanya_pde_solver	08.06.2021 11:52	Папка с файлами	
variables_csv	18.02.2020 13:58	Папка с файлами	
wolfram_NS_solver	08.11.2021 12:48	Папка с файлами	
b_gen.py	13.07.2022 17:06	Python source file	10 КБ
b_gen.zip	31.07.2018 23:15	Сжатая ZIP-папка	6 715 КБ
b_gen_1.py	01.03.2018 16:20	Python source file	2 КБ
b_genN_5d.py	07.11.2017 12:46	Python source file	8 КБ

Папка Script (принадлежит автору)

# Второй этап — «Гнев»



Первое граничное условие, на которое мы обычно «натываемся» совместная работа. В науке она выглядит как на рисунке.

В идеальном мире:

- Каждая строчка кода прокомментирована, если нужно.
- Код можно переиспользовать для новой, но близкой задачи.
- Можно откатиться к предыдущей версии кода.
- Если необходимо, код с различной добавленной функциональностью лежит в разных ветках (которые потом когда-то объединятся).
- Кодом может пользоваться кто-то кроме вас (есть документация и примеры использования).

В реальном мире:

- Ваш (в большинстве случаев, старший) коллега видит мир по своему, а поменять в коде ничего не может – происходит тысяча писем.
- Вы посылаете версию кода по почте – коллега её правит, вы правите параллельно. В итоге правки не совпадают.
- Иногда вам приходит письмо: «Нужно сделать так же, но для другой задачи» со старой версией кода, а последней версии в папке Script нет.

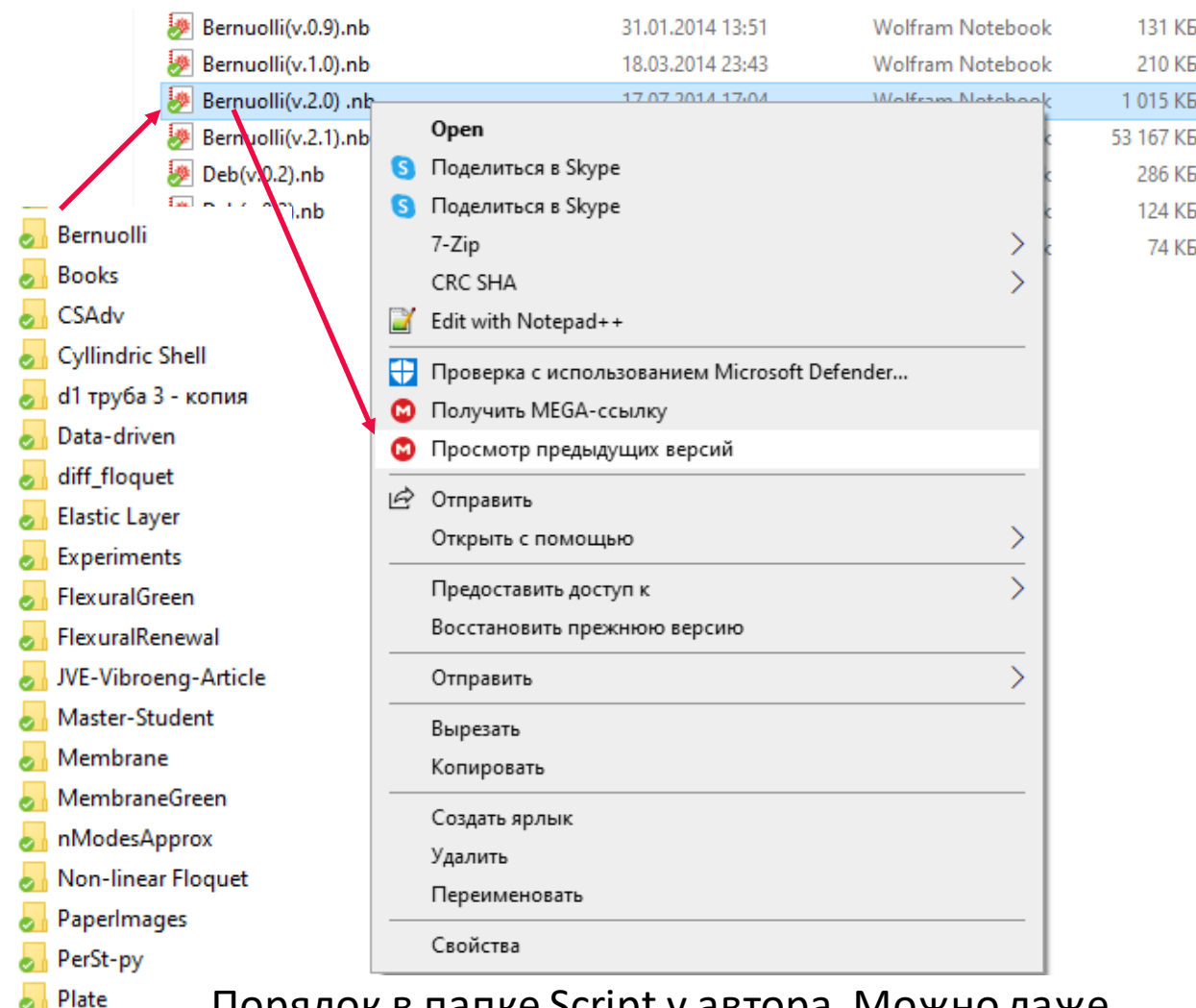
# Третий этап — «Депрессия»

Второе граничное условие – сроки жизни кода без комментариев.

- Можно навести порядок в папке Scripts.
- Можно создавать и помечать необходимые «ветки» кода.
- Современные облачные хранилища (даже бесплатные версии) позволяют работать с историей версий – можно «откатиться», если нужно.
- В целом есть и облачные решения для совместной работы – wolfram cloud, overleaf (для LaTeX), Google colab.

В реальном мире:

- Ваш код теряет свой смысл через короткое время, и в лучшем случае придётся тратить значительное время на восстановления хода мыслей, чтобы построить график с новыми параметрами или хоть как-то изменить код.



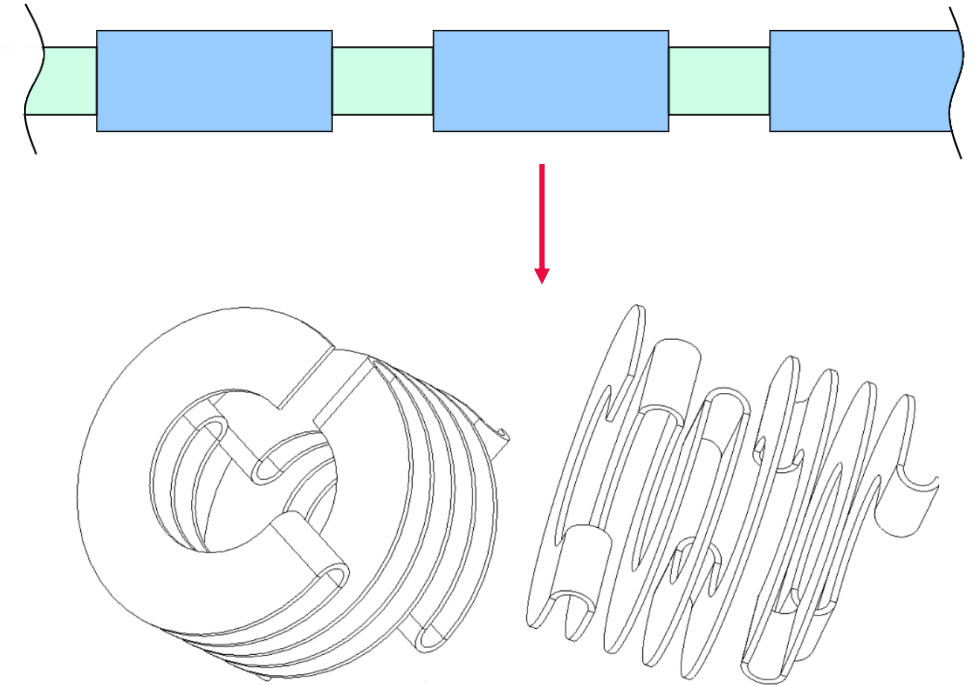
Порядок в папке Script у автора. Можно даже при желании откатиться до старых версий, если всё пошло не по плану.

# Четвёртый этап — «Торговля»

- Всегда возникает соблазн наклепать скриптов и держать всё в голове.
- Всегда возникает соблазн сказать, что это больше не будет использовано.
- Всегда возникает соблазн подождать побольше в colab, а не писать код в виде скриптов, чтобы быстро (или не очень) запускать на сервере.
- Всегда возникает соблазн использовать медленные CAS (computational algebra system), чтобы написать пару строк, проверить гипотезу, и закрыть.

В реальном мире:

- Гипотезы классно проверяются в CAS
- Все соблазны остаться в CAS при переходе от проверки к реальным экспериментам ведут к потере времени на написание «совершенно» нового кода
- Нет контроля, мы ограничены либо тем набором функций, что есть в CAS (или аналогов) или в пакетах-модулях «всё в одном»

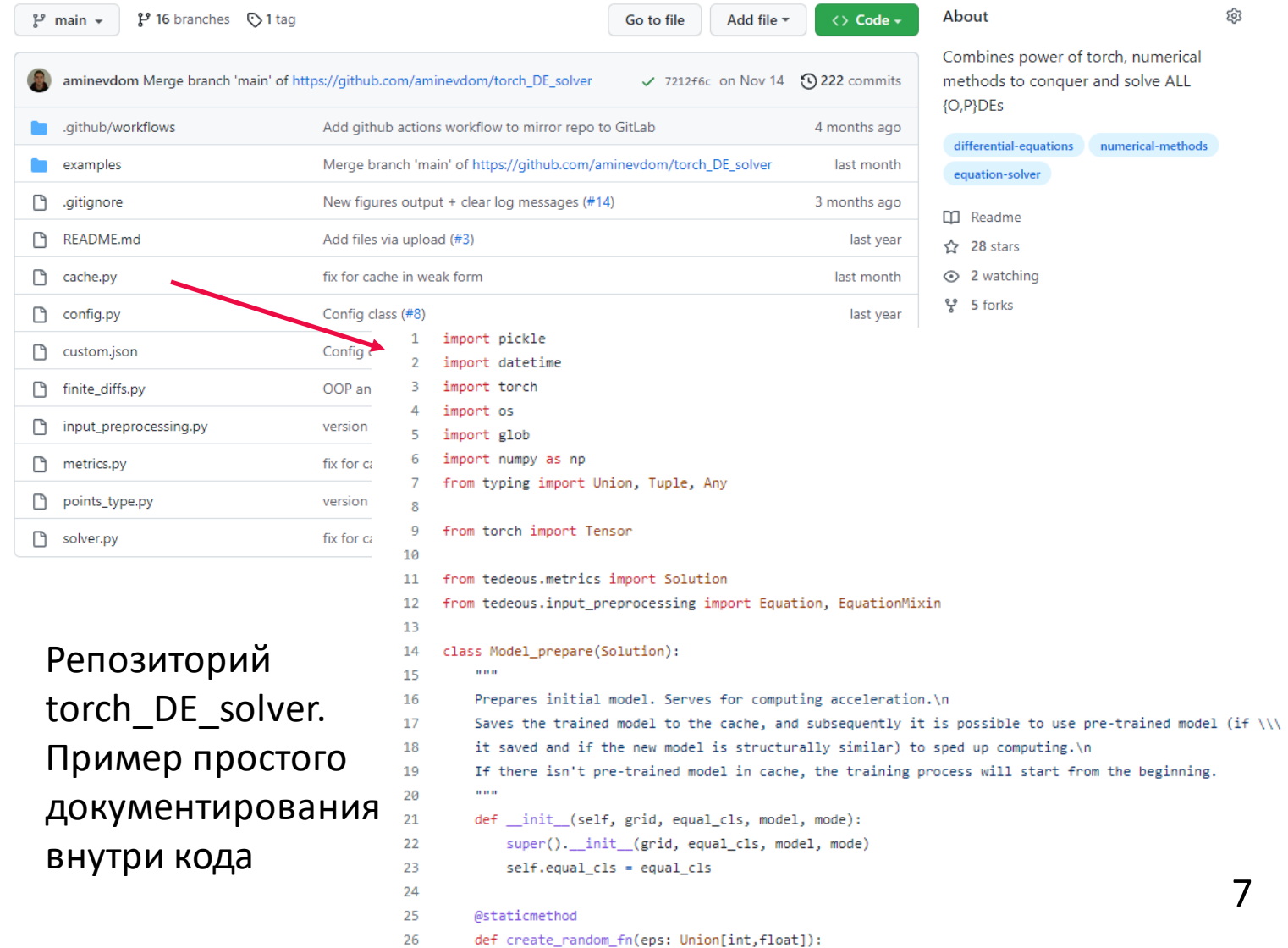


Переход от первой структуры ко второй не сильно меняет теорию (число уравнений становится больше 16 на ячейку вместо 4 на ячейку), но кардинально меняет подход к численным экспериментам.



# Пятый этап — «Смирение»

- Инструменты – большие и малые должны быть разложены по местам.
- Место кода – на репозиториях (git, bitbucket).
- Не обязательно каждый инструмент должен быть монументальным.
- Поделки на один раз тоже можно и нужно держать в отдельном месте, на случай, если они понадобятся снова.
- Потратив немного времени на документацию, можно сэкономить много времени на восстановление алгоритма в памяти при необходимости.



The screenshot displays the GitHub interface for the repository 'aminevdom/torch\_DE\_solver'. The top navigation bar includes links for 'main', '16 branches', and '1 tag', along with buttons for 'Go to file', 'Add file', and 'Code'. The repository description states: 'Combines power of torch, numerical methods to conquer and solve ALL {O,P}DEs'. It also shows tags for 'differential-equations', 'numerical-methods', and 'equation-solver', and statistics: '28 stars', '2 watching', and '5 forks'.

The file list on the left includes: '.github/workflows', 'examples', '.gitignore', 'README.md', 'cache.py', 'config.py', 'custom.json', 'finite\_diffs.py', 'input\_preprocessing.py', 'metrics.py', 'points\_type.py', and 'solver.py'. A red arrow points from 'config.py' to its content on the right.

The content of 'config.py' is as follows:

```
1 import pickle
2 import datetime
3 import torch
4 import os
5 import glob
6 import numpy as np
7 from typing import Union, Tuple, Any
8
9 from torch import Tensor
10
11 from tedious.metrics import Solution
12 from tedious.input_preprocessing import Equation, EquationMixin
13
14 class Model_prepare(Solution):
15     """
16     Prepares initial model. Serves for computing acceleration.\n
17     Saves the trained model to the cache, and subsequently it is possible to use pre-trained model (if \n\n
18     it saved and if the new model is structurally similar) to speed up computing.\n
19     If there isn't pre-trained model in cache, the training process will start from the beginning.
20     """
21     def __init__(self, grid, equal_cls, model, mode):
22         super().__init__(grid, equal_cls, model, mode)
23         self.equal_cls = equal_cls
24
25     @staticmethod
26     def create_random_fn(eps: Union[int, float]):
```

Репозиторий  
torch\_DE\_solver.  
Пример простого  
документирования  
внутри кода

# Вместо заключения

- Открытый код - наука будущего и немного настоящего.
- В бесконечном потоке статей нет времени проверять верность формул, верность вычислений – скоро будут требовать проверять их автоматически.
- Даже самые «экзотические» инструменты интересны, если ими можно воспользоваться.
- Никто не мешает проверять гипотезы «как попало», но проверять гипотезы в рамках одного «фреймворка» приятнее.



Спасибо за внимание!

it's **MO**re than a  
**UNIVERSITY**

The logo for ITMO University, consisting of the letters 'ITMO' in a bold, white, sans-serif font. The background is a dark purple grid with white wavy lines on the right and bottom edges.

**ITMO**

# **Зачем математику репозитории и коммиты?**

Александр Хватов, зав. лаб. (Лаборатория композитного ИИ)

[mailto:alex\\_hvatov@itmo.ru](mailto:alex_hvatov@itmo.ru)

ITMO OS Meetup, 16.12.2022