

# Итеративное улучшение тематической модели с обратной связью от пользователя

Алексей Ильич Горбулев

Московский физико-технический институт

*Курс:* Моя первая научная статья/Группа Б05-021а

*Эксперт:* д. ф.-м. н. К. В. Воронцов

*Консультант:* В. А. Алексеев

2023

# Итеративное улучшение модели

**Мотивация:** большая часть тем может вообще быть не связана с исследованием, который проводит пользователь

**Цель:** улучшить тематическую модель, сохранив ранее найденные релевантные темы, найти новые релевантные темы за счёт других

**Решение:** при обучении новой тематической модели:

- ▶ при инициализации матрицы  $\Phi$  зафиксировать столбцы, соответствующие релевантным темам
- ▶ использовать аддитивную регуляризацию тематической модели (ARTM):
  - ▶ использовать регуляризатор декоррелирования, чтобы уменьшить число дублирования тем

$$R(\Phi) = -\frac{\tau}{2} \sum_{t \in T} \sum_{s \in T \setminus t} \sum_{w \in W} \varphi_{wt} \varphi_{ws}$$

- ▶ использовать регуляризаторы сглаживания и разреживания для улучшения интерпретируемости тем