

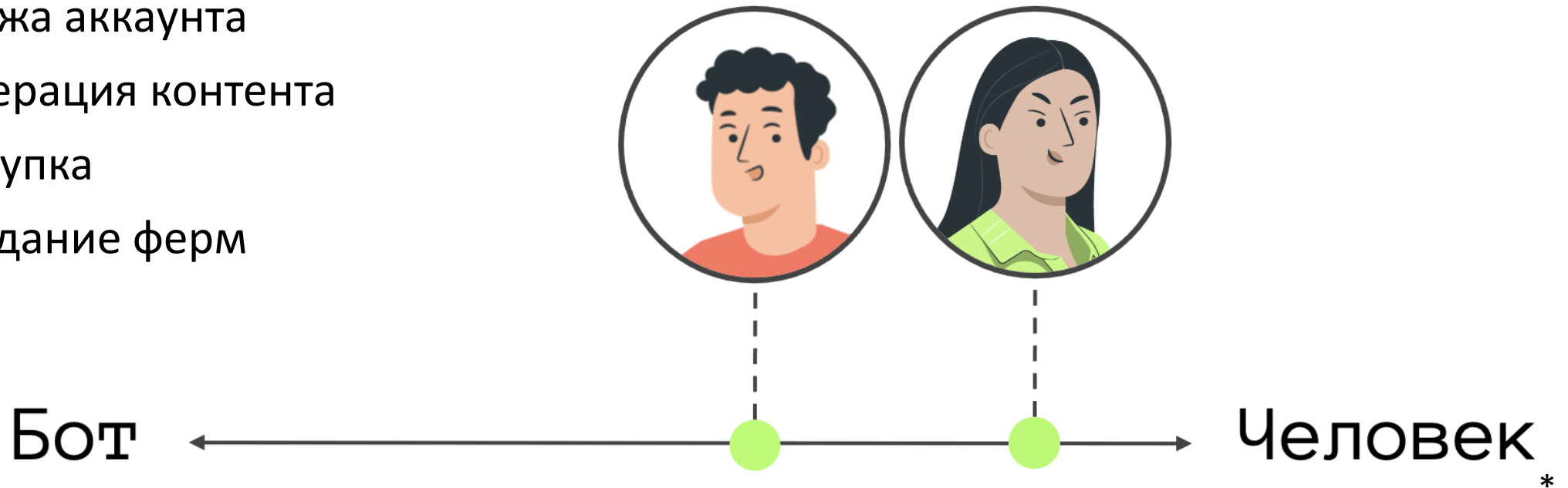
Оценка способ человека к обн ботов в социал

Коломеец Максим
PhD, с.н.с. СПб ФИЦ РАН



Методы эволюции ботов

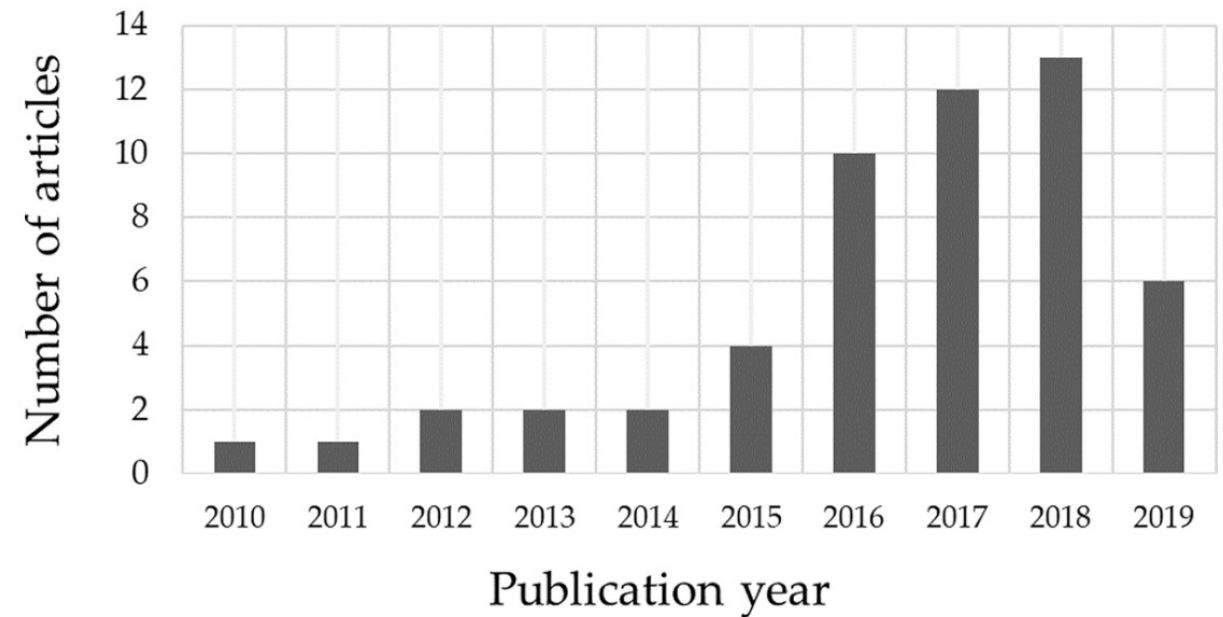
- Кража профиля
- Кража аккаунта
- Генерация контента
- Покупка
- Создание ферм



*Коломеец М.В., “Роботы против людей, эволюция ботов в социальных сетях”
Кампус // Бумага, 2021.

Методы анализа

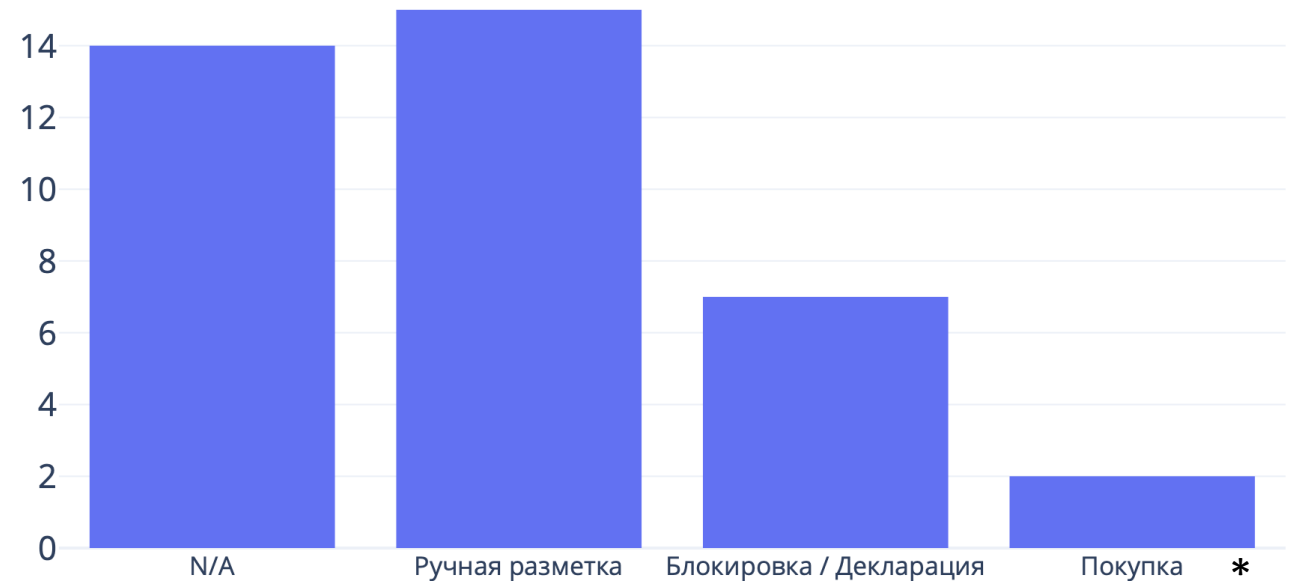
- Экспертный
(Oxford DEMTECH, SratCom, ...)
- Машинное обучение
(Botometr, Debot, ...)
 - С учителем (~80%)
 - Без учителя
 - Частичное



*M. Orabi, D. Mouheb, Z. Al Aghbari, and I. Kamel, "Detection of Bots in Social Media: A Systematic Review," *Inf. Process. Manag.*, vol. 57, no. 4, 2020, doi: 10.1016/j.ipm.2020.102250.

Методы аннотации и их недостатки

- N/A
невозможно верифицировать
- Ручная разметка
субъективно и дорого
- Блокировка / Декларация
эффективность гарантировано
ниже оригинала
- Покупка
сложность создания больших
наборов



*M. Orabi, D. Mouheb, Z. Al Aghbari, and I. Kamel, "Detection of Bots in Social Media: A Systematic Review," *Inf. Process. Manag.*, vol. 57, no. 4, 2020, doi: 10.1016/j.ipm.2020.102250.

Цель

Проверить теоретическое качество аннотации человеком

=

H0 – человек не может определить бота ($P \leq 0.5$)

H1 – человек может определить бота ($P > 0.5$)

*

Методология

Dataset	Type	Company	Count	Descriptions
[REDACTED]				
user_1	activists community		385	Users from group "velosipedization" that is dedicated to development of bicycle transport.
user_2	mass media community		298	Users from group "belteanews" that is dedicated to Belarussian news.
user_3	developers community		332	Users from group "tproger" that is dedicated to software development.
user_4	sport community		224	Users from group "mhl" that is dedicated to youth hockey.
user_5	mass media community		420	Users from group "true_lentach" that is dedicated to Russian news.
user_6	blogger community		251	Users from group "mcelroy_dub" that is dedicated to re-playing of funny videos.
user_7	commerce community		284	Users from group "sevcableport" that is dedicated to creative space.
user_8	festival community		259	Users from group "bigfestava" that is dedicated to cartoon festival.
user_9	sport community		181	Users from group "hcakbars" that is dedicated to fun community of hockey club.
user_10	developers community		397	Users from group "tnull" that is dedicated to software development and memes.

1) Боты = BOTS

2) Смещенные пользователи = SHIFT

3) Случайные пользователи = RAND

4) Верифицированные пользователи = STUDENT

Методология

1) BOTS

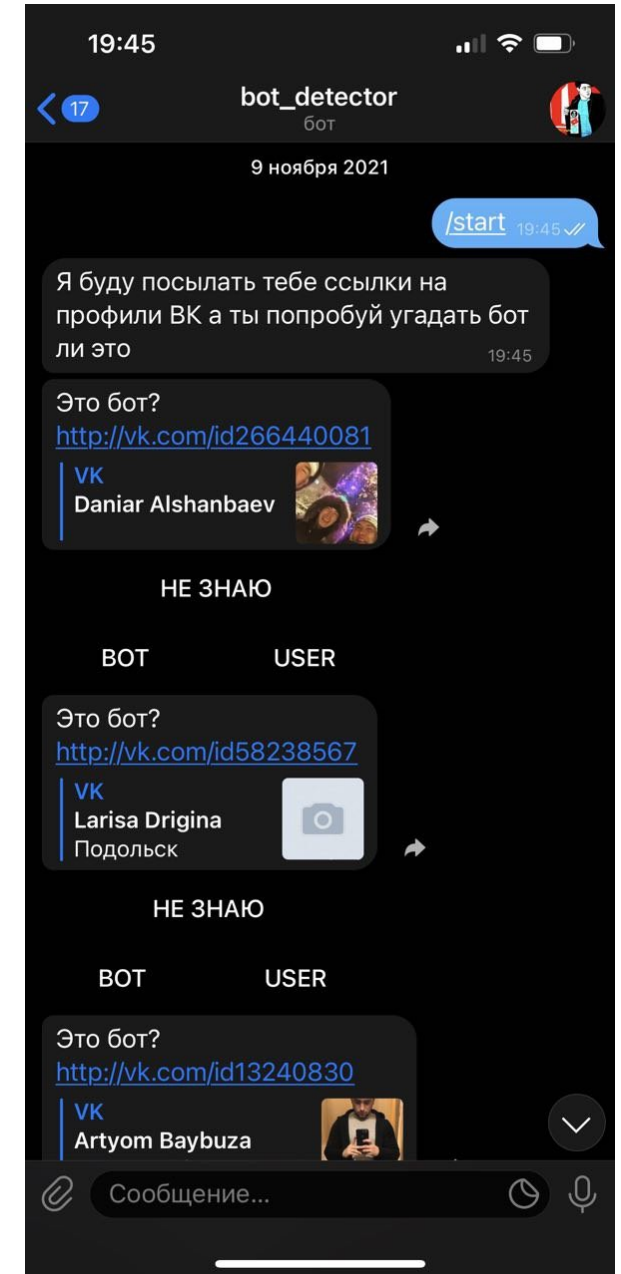
2) BOTS + SHIFT

3) BOTS + RAND

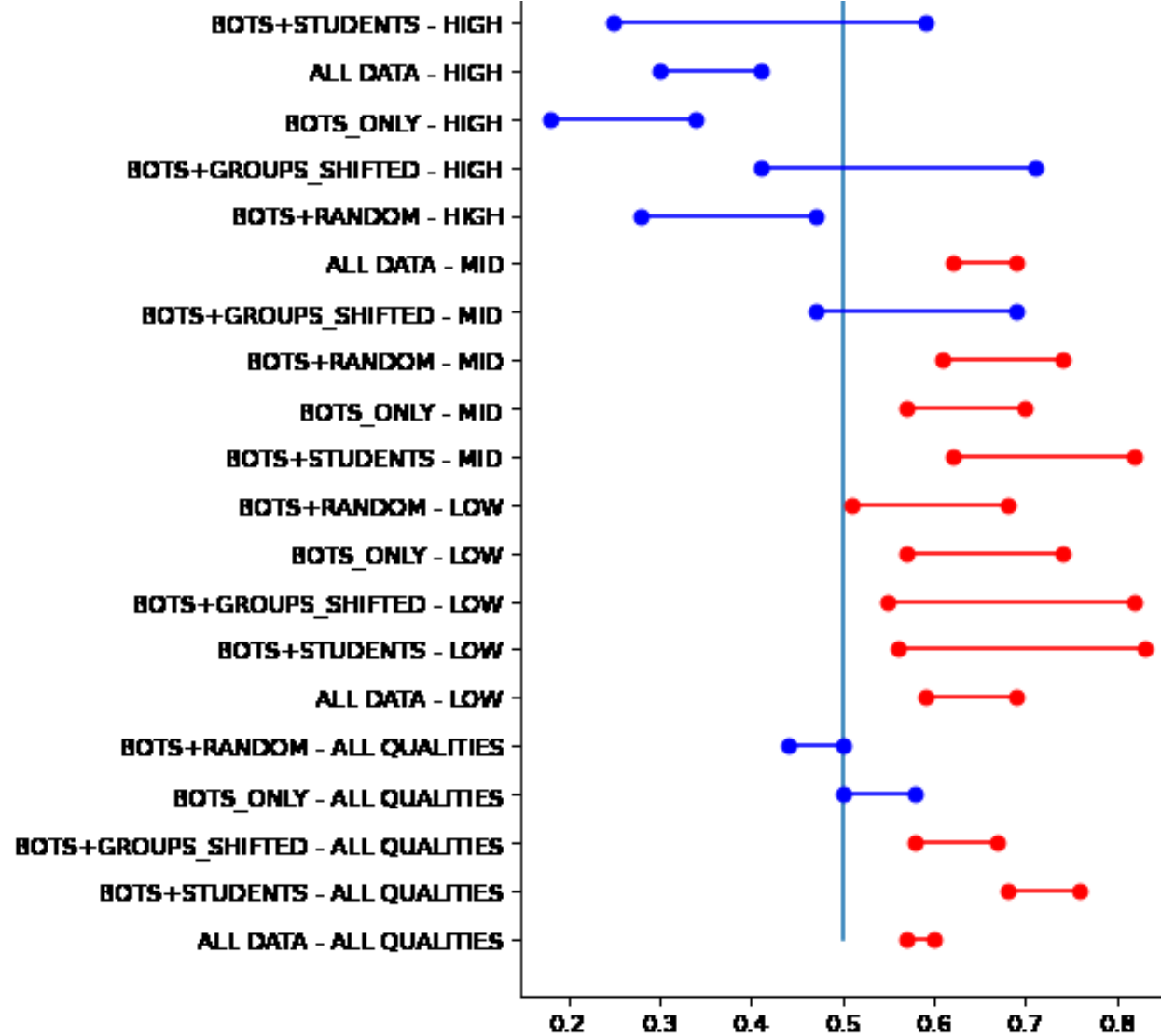
4) BOTS + STUDENT



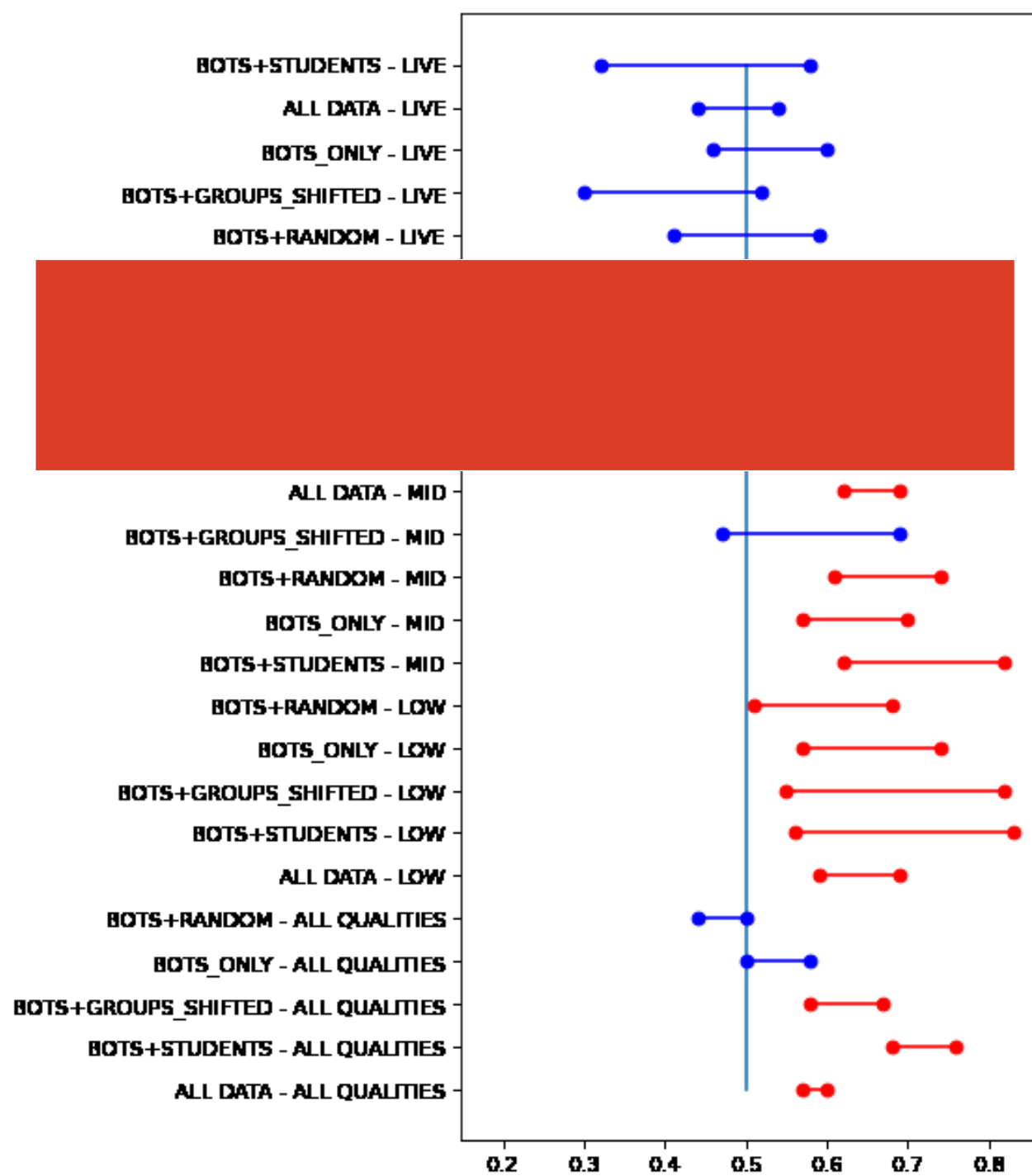
73
аннотатора



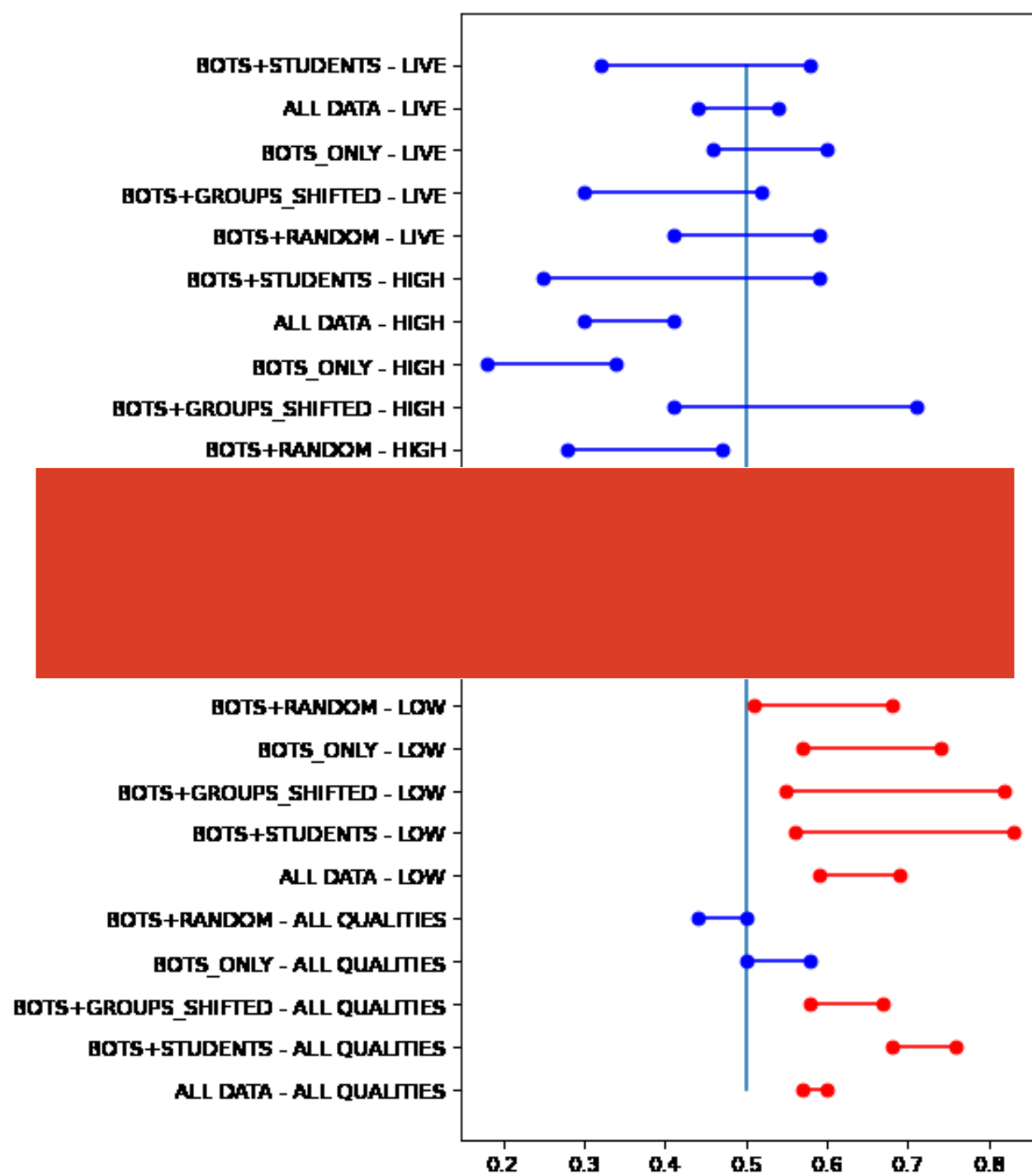
Результаты



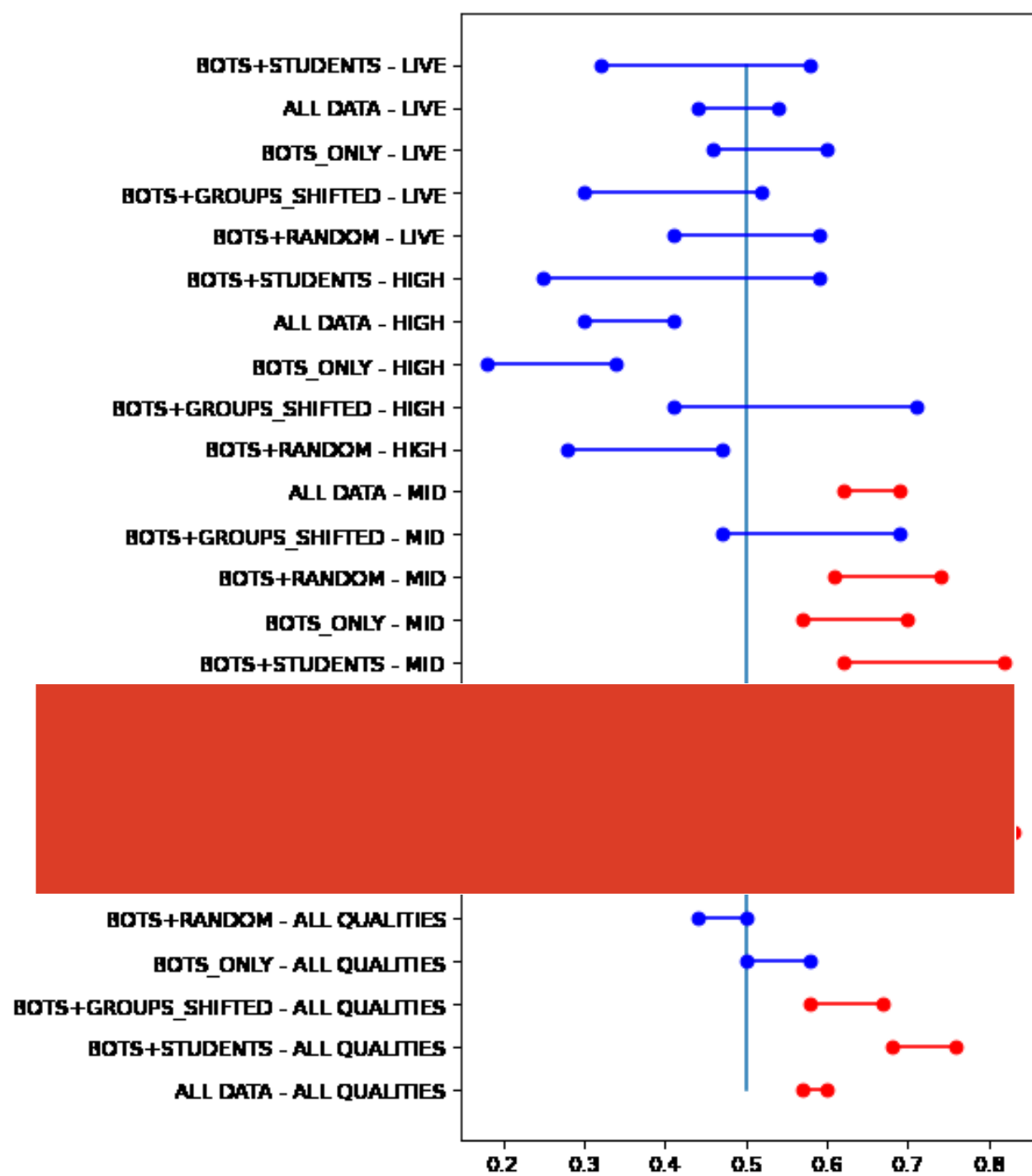
Результаты



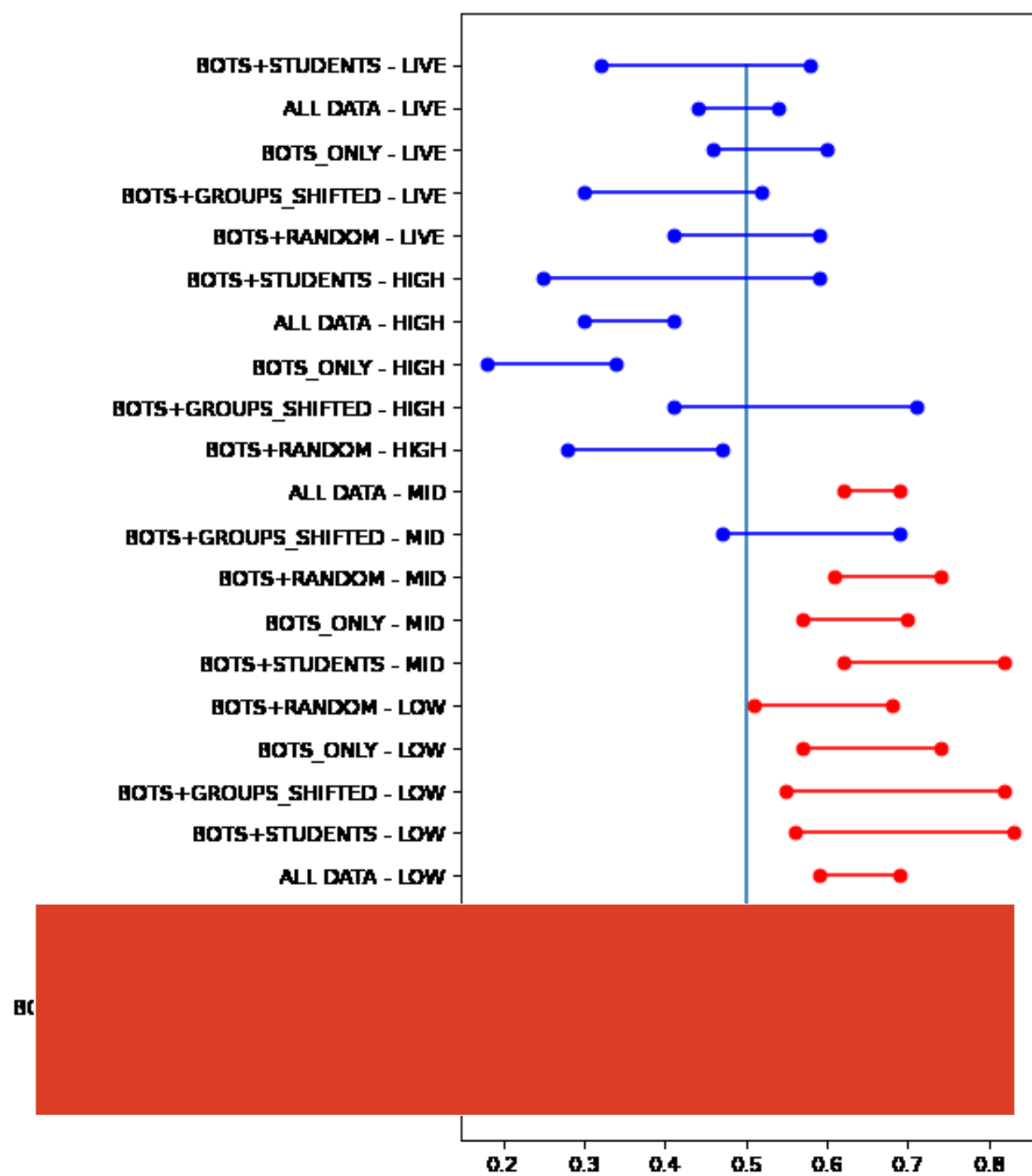
Результаты



Результаты



Результаты



Результаты

В реальности:

$$M = M_{clf} * M_{data}$$

Например:

$$Acc = 0.9_{clf} * 0.9_{data}$$

Представим себе

“идеальную модель”, с

$$Acc = 1_{clf} * X_{data}$$

Набор ботов	Качество	Balanced Accuracy	FPR	FNR
BOTS	LOW	0.901	0	0.099
	MID	0.814		0.186
	HIGH	0.296		0.704
	LIVE	0.623		0.377
BOTS + RAND	LOW	0.578	0.611	0.233
	MID	0.583		0.224
	HIGH	0.402		0.586
	LIVE	0.472		0.445
BOTS + SHIFT	LOW	0.817	0.161	0.205
	MID	0.802		0.235
	HIGH	0.749		0.341
	LIVE	0.662		0.515
BOTS + STUDENT	LOW	0.936	0.075	0.053
	MID	0.856		0.212
	HIGH	0.704		0.517
	LIVE	0.717		0.491

Выводы

- Эффективность классификации систем на основе аннотации ниже декларируемой
- Такие системы вероятно не смогут распознать большинство ботов высокого качества
- Такие системы не применимы для задач глобального мониторинга

Возможности

- Разница между методом “закупки” и “аннотацией” позволяет производить качественный анализ
- Метод “закупки” позволяет производить контроль качества при аннотации (требуется проверка)