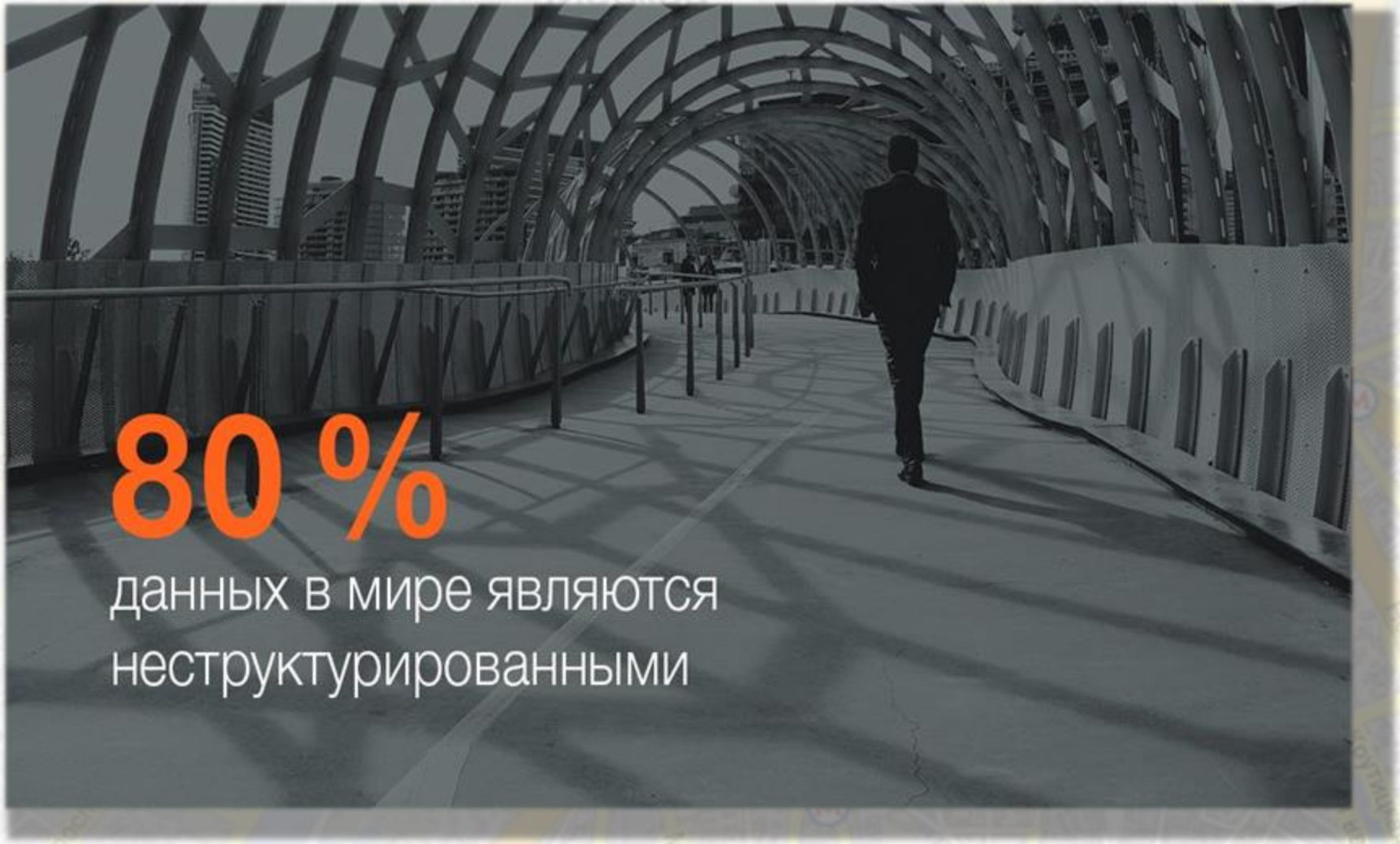
Средства повышения качества данных в задачах машинного обучения

О чем это мы или плохие новости 2 от IBM



Причины возникновения проблем

1.Миграционные потоки.

Большое смешение смысловых и семиотических полей при большом количестве ошибок, т.е. нарушение синтаксиса (грамматик) и семантики текстов

2.Качество образования, сленг

Любая система общего образования дает различное качество подготовки. Это зависит как от самих учащихся, так и школ. В любом обществе всегда существуют социальные группы, имеющие «собственный» язык. Чаще всего, возникает простая синонимия, возникновение новых «смыслов» происходит реже.

1.Различие в фонемных рядах разных языков.

Невозможность предсказывать опечатки, а значит невозможно создать «полную» базу вариантов написаний

2.Гаджетизация

На сегодняшний день сматфоны и планшеты есть у всех. В результате развитой системы подсказок и исправлений текстов возникает новый класс ошибок. Выпадающие слова из контекста.

3.Неоднозначность понятий.

В рамках России это озвученная, например:

- порталом «Государственных услуг» проблема, когда ведомства дают наименования по сути одних и тех же услуг по разному. При этом они подаются в сильно «забюрократизированном», формальном виде или очень длинные названия. Понять нормальному человеку это невозможно.
- Наименования номенклатуры у разных ритейлеров разнится, пример: YANDEX.Market

Свойства сознания

- **любопытство**

стремление разносторонне познать то или иное явление в существенных отношениях. Это качество ума лежит в основе активной познавательной деятельности;

- **глубина ума**

заключается в способности отделять главное от второстепенного, необходимое от случайного;

- **гибкость и подвижность ума**

способность человека широко использовать имеющийся опыт, оперативно исследовать предметы в новых связях и отношениях, преодолевать шаблонность мышления:

- **логичность мышления**

характеризуется строгой последовательностью рассуждений, учетом всех существенных сторон в исследуемом объекте, всех возможных ею взаимосвязей;

- **доказательность мышления**

характеризуется способностью использовать в нужный момент такие факты, закономерности, которые убеждают в правильности суждений и выводов;

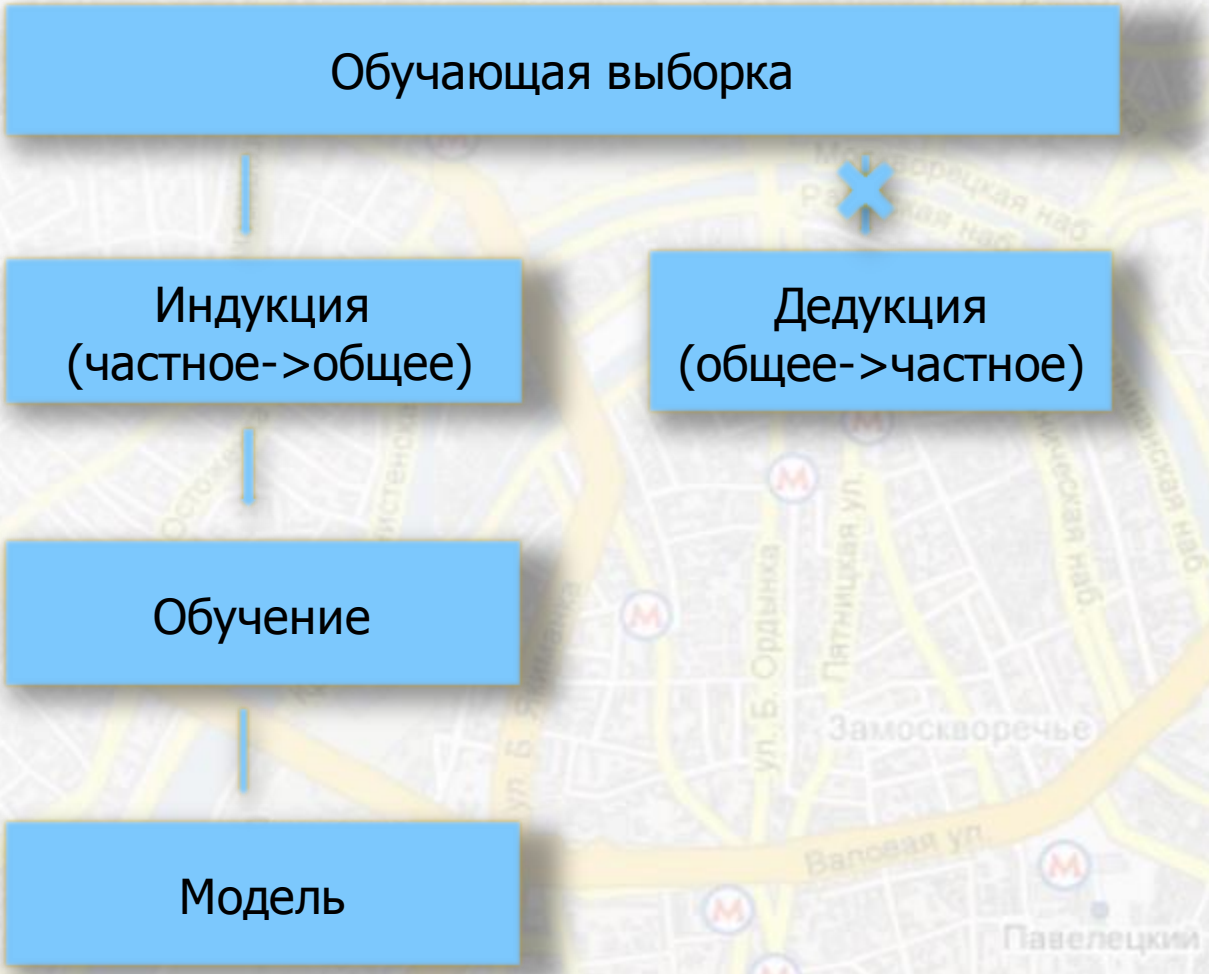
- **критичность мышления**

предполагает умение строго оценивать результаты мыслительной деятельности, подвергать их критической оценке, отбрасывать неправильное решение, отказываться от начатых действий, если они противоречат требованиям задачи:

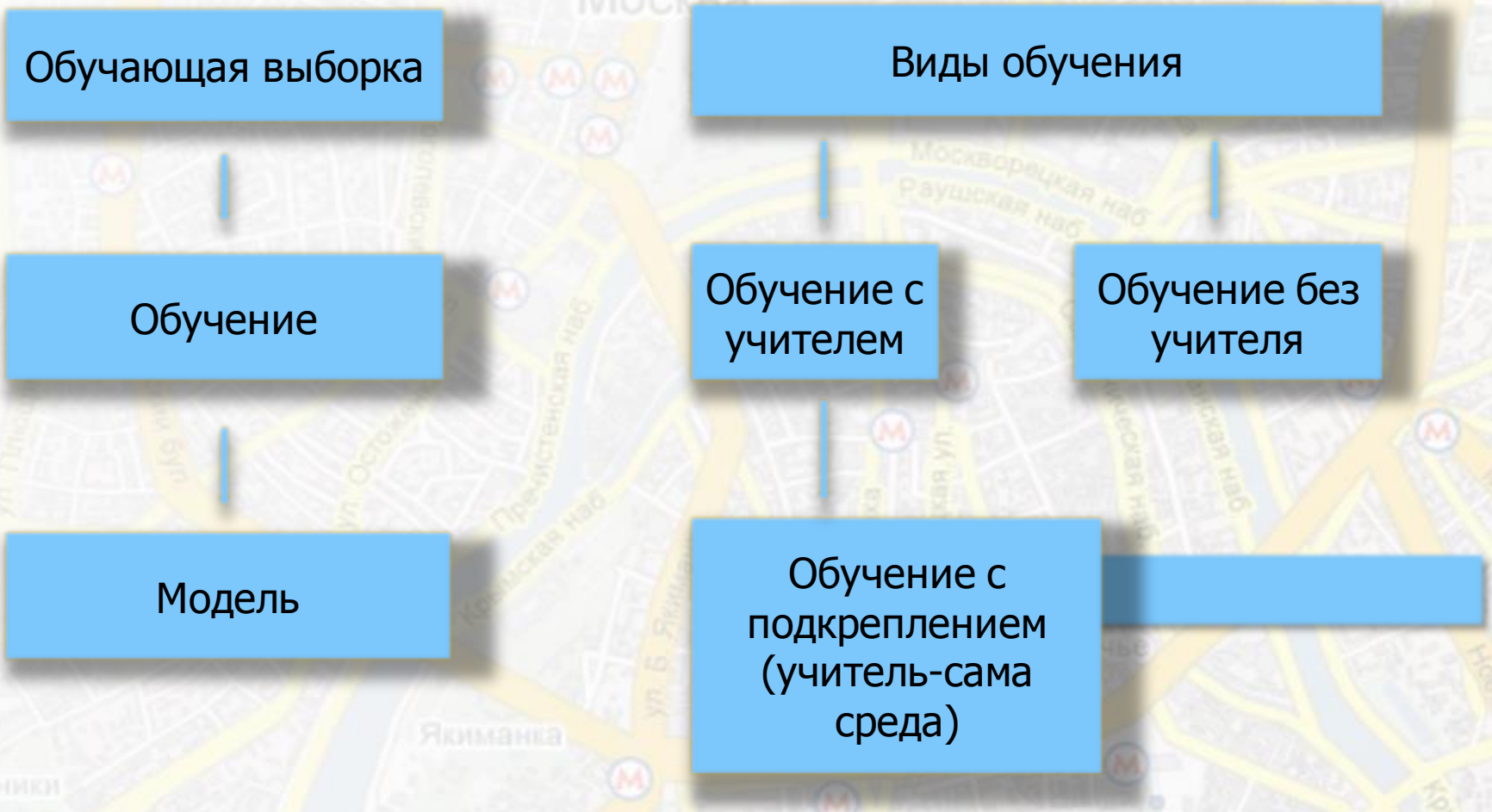
- **широта мышления**

способность охватить вопрос в целом, не теряя из виду исходных данных соответствующей задачи, видеть многовариантность в решении проблемы.

Машинное обучение



Машинное обучение



Проблема «Разметчика»

1.Низкая квалификация.
Невозможно подобрать высококлассных экспертов для примитивной работы
2.Однотонность труда
Снижение качества даже простых операций в течение смены
1.Быстрое выгорание
Рост процессов профорганизации, рост кризисных мосент

Семантический анализ.

Спасибо!

ООО «АйКью Системс» (IQ Systems)

http://www.iqsystems.ru

http://www.iqdq.ru

https://www.facebook.com/IQSystems.ru

pavlyuts@iqsystems.ru

+7(499) 501-79-93