

[КАК СТАТЬ АВТОРОМ](#)[Лучшие ИТ-работодатели России 2023](#) [Добавь свой ...](#)

★ 4.29

Оценка

538.6

Рейтинг

VK

Технологии, которые объединяют



levashove

7 фев в 15:36

## Иерархия ИИ-потребностей

⌚ 6 мин 👁 771

Блог компании VK, Big Data\*, Машинное обучение\*, Искусственный интеллект

[Перевод](#)

Автор оригинала: Monica Rogati



Как это часто бывает с быстро развивающимися технологиями, ИИ породил массовые проявления синдрома упущенных возможностей, страха, неуверенности, сомнений и междоусобиц. Некоторые из них имеют под собой основания, некоторые — нет, но всё это бросается в глаза. Все от мала до велика, от незаметных стартапов до финтех-гигантов и государственных учреждений, сформировали команды специалистов, которые

↑ +12 ↓

📄 12



💬 0

Команда VK Cloud перевела статью о том, как использовать ИИ и машинное обучение, чтобы лучше справляться с тем, что мы делаем.

## Введение

Сплошь и рядом компании **не готовы к ИИ**. Может, они наняли своего первого дата-сайентиста и получили более чем посредственные результаты или грамотность работы с данными находится где-то на периферии их корпоративной культуры. Но самый распространённый сценарий — это когда они ещё не создали инфраструктуру, позволяющую внедрять и пожинать плоды самых простых алгоритмов Data Science и тем паче машинного обучения.

Как консультант по Data Science и ИИ я говорила об этом бесчисленное множество раз, особенно в последние два года. Другие специалисты со мной согласны. Сложно быть душителем всеобщего энтузиазма, особенно если ты сам разделяешь этот энтузиазм. И как сказать компании, что она не готова к искусственному интеллекту, и при этом не показаться эдаким заносчивым самозванцем, вахтёром, который упорно не даёт людям пройти?

Вот объяснение, которое нравится мне больше всего:

Воспринимайте ИИ как вершину пирамиды потребностей. Да, самореализация (ИИ) — это здорово, но сначала вам нужны еда, вода и крыша над головой (инфраструктура, сбор и грамотная работа с данными).

### ИЕРАРХИЯ ПОТРЕБНОСТЕЙ В DATA SCIENCE

ОБУЧЕНИЕ / ОПТИМИЗАЦИЯ

АГРЕГИРОВАНИЕ / РАЗМЕТКА

ИЗУЧЕНИЕ / ПРЕОБРАЗОВАНИЕ

ПЕРЕМЕЩЕНИЕ / ХРАНЕНИЕ

СБОР



*Чтобы эффективно реализовывать проекты на основе ИИ и машинного обучения,*

*нужно сначала построить твёрдый фундамент для данных*

## Базовые потребности: вы умеете считать?

На нижнем ярусе пирамиды расположился **сбор данных**. Какие данные вам нужны, какие у вас есть? Если это продукт для пользователей, регистрируете ли вы все соответствующие взаимодействия с ними? Если это датчик, как и какие данные через него проходят? Насколько легко зафиксировать взаимодействие, для которого ещё нет инструментов? В конце концов, последние достижения в машинном обучении стали возможны именно благодаря правильным датасетам.

Далее: как данные проходят по системе? У вас надёжные потоки данных / ETL? Где вы их храните, легко ли получить к ним доступ и анализировать их? Джей Крепс вот уже лет десять твердит, что надёжный поток — это альфа и омега любых операций с данными. Кстати, я искала точную цитату и нашла её в его шедевре *I love logs*. Потом я заметила, что параграфом выше он приводит точное сравнение с пирамидой Маслоу, добавив для убедительности, что это нечто очевидное (спасибо, Джей!). Если вспоминать другие работы на эту тему, ещё я натолкнулась (спасибо Дэниелу Танкелангу за информацию) на отличный пост Хилари Мэйсон и Криса Виггинса о том, чем занимается дата-сайентист. Несколько дней назад Шон Тейлор показал собственную пирамиду потребностей в Data Science (иронично названную «разъединённый треугольник Data Science»). Конечно, в ней другие потребности. Может, нам пора завести микроблог в Tumblr?

Данные можно **исследовать и преобразовывать**, только когда они доступны. Сюда относится и пресловутая «очистка данных», недооценённая сторона Data Science, которой будет посвящена отдельная статья. Это когда оказывается, что вам не хватает многих данных, ваши датчики выдают ненадёжную информацию, из-за изменения версии у вас пропали события, вы неправильно интерпретировали флаг — и вы снова спускаетесь к основанию пирамиды, чтобы она не рухнула.

Когда вы способны достоверно исследовать и очистить данные, можно переходить к тому, что традиционно считают BI или **аналитикой**: определить метрики, которые нужно отслеживать, их сезонность и чувствительность к разным факторам. Может, приблизительно сегментировать пользователей и посмотреть, не выскочит ли что-нибудь. Но поскольку вы настроены на ИИ, теперь вы создаёте, что называется, **признаки** для модели машинного обучения. На этом этапе вы уже знаете, что хотели бы предсказывать или обучать, так что можно приступить к подготовке **датасета для обучения**, генерируя метки автоматически (какие клиенты ушли?) или с участием человека.

## Так, считать умеем. А дальше что?

У нас есть обучающие данные — так что, теперь можно заняться машинным обучением? Возможно, да, если вы собираетесь у себя в компании прогнозировать уход клиентов; нет, если вы собираетесь предлагать результат этого машинного обучения заказчикам вне компании. Нам нужен пусть даже примитивный фреймворк для A/B-тестирования или **экспериментов**, чтобы можно было выполнять деплоймент шаг за шагом: так мы сможем избежать серьёзных сбоев и примерно оценить последствия изменений до того, как они на всех повлияют. В это же время стоит внедрить самые **простые базовые показатели** (для рекомендательных систем это «самый популярный», потом «самый популярный для вашего сегмента пользователей» — такой подход «сначала стереотип, потом персонализация» очень раздражает, но срабатывает эффективно).

Простую эвристику на удивление чрезвычайно трудно превзойти, ведь с её помощью можно полностью отладить систему без мистических чёрных ящиков машинного обучения с гипернастроенными гиперпараметрами внутри. И именно поэтому мой любимый алгоритм Data Science — это деление.

На этом этапе можно развернуть очень простой ML-алгоритм (вроде логистической регрессии или того же деления), потом подумать о новых сигналах и признаках, которые могут повлиять на результаты. Данные о погоде и переписи населения — мои любимые помощники. И нет, каким бы мощным ни было глубокое обучение, оно не сделает это за вас автоматически. Появление новых сигналов (создание, а не конструирование признаков) поможет в разы повысить эффективность. Здесь имеет смысл притормозить, хотя как дата-сайентисты мы с нетерпением рвёмся на следующую ступеньку пирамиды.

## Вот он, искусственный интеллект!

Вы всё сделали. Подготовили инструменты. Ваш ETL работает как часы. У вас есть организованные и очищенные данные, дашборды, метки и отличные признаки. Вы измеряете то, что надо. Вы можете экспериментировать каждый день. У вас отлаженный со всех сторон базовый алгоритм, который работает в продакшен-среде, и вы меняли его с десятков раз. Вы готовы. Вперёд, к вершинам.

Попробуйте всё, от собственных экспериментов до сотрудничества с компаниями, которые специализируются на машинном обучении. Может, вы добьётесь существенного улучшения производительности. А может, и нет. В худшем случае вы просто освоите новые методы, подготовите заключения и приобретёте опыт и с чистой совестью сможете рассказывать инвесторам и клиентам о своей работе с искусственным интеллектом. В лучшем случае вы кардинально измените взаимодействие с пользователями и клиентами и работу самой компании — у вас получится истинная история успеха в сфере машинного

обучения.

## Минуточку, а как же MVP, Agile, Lean и вот это всё?

**Иерархия потребностей в Data Science — это не повод целый год строить разрозненную, чрезвычайно усложнённую инфраструктуру.** Когда вы создаёте традиционный минимально жизнеспособный продукт, вы начинаете с малого вертикального разреза продукта и добиваетесь, чтобы всё в нём работало. Можно построить пирамиду продукта и наращивать её в горизонтальной плоскости. Например, у себя в Jawbone мы начали с данных о сне и построили пирамиду: инструменты, ETL, очистка и организация, фиксация и определение меток, метрики (сколько часов в среднем люди спят каждую ночь? А сколько времени они дремлют? А что значит «вздремнуть»?), межсегментный анализ вплоть до создания историй, основанных на данных, и Data Products на базе машинного обучения (автоматическая регистрация сна). Потом мы занялись шагами, едой, погодой, тренировками, потом добавили социальные сети и коммуникацию, раз за разом, одно за другим. Мы не строили всеобъемлющую инфраструктуру, даже не попытавшись запустить её в работу комплексно.

## Потенциал инструментов машинного обучения

«Минуточку, а как же этот Amazon API, или TensorFlow, или ещё какая-нибудь Open-Source-библиотека? А как же компании, которые продают ML-инструменты или автоматически извлекают аналитические данные и признаки?»

Всё это чудесно и очень полезно. (Некоторые компании в конечном счёте изо всех сил рвутся соорудить пирамиду целиком, чтобы было что показать миру. Это просто герои.) Но, подпав под влияние шумихи вокруг ИИ, люди не брезгуют неочищенными и неполными данными, на изменение формата и значения которых уйдут годы, которые ещё не осмыслены, которые структурированы так, что становятся бессмысленными, — и ожидают, что эти инструменты сотворят с такими данными чудо. Может быть, когда-то это чудо и случится; я вижу работу в этом направлении и аплодирую этим усилиям. А пока что лучше возвести твёрдый фундамент для пирамиды ИИ-потребностей.

Попробуйте ML Platform от VK Cloud — она помогает построить процесс работы с ML-моделями от дизайна до деплоя, контролировать качество экспериментов и моделей. Для тестирования мы начисляем новым пользователям 3000 бонусных рублей и будем рады вашей обратной связи.

Stay tuned

Присоединяйтесь к телеграм-каналу «Данные на стероидах». В нем вы найдете все об инструментах и подходах к извлечению максимальной пользы из работы с данными: регулярные дайджесты, полезные статьи, а также анонсы конференций и вебинаров.

**Теги:** vk cloud, ИИ, ML, Data Science

**Хабы:** Блог компании VK, Big Data, Машинное обучение, Искусственный интеллект

Если эта публикация вас вдохновила и вы хотите поддержать автора — не стесняйтесь нажать на кнопку

Задонатить

## Редакторский дайджест



Присылаем лучшие статьи раз в месяц

Электронпочта



VK

Технологии, которые объединяют

[Twitter](#) [ВКонтакте](#) [Telegram](#)



24

4.8

Карма

Рейтинг

**Евгений Левашов** @levashove

Лид-редактор, контент-менеджер, техпис, ИТ-блогер

[Сайт](#) [Telegram](#)

 [Комментировать](#)

## Публикации

**ЛУЧШИЕ ЗА СУТКИ**

**ПОХОЖИЕ**

**Bright\_Translate**

2 часа назад

## На что способен самодельный очиститель воздуха, который можно собрать за 30 секунд?

**Простой**

7 мин



2.8K

Обзор

Перевод

**+31**

21



6

**MaFrance351**

7 часов назад

## Оживляем автоинформатор из подмосковного автобуса

**Простой**

9 мин



2.2K

Обзор

**+29**

9



18

**Van4iniy**

19 часов назад

## ChromeOS: почему я отказал своей мечте



7 мин



14K

Из песочницы

**+27**

28



37

**oldadmin**

6 часов назад

## Будни техпода. Ошибки при подключении по RDP

**Простой**

8 мин



2.9K

Тutorial

**+23**

23



8

**it\_union**

9 часов назад

## Внимание. Ведется аудиоконтроль



2 мин



2.6K

 +20 9 40

SLY\_G

20 часов назад

## Раскрыта причина, по которой насекомые кружат вокруг огней по ночам

 5 мин 5.8K

Перевод

 +18 23 14

LukaSafonov

21 час назад

## Web Application and API Protection (WAAP): эволюция WAF (Web Application Firewall)

 Средний 5 мин 1.6K

Аналитика

 +18 21 2

AlexanderKas1505

4 часа назад

## Синдром айтишника 2

 Простой 5 мин 5.5K

Из песочницы

 +15 23 45

PatientZero

8 часов назад

## Поисковый движок в 80 строках Python

 11 мин 2.2K

Тutorial

Перевод

 +15 44 0

Ilya12c

5 часов назад



# Роман Тезиков про CV-проекты и промт-инжиниринг как базовый навык каждого человека

🕒 7 мин

👁 461

Интервью

💎 +12

🔖 17

💬 1

Показать еще

ВАКАНСИИ КОМПАНИИ «VK»

- Pre-sale Data Platform Architect в VK Cloud  
VK · Москва · Можно удаленно
- MLOps (команда VK Видео)  
VK · Москва
- VK Карты, C++ разработчик графического движка в VK Карты, удаленно  
VK · Можно удаленно
- Дизайнер продукта (VK маркет)  
VK · Москва
- Go Developer (VK CXhub)  
VK · Можно удаленно
- Больше вакансий на Хабр Карьере

ИНФОРМАЦИЯ

Сайт	team.vk.company
Дата регистрации	9 августа 2008
Дата основания	15 октября 1998
Численность	свыше 10 000 человек
Местоположение	Россия
Представитель	Руслан Дзасохов

ВИДЖЕТ



Приглашаем на митап RuStore  
«Монетизация приложений»

#### ССЫЛКИ

---

VK Team — официальный youtube-канал  
[www.youtube.com](https://www.youtube.com)

VK — технологии, которые объединяют  
vk.team

all cups — чемпионаты для программистов  
cups.online

Tarantool — middleware for data  
www.tarantool.io

VK Cloud — платформа бизнес-класса для компаний, которые строят ИТ-решения в облаке  
mcs.mail.ru

## ПРИЛОЖЕНИЯ

---



### ВКонтакте: музыка, видео, чаты

ВКонтакте объединяет десятки миллионов людей и даёт безграничные возможности для общения, знакомств, развлечений, бизнеса и обмена новостями из любой точки мира.

iOS Android



### Почта Mail.ru

Крупнейший в русскоязычном сегменте сети сервис бесплатной электронной почты.

iOS Android



### Облако Mail.ru

Бесплатный сервис для хранения фотографий, видео, музыки, презентаций, документов и других файлов.

iOS Android



### Маруся

Дружелюбный голосовой помощник.

iOS Android

## БЛОГ НА ХАБРЕ

---

7 фев в 15:36

### Иерархия ИИ-потребностей



771



0

6 фев в 15:23

### Аналитика в игре на Godot: подключаем MyTracker



1.2K



0

5 фев в 14:29

Тестирование менеджера транзакций

 2.2K  1

2 фев в 13:18

Проверка продуктовых гипотез и немодерируемые UX-исследования с применением внутренних инструментов: кейс команды ОК

 598  0

1 фев в 16:52

Figma to Frontend: как мы автоматически синхронизируем дизайн и код

 14K  7

Ваш аккаунт

Войти  
Регистрация

Разделы

Статьи  
Новости  
Хабы  
Компании  
Авторы  
Песочница

Информация

Устройство сайта  
Для авторов  
Для компаний  
Документы  
Соглашение  
Конфиденциальность

Услуги

Корпоративный блог  
Медийная реклама  
Нативные проекты  
Образовательные  
программы  
Стартапам



Настройка языка

Техническая поддержка

© 2006–2024, Habr