

Преподавание фундаментальных основ искусственного интеллекта

Кафедра теоретической информатики

Концепция преподавания основ ИИ

- В настоящее время мир достиг поворотного момента в развитии методов и средств искусственного интеллекта (ИИ), когда создаваемые технологии перешли из области теории и науки в «реальный мир», т.е. стали доступны практически всем отраслям экономики.
- Этот сдвиг является весьма эффективным, но, в то же время, непростым, поскольку сочетает в себе сложности применения ИИ со сложностями организации человеческой деятельности.

Концепция преподавания основ ИИ

- **Основа знания:** «**Наука о данных**» – наука о представлении данных, их хранении, передаче и анализе, как концептуальной основе моделей жизненных процессов.
- **Цель знания:** осмысление и представление жизненных процессов в научной, цифровой форме, ориентированной на дальнейший анализ и управление процессами, как в природных, так и в созданных людьми системах развитыми средствами ИИ.
- **Задача знания:** успешное решение различных задач человеческой деятельности методами и средствами искусственного интеллекта, базирующимися на достижениях «Науки о данных» как на фундаментальной основе.
- **Носители знания:** ученые по данным, специалисты в области ИИ.

Очень краткая история «науки о данных» (Data Science)



1962 John W. Tukey writes in “The Future of Data Analysis”

- «Data analysis, and the parts of statistics which adhere to it, must...take on the characteristics of science **rather than those of mathematics... *data analysis is intrinsically an empirical science***»



- В 1947 году Тьюки придумал термин «бит», который Клод Шеннон использовал в своей статье 1948 года «Математическая теория коммуникаций».
- В 1965 году в статье с J W Cooley, опубликованной в Mathematics of Computation, Тьюки ввел алгоритм быстрого преобразования Фурье.

1974 Peter Naur опубликовал *Краткий обзор компьютерных методов*

(Concise Survey of Computer Methods)

- «Даталогия, **наука о данных** и об их обработке, и её место в образовании»
- Наур предлагает следующее определение науки о данных: «Наука именно об обработке полученных данных, в то время как их интерпретация делегирована другим областям знания и наукам» (“The science of dealing with data, once they have been established, while the relation of the data

what they represent is delegated to other fields sciences”).

- ALGOL 58 / 60 / 68,
- Форма Бэкуса — Наура (<определяемый символ> ::= <посл.1> | <посл.2> | ... | <посл.n>).



Очень краткая история «науки о данных» (Data Science)

- 1977 The International Association for Statistical Computing (IASC) is established as a Section of the ISI (Institute for Scientific Information).
- 1989 Gregory Piatetsky-Shapiro organizes and chairs the first Knowledge Discovery in Databases (KDD) workshop.
- 1994 BusinessWeek publishes a cover story on “Database Marketing”
- 1996 Members of the International Federation of Classification Societies (IFCS) meet in Kobe, Japan, for their conference “**Data science**, classification, and related methods”.
- 1996 Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth publish “From Data Mining to Knowledge Discovery in Databases.”
- ...
 - 2002 Launch of Data Science Journal
- ...



Цифровая эра (digital era) превращается в эру интеллекта (intelligence era)



Mainframe & PCs
1960s – 1980s



Client Server & Internet
1990s – 2000s



Cloud, Mobile, & Big Data
2000s – 2010s



Intelligent Technologies
2010s – 2020s

ENABLING TECHNOLOGIES

- Transistors & silicon revolution
- Large scale mainframe computing adoption
- Emergence of PCs
- Plant floor automation

- Widespread PC adoption
- Broadband Internet
- ERP and business process technologies

- Mobile & smartphone ubiquity
- Cloud computing
- Social networks
- Big Data

- Machine learning & Artificial Intelligence
- IoT & distributed computing
- Blockchain

CUSTOMER VALUE CREATION

Industrial Automation

Business Process Automation

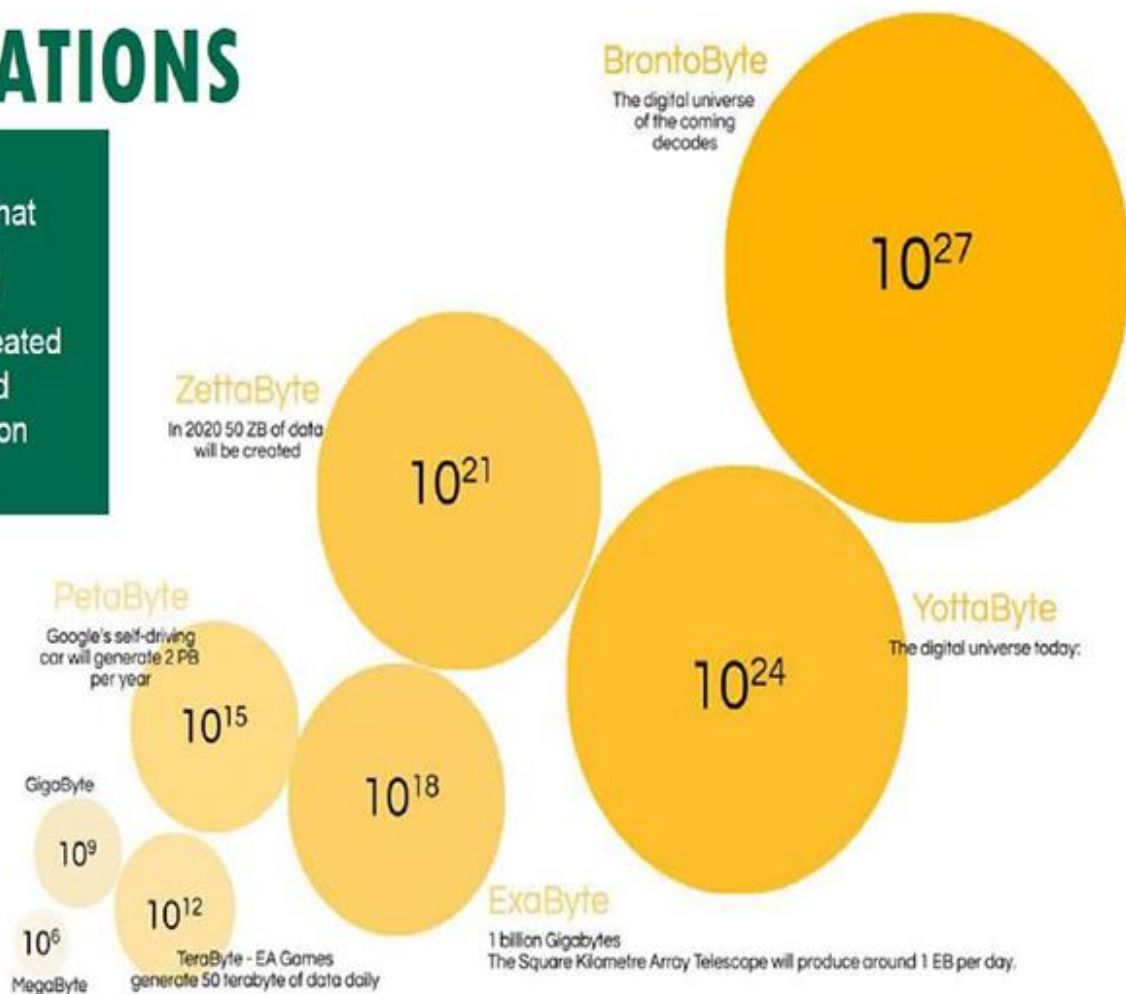
Digital Transformation

Intelligent Enterprise

Intelligence era: рост объема информации

OBSERVATIONS

By 2020,
it's estimated that
1.7MB
of data will be created
every second
for every person
on earth



Согласно исследованиям, человеческий мозг может хранить около 2,5 петабайт данных.

Intelligence era: эра зеттабайт (zettabyte)

- один зеттабайт = один триллион гигабайт



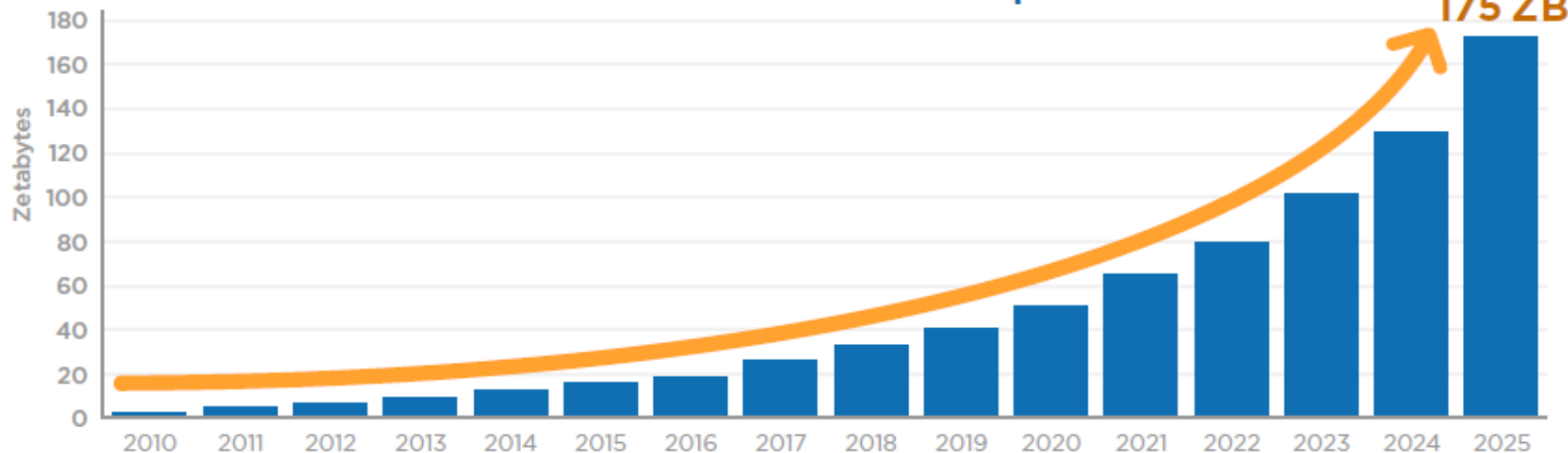
The Digitization of the World
From Edge to Core

David Decker - John Carlo - John Dylung
November 2018

In IDC White Paper #20180218, sponsored by SEAGATE
IDC

Figure 1 - Annual Size of the Global Datasphere

Annual Size of the Global Datasphere



Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

- *В 2020 году было создано или реплицировано 64,2 ЗБ данных, что позволило противостоять давлению, вызванному пандемией Covid-19, во многих отраслях, и его влияние будет ощущаться в течение нескольких лет»*
- Согласно пересмотренной модели, среднегодовой темп роста создания и репликации глобальных данных составит **23%** в течение прогнозируемого периода 2020–2025 годов .

Intelligence era:

статистика больших данных за 2020 год

- 90% всех данных было создано за последние два года;
- в Google ежедневно выполняется более 3,5 миллиардов запросов (это означает 1,2 триллиона запросов в год и более 40 000 поисковых запросов в секунду) ;
- пользователи WhatsApp ежедневно обмениваются приблизительно 65 миллиардами сообщений ;
- интернет-пользователи генерируют около 2,5 квинтиллионов байтов данных (0.0025 зеттабайт) каждый день;
- 80-90% данных, которые мы генерируем сегодня, неструктурированы.

Intelligence era: эра зеттабайт (zettabyte)

- «Объем цифровых данных, созданных в течение следующих пяти лет, более чем в два раза превысит объем данных, созданных с момента появления цифровых хранилищ».

<https://www.idc.com/getdoc.jsp?containerId=prUS47560321>

Intelligence era: цифровая трансформация

- Цифровая трансформация — это процесс интеграции цифровых технологий во все аспекты бизнес-деятельности, требующий внесения коренных изменений в технологии, культуру, операции и принципы создания новых продуктов и услуг.

Intelligence era: цифровая экономика



ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ

РАСПОРЯЖЕНИЕ

от 28 июля 2017 г. № 1632-р

МОСКВА

Утвердить прилагаемую программу "Цифровая экономика Российской Федерации".

Цифровая экономика и большие данные

- В последние годы в мировом сообществе о больших наборах данных сложилось представление как о наборах данных, характеризующихся следующими основными особенностями:
- объемом (Volume);
- скоростью обновления (Velocity);
- разнообразием и неоднородностью (Variety);
- проблемами с достоверностью (Veracity);
- стоимостью обработки (Value);
- изменчивостью (Variability);
- потребностью в визуализации (Visualization).

3V (Gartner 2001):

Velocity (скорость обновления),

Variety (разнообразие и неоднородность)

Volume (объем)

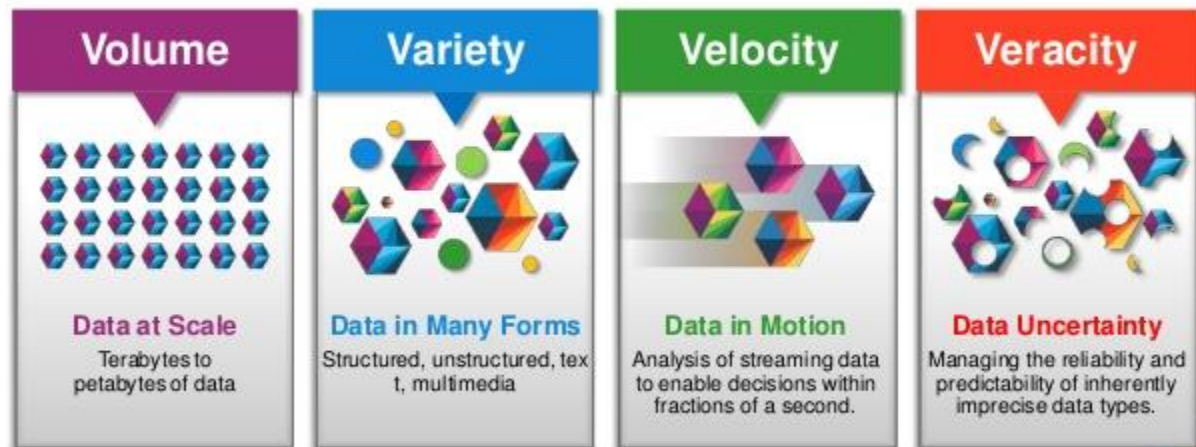


4V (IBM 2012):

Velocity, Variety and Volume +
Veracity (достоверность)

The Era of Big Data Demands Confidence

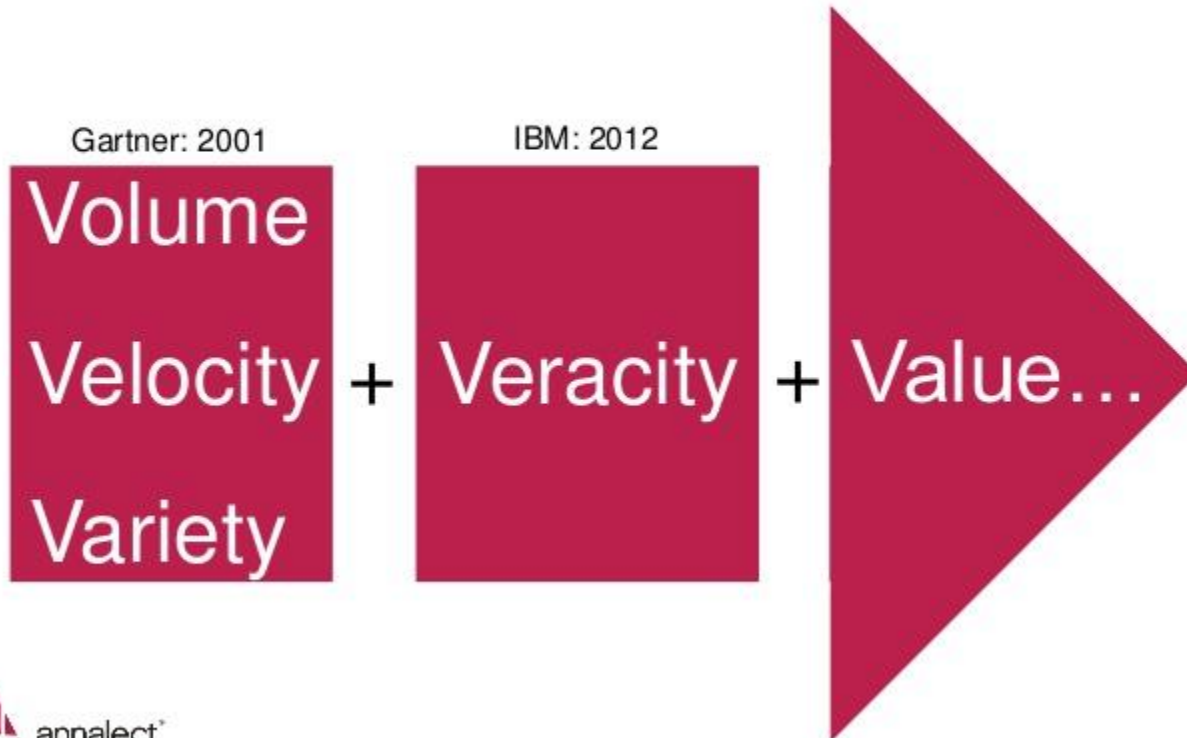
IBM



5V:

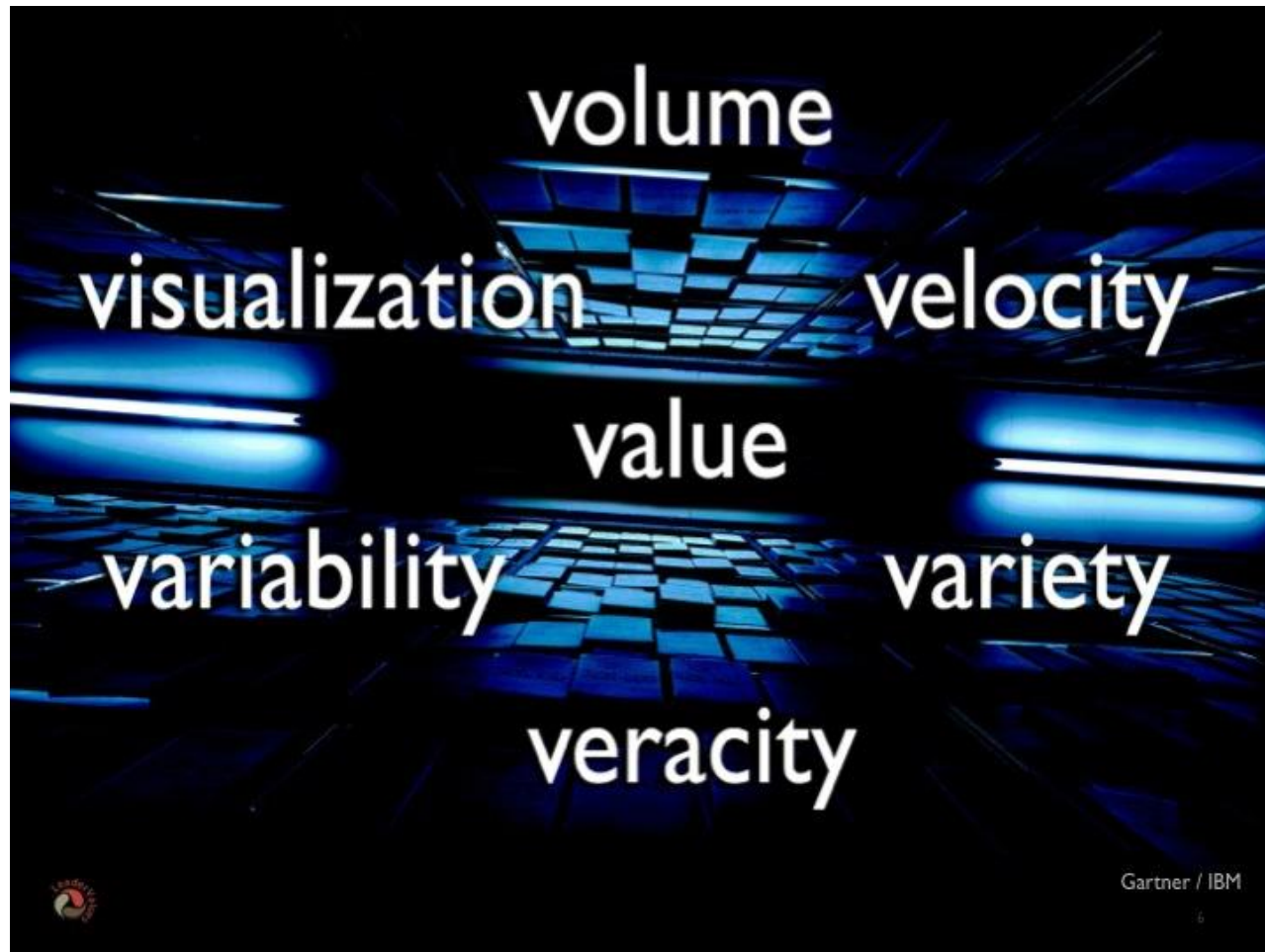
Velocity, Variety, Volume and Veracity + Value (стоимость)

More V's - even the definition has expanded



7V:

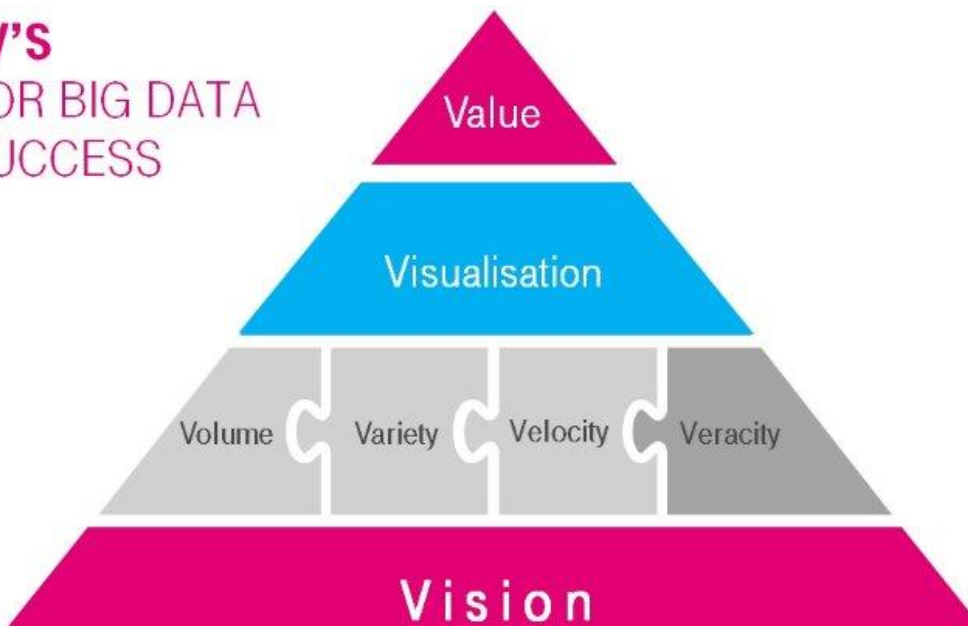
Velocity, Variety, Volume, Veracity and Value +
Variability (Изменчивость)
Visualization (Визуализация данных)



7V:

Velocity, Variety, Volume, Veracity and Value +
~~Variability (Изменчивость)~~ Vision (видение)
Visualization (Визуализация данных)

7V'S
FOR BIG DATA
SUCCESS



- Vision – у каждой компании, которая начинает работать с Big Data, должно быть видение, что с ними делать.
- Компания должна быть готова к цифровым преобразованиям. Если руководство не поймет, что может предложить Big Data, успех не будет вообще.
- Видение должно сопровождаться также изменением текущих / старых процессов.

«Наука о данных»

- Рассматривается как академическая дисциплина, а с начала 2010-х годов, во многом благодаря популяризации концепции «больших данных», — и как практическая межотраслевая сфера деятельности.
- Специализация «учёного по данным» / «специалиста по работе с [большими] данными» (data scientist) сейчас считается одной из самых привлекательных, высокооплачиваемых и перспективных профессий.

«Наука о данных»: темы

- типы данных, структуры данных, модели данных;
- представление данных, хранение и передача данных;
- методы и алгоритмы первичной обработки данных, базы данных, языки манипулирования данными;
- проектирование баз данных, языки определения данных, нормальные формы в проектировании реляционных баз данных;
- структурированные и неструктурированные данные, хранилища данных;
- анализ больших (неструктурированных) наборов данных, технологии распараллеливания обработки и сжатия информации;
- продвинутое техники баз данных, In-Memory базы данных как технологическая платформа для обработки больших наборов данных;
- базы данных NoSQL как набор технологических платформ для обработки больших наборов данных;

«Наука о данных»: темы

- задачи интеллектуального анализа больших наборов данных и проблемы больших объемов и размерностей;
- вероятностные методы первичного сжатия данных, хеширование и статистические оценки;
- задача обнаружения схожих документов, предлагаемые методы и алгоритмы, применение технологий распараллеливания обработки;
- метрические пространства, кластерные методы в снижении размерности задачи;
- рекомендательные системы, матричное представление данных, алгоритмы линейной алгебры и их использование в снижении размерности задачи;
- “всемирная паутина” (WWW), методы сбора данных и первичного анализа;
- структура “всемирной паутины” и ее использование в задачах ранжирования информации;
- интеллектуальный анализ информационных процессов.

«Наука о данных»: основные научно-инженерные направления

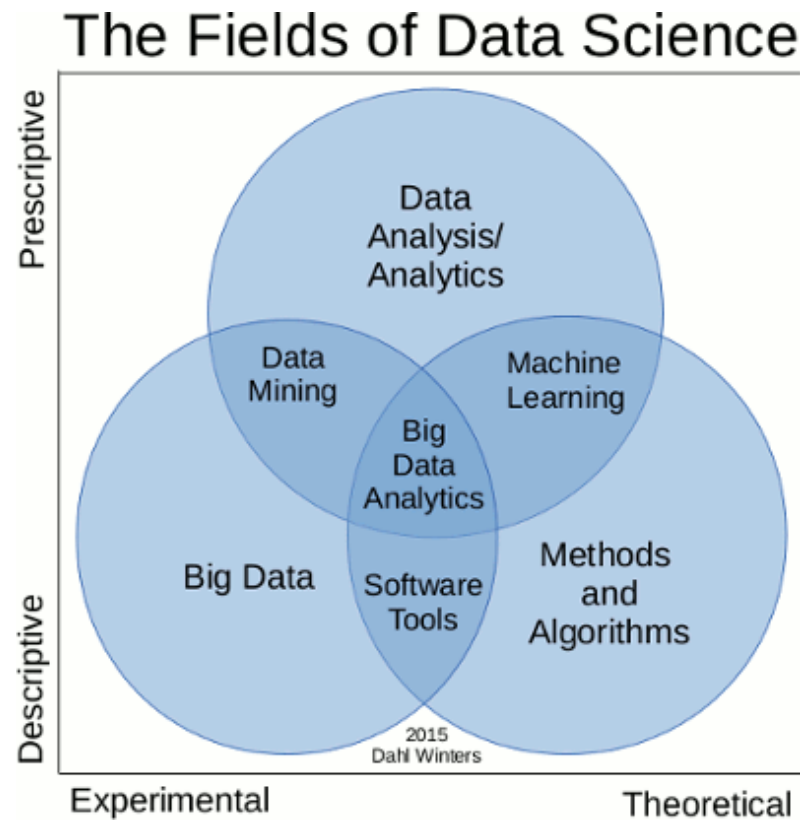
- Большие наборы данных. Сбор и обработка больших объемов данных;
- Анализ данных. Построение средств поддержки принятия решений;
- Наука о данных. Разработка и использование статистических и математических моделей, алгоритмов и визуализаций;
- Интеллектуальный анализ данных. Извлечение новых знаний;
- Бизнес-аналитика;
- Эконометрика;
- Статистика;
- Машинное обучение;
- Искусственный интеллект;
- Математическое моделирование.

«Наука о данных»: основные научно-инженерные направления

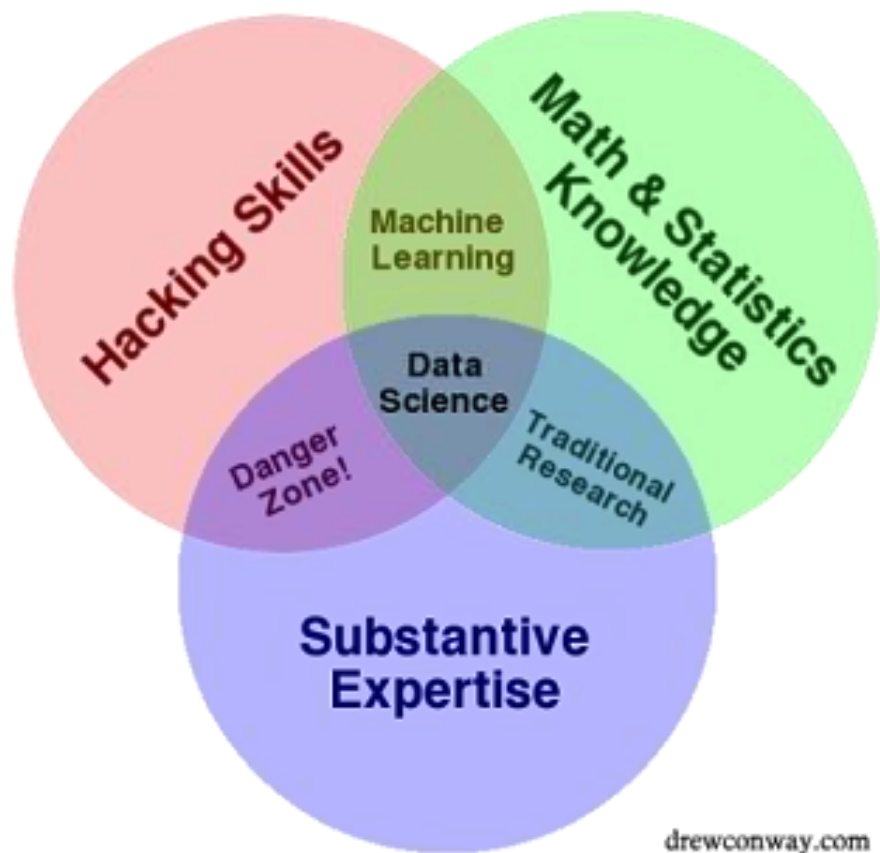
- Классификация не вполне отражает особенности математического содержания применяемых в рассматриваемой сфере методов и алгоритмов;
- Используются функциональные подходы.

«Наука о данных»: области исследований

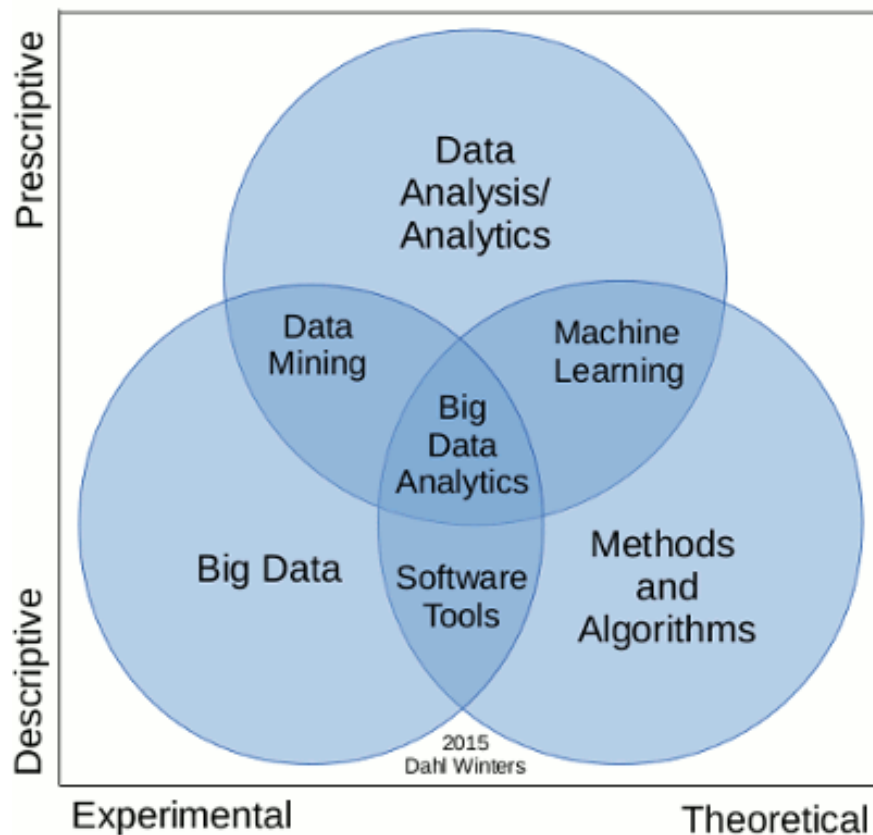
- по вертикальной оси
 - описательные - предписывающие;
- по горизонтальной оси
 - экспериментальные - теоретические.
- Термины:
 - **“Big Data”** - большие наборы данных;
 - **“Data Analysis/Analytics”** – анализ/аналитика данных;
 - **“Methods and Algorithms”** – методы и алгоритмы;
 - **“Data Mining”** – интеллектуальный анализ данных;
 - **“Machine Learning”** – машинное обучение;
 - **“Software Tools”** – программные средства;
 - **“Big Data Analytics”** – аналитика больших наборов данных.



«Наука о данных»: НАВЫКИ



The Fields of Data Science



«Наука о данных»: определение

- Наука о данных (data science/ datalogy) — раздел информатики, изучающий проблемы
 - анализа,
 - обработки и
 - представления данных в цифровой форме.
- Объединяет
 - методы по обработке данных в условиях больших объёмов и высокого уровня параллелизма,
 - статистические методы,
 - методы интеллектуального анализа данных и приложения искусственного интеллекта для работы с данными, а также
 - методы проектирования и разработки баз данных.

«Наука о данных»: Review of Controls for Certain Emerging Technologies

- (5) Advanced computing technology, such as:
 - (i) Memory-centric logic.
- (6) Data analytics technology, such as:
 - (i) Visualization;
 - (ii) Automated analysis algorithms; or
 - (iii) Context-aware computing.

Концепция преподавания основ ИИ

- В настоящее время принята следующая классификация специалистов, работающих в области ИИ:
 - Ученые по данным (Data Scientists/AI developers);
 - Инженеры по данным (Data Engineers);
 - Инженеры-программисты (Software Engineers);
 - Специалисты по эксплуатации и развитию (DevOps);
 - Специалисты по кибербезопасности (DevSecOps);
 - Менеджеры по управлению рисками / аудиторы;
 - Архитекторы данных.

Ученые по данным (Data Scientists/AI developers)

- Специалисты по обработке данных обычно выступают в роли действующих лиц при разработке ИИ по умолчанию. Большинство специалистов по данным занимаются как ручной аналитикой, так и аналитикой на основе ИИ, хотя ряд ученых по данным предпочитают специализироваться на ИИ. Например, конкретные дисциплины, такие, как нейронные сети, часто требуют специального опыта.
- Одной из причин возникновения таких специализированных должностей, как инженеры по обработке данных, советники по этике/политике и др., является то, что они позволяют ученым по данным сосредотачиваться именно на развитии новых технологий.

Современный Data Scientist: набор навыков

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative

PROGRAMMING & DATABASE

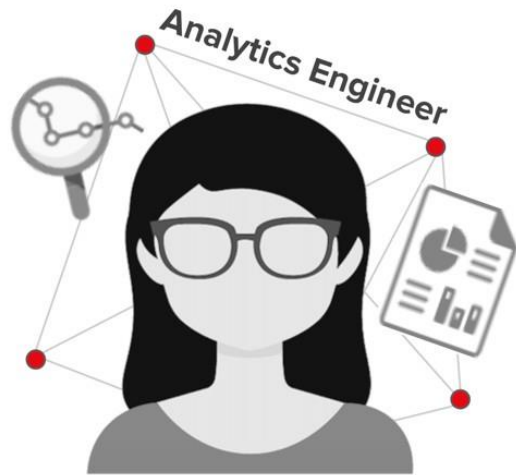
- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau



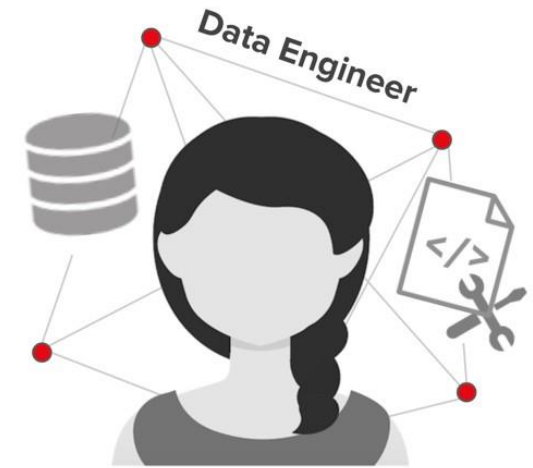
Современный Data Scientist - это специалист, работающий в области ИИ



tools: Sublime, Atom, Tableau
languages: SparkSQL, Presto, Python

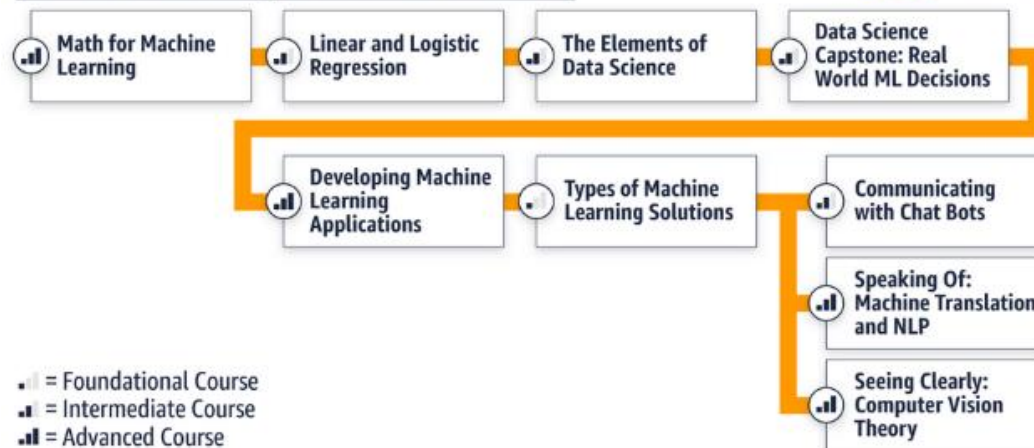


tools: Jupyter, RStudio, PyCharm
languages: Python, Presto, R, PySpark



tools: IntelliJ, PyCharm, Sublime
languages: Scala, Spark, Python, SQL

Machine Learning Path: Data Scientist



■ = Foundational Course
 ■ = Intermediate Course
 ■ = Advanced Course

Преподавание основ ИИ: учебные планы

- CC2005 – Computing Curricula
- SE2014 – Software Engineering
- CS2013 – Computer Science
- CE2016 – Computer Engineering
- IT2017 – Information Technology
- CSEC2017 – Cybersecurity

- Data Science (2021)

Computing Competencies for Undergraduate Data Science Curricula

ACM Data Science Task Force

January 2021

Andrea Danyluk, Co-chair

Paul Leidig, Co-chair



Association for
Computing Machinery

Преподавание основ ИИ для математиков: подходы, основанные на

- ориентации на фундаментальность подхода к построению и использованию математических теорий, понятий и моделей;
- строгой математической постановке задач исследования;
- разработке методов и алгоритмов их решения;
- применении разработанных алгоритмов в решении задач обработки и анализа данных.

Преподавание основ ИИ для математиков: курсы

- 1. Модели данных и базы данных – годовой:
 - 1.1. Модели данных и основы систем баз данных – полугодовой;
 - 1.2. Базы данных. Дополнительные главы – полугодовой курс ЕНС.
- 2. Аналитика больших данных – годовой:
 - 2.1. Аналитика больших данных. Основные алгоритмы – полугодовой;
 - 2.2. Аналитика больших данных. Дополнительные главы – полугодовой.

Преподавание основ ИИ для математиков: курсы

- включают в себя
 - теоретическую и
 - практическую составляющие;
- являются,
 - с одной стороны, взаимозависимыми, а
 - с другой – не требуют обязательного предварительного изучения содержания остальных спецкурсов из предлагаемого набора;
- отражают
 - как уже ставшие классическими модели и алгоритмы,
 - так и современные взгляды и понятия.

Преподавание основ ИИ для математиков & линейная алгебра

- сингулярное разложение матриц (SVD);
- собственное разложение матриц, главные компоненты;
- LU-разложение матриц;
- QR-разложение / факторизация матриц;
- симметричные матрицы;
- ортогонализация и ортонормализация;
- матричные операции;
- проекции;
- собственные значения и собственные векторы;
- векторные пространства и нормы.

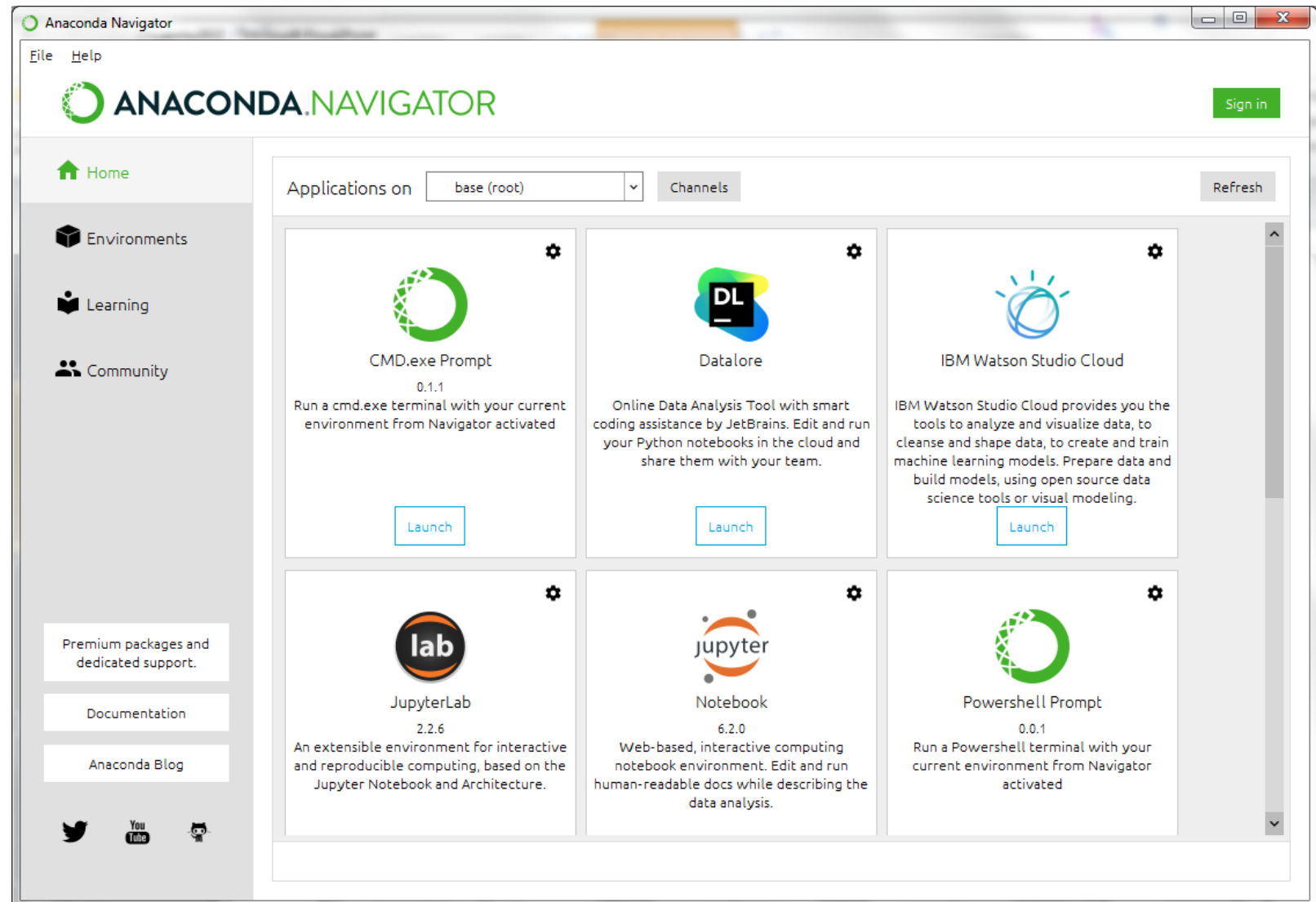
Преподавание основ ИИ для математиков: практическая поддержка (desktop RDBMS)

- моделированием данных в рамках концептуальной модели данных «сущность-связь» (ER);
- проектированием реляционных баз данных (домены, таблицы, первичные и внешние ключи, индексы, ограничения целостности и т.д.);
- декларативным (оператор SELECT, INSERT, UPDATE, DELETE) и императивным (хранимые процедуры) подходами к манипулированию данными с помощью языка SQL;
- агрегированными запросами (COUNT, SUM, MAX, MIN, AVG, ROLLUP/CUBE/WINDOW clauses) и материализованными представлениями данных (Materialized views).

Преподавание основ ИИ для математиков: практическая поддержка (SAP HANA)

- Predictive Analysis Library/PAL
 - кластеризация (Clustering);
 - классификация (Classification);
 - регрессия (Regression);
 - ассоциация (Association);
 - временные ряды (Time Series);
 - статистика (Statistics);
 - анализ социальных сетей (Social Network Analysis).
- Text mining
 - функции ранжирования терминов (ключевых слов) и документов;
 - категоризацию (Categorization) документов с использованием метода k-ближайших соседей.

Преподавание основ ИИ для математиков: практическая поддержка (Python)



Преподавание основ ИИ для математиков: практическая поддержка (Python)

```
import numpy as np
from numpy import linalg as la
```

- **NumPy** - пакет для векторного и матричного умножения
 - **NumPy** также поставляется с рядом встроенных процедур для расчетов линейной алгебры. Их можно найти в submodule **linalg**.
- **SciPy** - это научная вычислительная библиотека с открытым исходным кодом для языка программирования Python.

Научно-образовательная школа «Мозг, когнитивные системы и искусственный интеллект»



МОСКОВСКИЙ
ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ИМЕНИ
М.В. ЛОМОНОСОВА

Главная Школы ▾

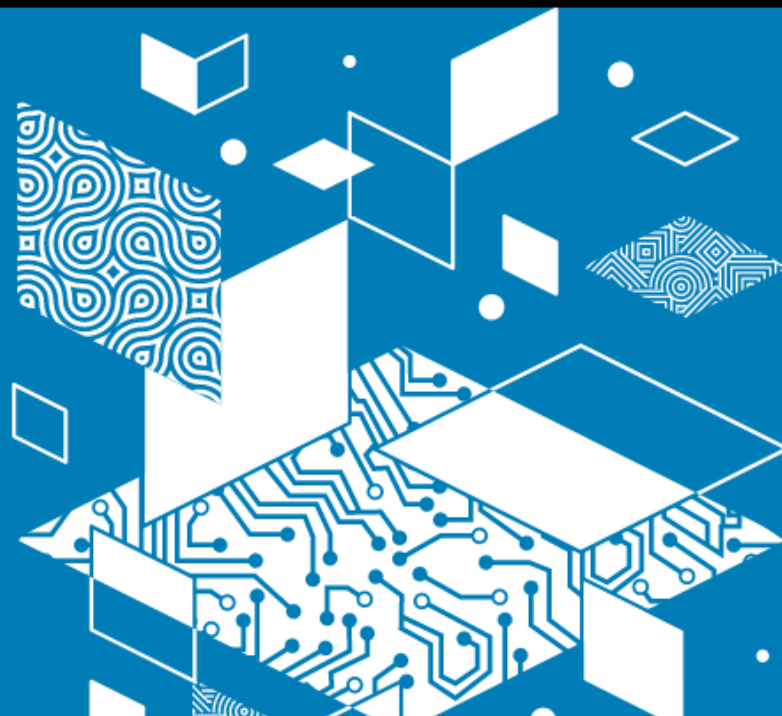


НАУЧНО-
ОБРАЗОВАТЕЛЬНЫЕ
ШКОЛЫ МГУ

Мозг, когнитивные системы, искусственный интеллект

Приоритетное направление:

переход к передовым цифровым, интеллектуальным производственным технологиям, роботизированным системам, новым материалам и способам конструирования, создание систем обработки больших объемов данных, машинного обучения и искусственного интеллекта



Вопросы ?

Математики МГУ предложили концепцию преподавания основ искусственного интеллекта

msu.ru/science/main_themes/matematiki-mgu-predlozhili-kontseptsiyu-prepodavaniya-osno...

Домашняя страница Bookmarks Начальная страница Mail.Ru Support Packages a... client Другие закладки


Рус Eng 中文

Новости Объявления Пресс-служба Сайты МГУ Адреса Карта сайта Поиск

Главная / Научная работа / Главные темы / Математики МГУ предложили концепцию преподавания основ искусственного интеллекта

17/02/22 математика НОШ МГУ

Математики МГУ предложили концепцию преподавания основ искусственного интеллекта



Ученые Московского университета предложили авторскую концепцию построения образовательных траекторий при подготовке специалистов в сфере искусственного интеллекта. Исследование выполнено при поддержке Междисциплинарной научно-образовательной школы МГУ «Мозг, когнитивные системы, искусственный интеллект». Концепция описана в журнале «Интеллектуальные системы».

Мир достиг поворотного момента в развитии методов и средств искусственного интеллекта (ИИ), когда создаваемые технологии перешли из области теории и науки в «реальный мир», то есть стали доступны практически всем отраслям экономики. Этот сдвиг является весьма эффективным, но в то же время непростым, поскольку сочетает в себе сложности применения ИИ со сложностями организации человеческой деятельности. Сегодня ни у кого не вызывает сомнений, что успехи в развитии экономики страны, в особенности в ее цифровизации, в очень большой степени связаны с внедрением технологий ИИ. С одной стороны, есть проблема обеспечения междисциплинарного подхода, а с другой — недостаток квалифицированных многопрофильных специалистов.

«Университетам нужно развивать научно-образовательные программы, по которым студенты будут обучаться всему, что им необходимо для дальнейшей успешной карьеры в выбранной области, а также участвовать в разработке реальных проектов (например, в студенческих конструкторских бюро)», — отметил [Илья Бурькин](#), научный сотрудник кафедры теоретической информатики механико-математического факультета МГУ.

Дефицит человеческих талантов является наиболее заметным дефицитом ИИ в стране и становится самым серьезным препятствием для приобретения, создания и внедрения технологий на базе искусственного интеллекта.

«Нужно строить новые конвейеры талантов, создать новые образовательные программы, а также важно улучшать как STEM-образование, так и систему приема и удержания высококвалифицированных кадров», — добавил [Сергей Главацкий](#), доцент кафедры теоретической информатики механико-математического факультета МГУ.

Сотрудники НОШ МГУ «Мозг, когнитивные системы, искусственный интеллект» описали подход к подготовке специалистов по ИИ основных направлений: ученый по данным, инженер по данным, аналитик данных и инженер-программист. Кроме того, авторы рассмотрели дополнительные специальности, которые необходимо ввести для решения задач ИИ в производстве: специалист

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В.ЛОМОНОСОВА

ОСНОВАН В 1755 ГОДУ

ГЛАВНАЯ СТРАНИЦА
ОБЩИЕ СВЕДЕНИЯ
УЧЕБА
ПОСТУПАЮЩИМ
НАУКА
НАУЧНО-ОБРАЗОВАТЕЛЬНЫЕ ШКОЛЫ
ПРИОРИТЕТНЫЕ НАПРАВЛЕНИЯ
РЕЙТИНГ НАУЧНОГО СОТРУДНИКА
КОНКУРСЫ НИР
ДИССЕРТАЦИОННЫЕ СОВЕТЫ
ПРЕМИИ И НАГРАДЫ
КОНФЕРЕНЦИИ
ТЕКУЩИЕ КОНКУРСЫ, ГРАНТЫ, СТИПЕНДИИ
ОБЩАЯ ИНФОРМАЦИЯ
МЕЖДУНАРОДНОЕ