КАФЕДРА
ТЕЛЕМАТИКА

Санкт-Петербургский
Государственный
Политехнический
Университет

Институт прикладной
математики и механики

Семинар по специальности на английском языке

тема

Trust in science: separating the explanations and understanding from the models

Workshop

Лекция 10

11 November
2020 г.

Introduction: Science and Conservation Law

All Sciences are divided into natural, “unnatural” and social (humanitarian) classes

Natural sciences study the laws of Nature, which manifest themselves independently of the human observer, and their understanding is expressed using various models that reflect the essence of conservation laws.

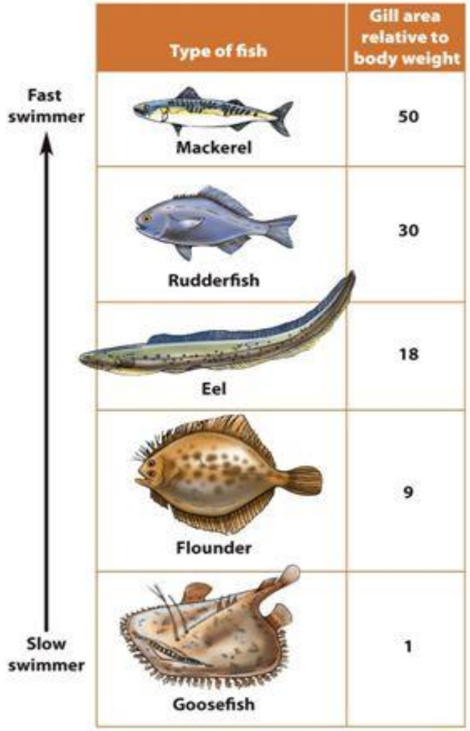
“Unnatural” science or engineering study the ways how to speed up evolution processes and design useful for people new object/structure/matter….

The social sciences are characterized by the objective impossibility of separating the object of research from the human observer, which leads to the need to take into account in the models subjective factors that enhance the situational nature of humanitarian knowledge**.**

This feature of humanitarian knowledge can be explained with the help of the phenomenological law of the **conservation of risks: the general risks of a social system in any of its transformation remain constant.**
In fact, the risks of the system “accumulate” to a critical volume, which leads to the transfer of the system to a new “stable state”.

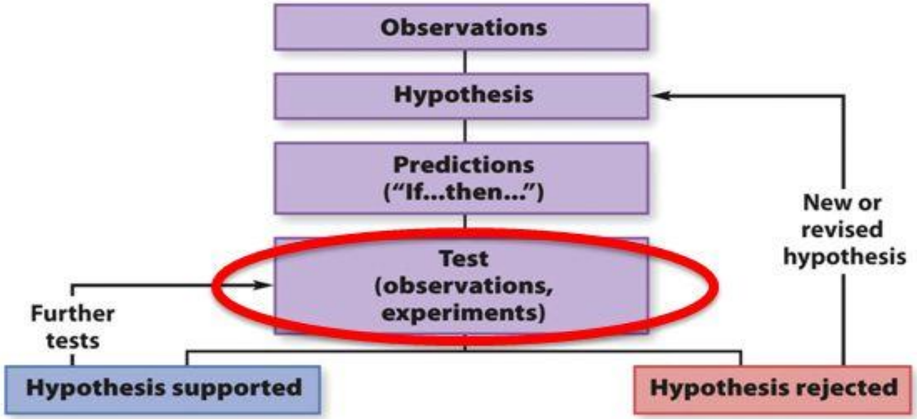
Natural Science Explanations

An Example of the Scientific Method



	Type of fish	Gill area relative to body weight
Fast swimmer	Mackerel	50
	Rudderfish	30
	Eel	18
	Flounder	9
Slow swimmer	Goosefish	1

Discover Biology, 5/e Figure 28.14
© 2012 W. W. Norton & Company, Inc.



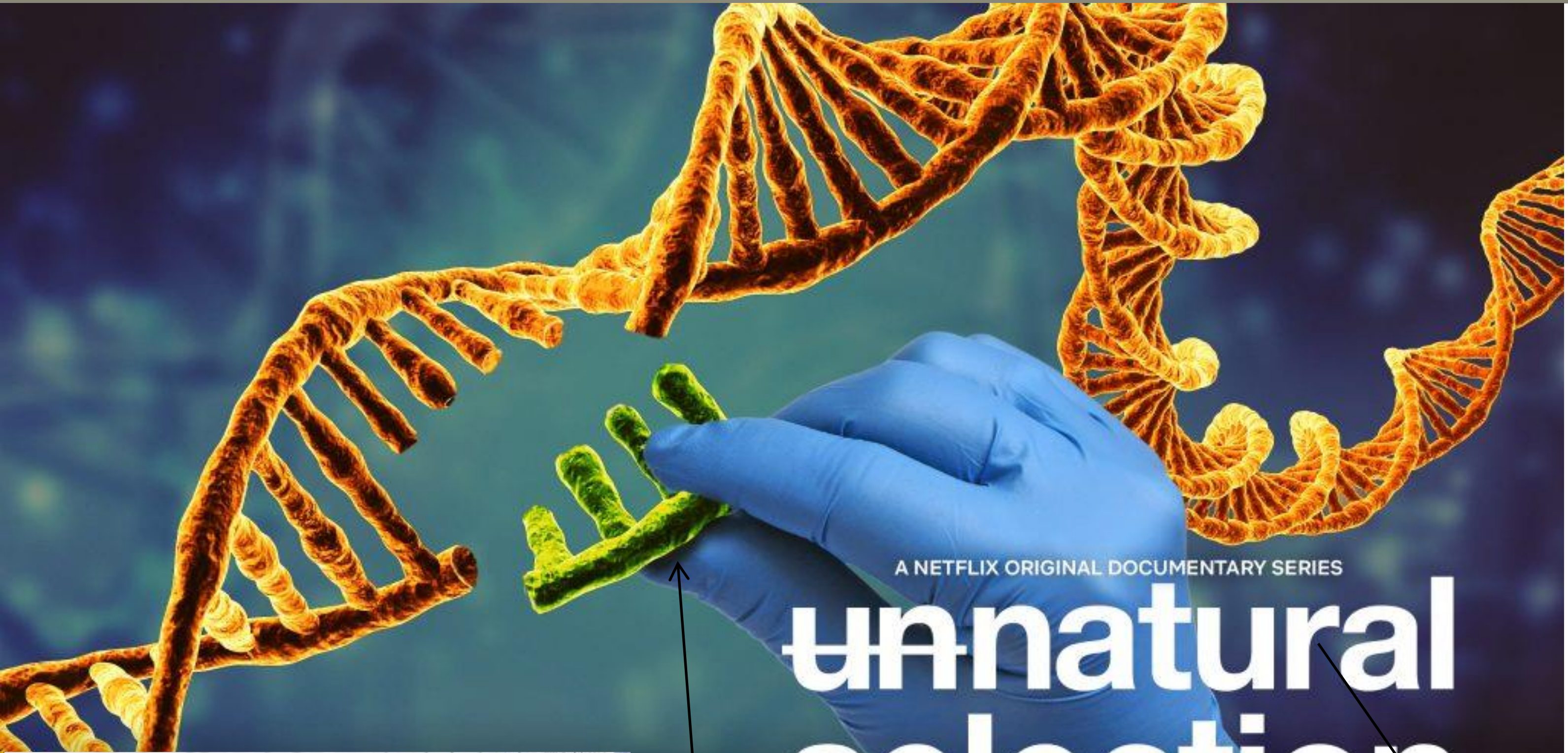
Discover Biology, 5/e Figure 1.2
© 2012 W. W. Norton & Company, Inc.

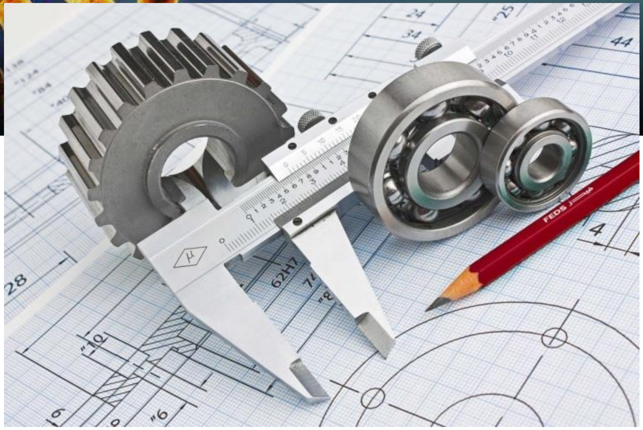
What were your results?

Is your hypothesis supported or not?

Why do you trust to obtain result ?

“Nunatural” science or Engineering

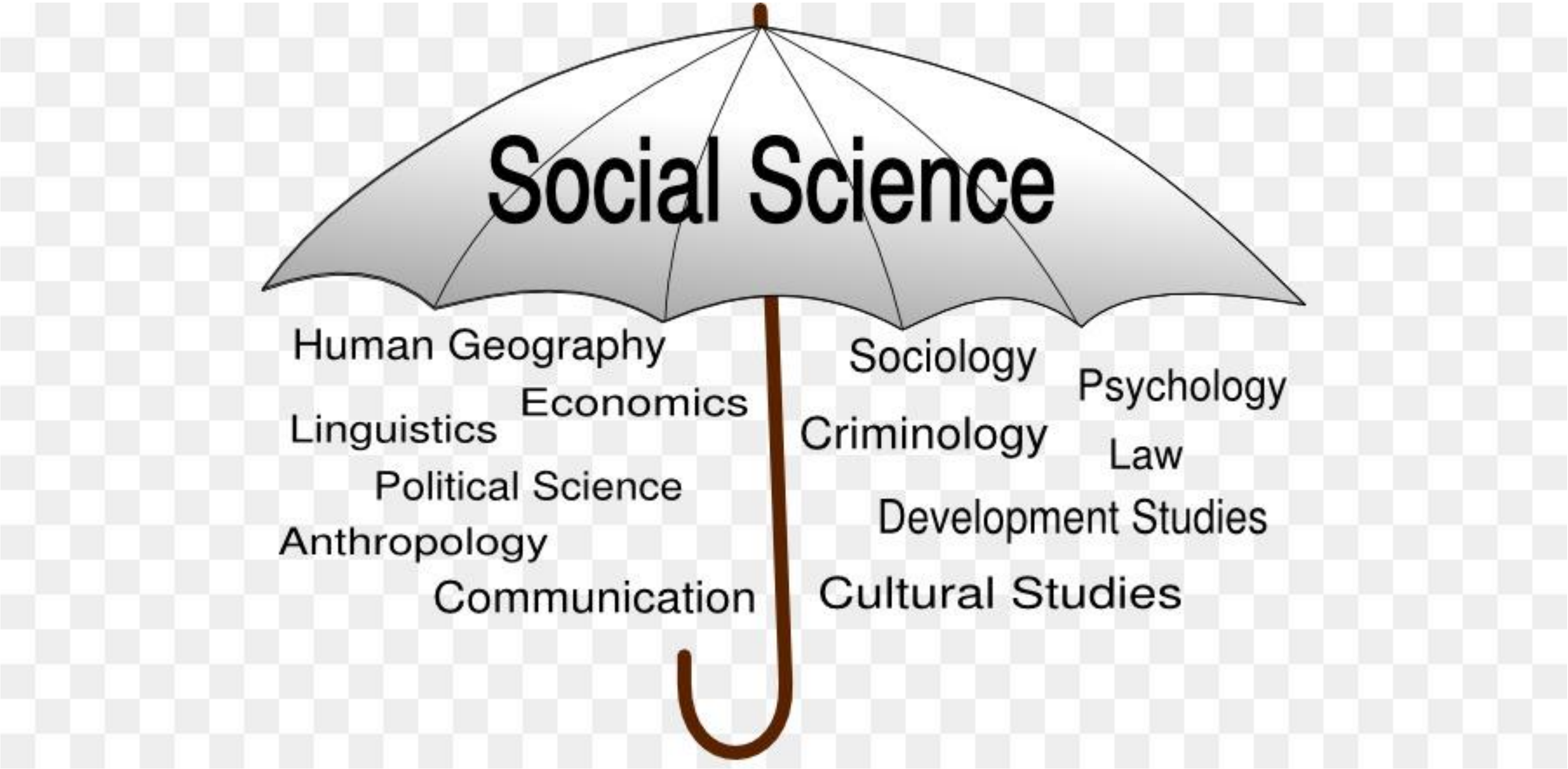




Engineering of new objects and design new mater

Explanation
Why, Where,
When….

Social (humanitarian) sciences



Are any fundamental lows of humanitarian sciences ?
Low of what ?
(energy, meter, information, relationship,….

What we need

- Main idea : model-agnostic interpretation methods
- Model-agnostic explanations method can be used for any type of model.
- High-Precision Model-Agnostic Explanations

Knowledge as a fact of ..... "glorified statistics"

Interpretation:

Math statistics - functions of experimental SAMPLING

“Glorified statistics“ – intellectual function, which includes:

- perception (obtaining data through communication channels),
- understanding (matching the utility function to the data processing process)
- knowledge itself (собственно познание) (building a logical model of perceived data and an algorithm for calculating a subset of the model state space on which utility function reaches maximum).

Intellect as a function: set of “computational” operations that can deliver the solution to **the problem - **constructing an or models of“ possible worlds ”, on which the maximum utility function is achieved.

Main aspects of trust in formal explanation system

Model flexibility: The interpretation method can work with any, such as random forests and deep neural networks.

Explanation flexibility: You are not limited to a certain form of explanation. In some cases it might be useful to have a linear formula, in other cases a graphic, etc.

Representation flexibility: The explanation system should be able to use a different feature representation as the model being explained.

High level look at model-agnostic interpretability:

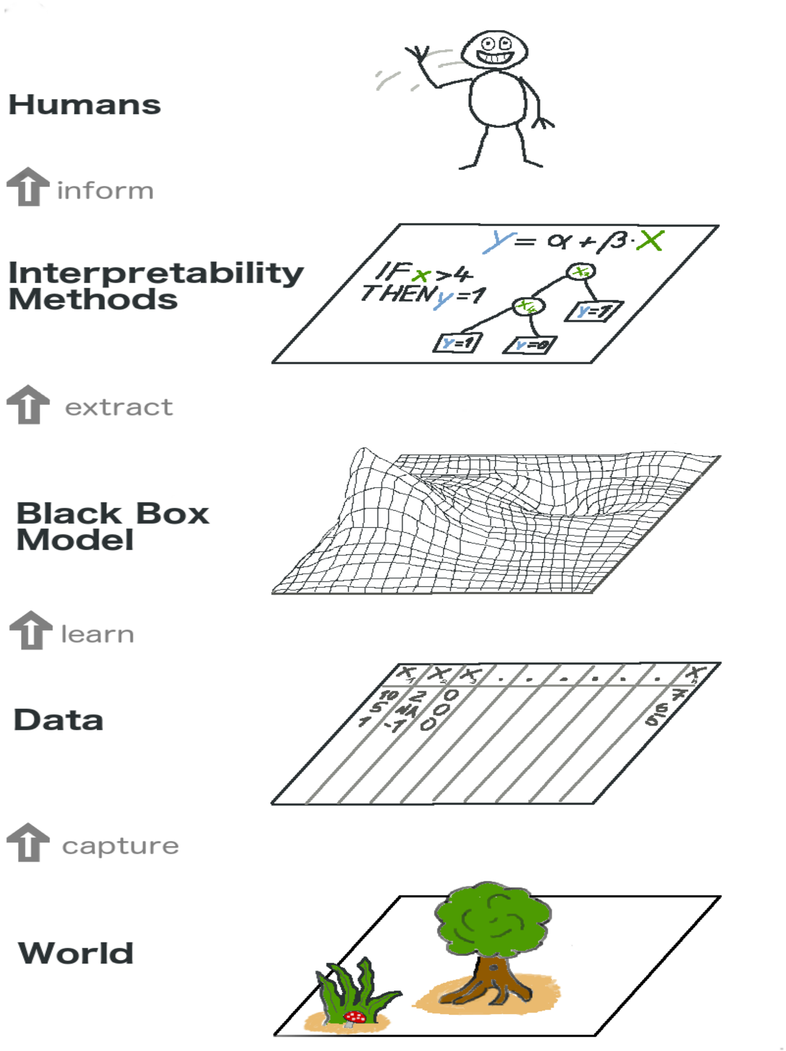
Humans - consumers of the explanations and knowledgrs . Why we trust in …..

Interpretability Methods layer, which helps human deal with the opacity of machine learning models (how machine calculate explanations)

Black Box Model layer - algorithms using data from the real world to make predictions, find structures or invariants

The Data layer contains ‘digital twins” anything from images, texts, tabular data and so on in order to make it process able for computers and also to store information.

The World layer contains everything that can be observed and is of interest.



Humans

inform

Interpretability Methods

extract

Black Box Model

learn

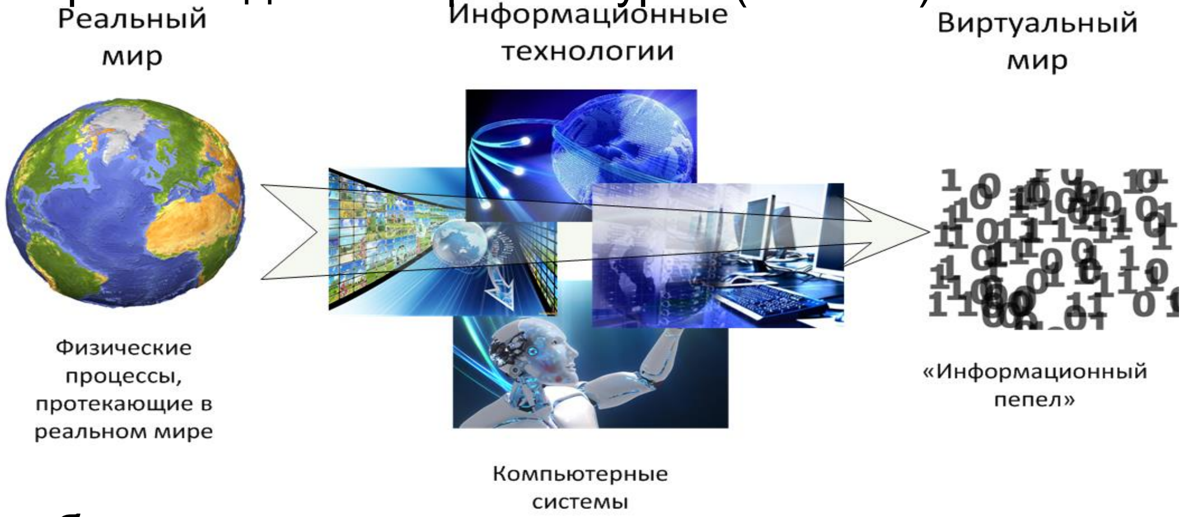
Data

capture

World

deal with the opacity - solving “inverse” problems in the space of “augmented reality”

Direct problem _ прямая задача вычислений – моделирование объектов с помощью компьютеров с заданной архитектурой (АО+ПО):



Реальный мир

Информационные технологии

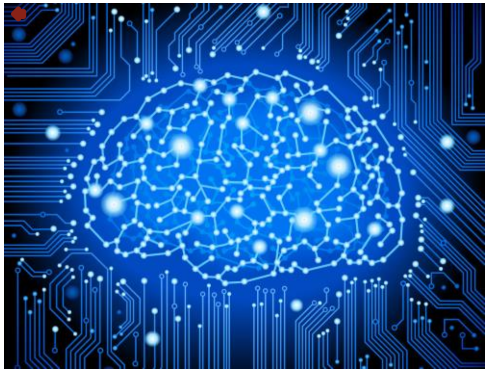
Виртуальный мир

Физические процессы, протекающие в реальном мире

Компьютерные системы

«Информационный пепел»

Invers problem – выбор такого алгоритма вычисления, «генерирующий» данные, на которых функция полезности достигает максимума :

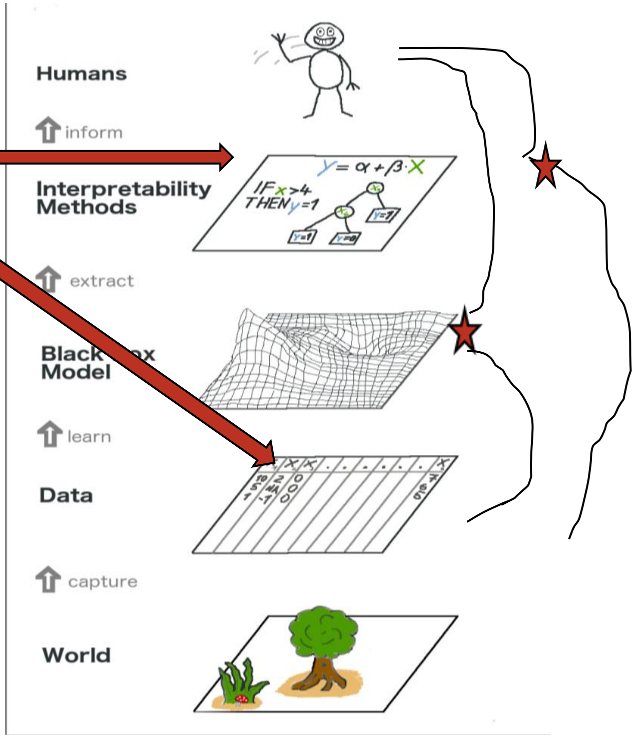


This multi-layered abstraction

We need to understand the differences in approaches between statisticians and machine learning approaches.
Statisticians deal with the Data layer, such as planning, estimation, predictions, skip the Black Box Model layer and go right to the Interpretability Methods layer.

Machine learning specialists also deal with the Data ∠ layer, train a black box machine learning mode and skip The Interpretability Methods layer, so Humans directly deal with the Black Box model predictions.

It’s great that interpretable machine learninп merge (unite, join) the work of statisticians and machine learning specialists.



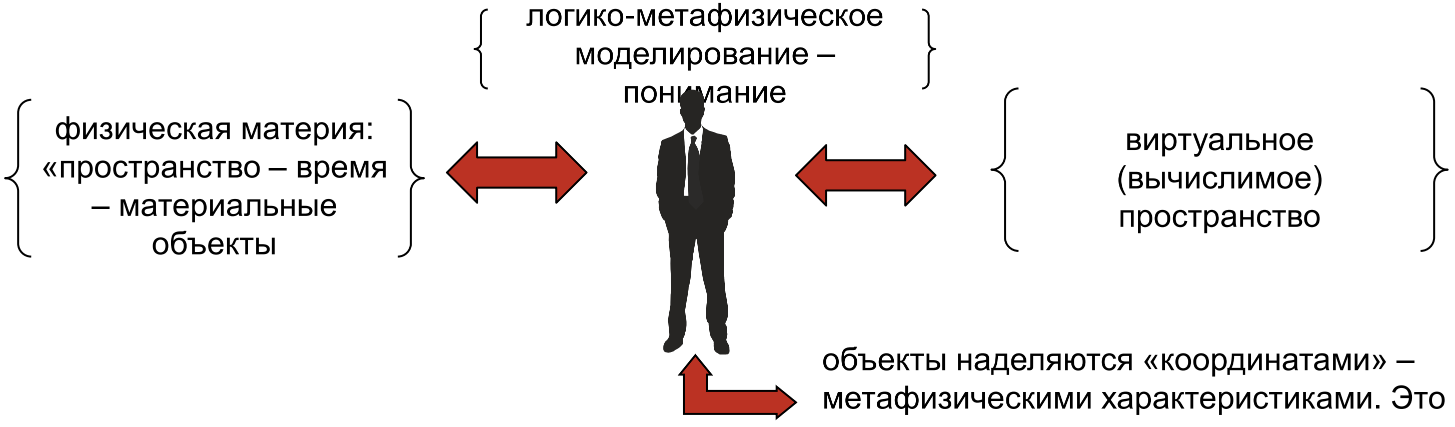
Суть проблемы «моделей»

Согласно др. Дж. Диспензе (исследователь в области нейрофизиологических процессов), наше прошлое «записано» в нейросетях мозга, которые формируют модели того, как мы воспринимаем и ощущаем мир в целом и его конкретные объекты в частности.

Итого: в 99% случаев мы воспринимаем реальность не такой, какая она есть, а интерпретируем ее на основе готовых моделей (образцов) из прошлого.

Model-agnostic approach

шаг 1



логико-метафизическое моделирование – понимание

физическая материя: «пространство – время – материальные объекты»

виртуальное (вычислимое) пространство

объекты наделяются «координатами» – метафизическими характеристиками. Это называется процессом «координатизации» или алгебраизации.

шаг 2

На множестве объектов с координатами задается алгебра кардиналов множеств: операции сложения и умножения, вычисления характеристических функций – предикатов: ND->2D: $f(x_1, x_2, \dots, x_n) = \begin{cases}1\\0\end{cases}$

{ реальный физический мир } $\xrightarrow{\text{морфизм}}$ { виртуальный мир классов, построенных на различных принципах координатизации, отношений порядка и топологий (преобразований) с инвариантами }

Существует ли «физика» .... Мышления (consciousness) ?

Так, чтение или письмо – есть тренировка для головного мозга, в особенности если при этом вы узнаёте или выражаете нечто новое.

о:

- A change in consciousness in the process of thinking leads to changes in the physical body of an intellectual subject.
- "Machine" will acquire the ability to think "if it acquires the properties of a" data-driven processor "