



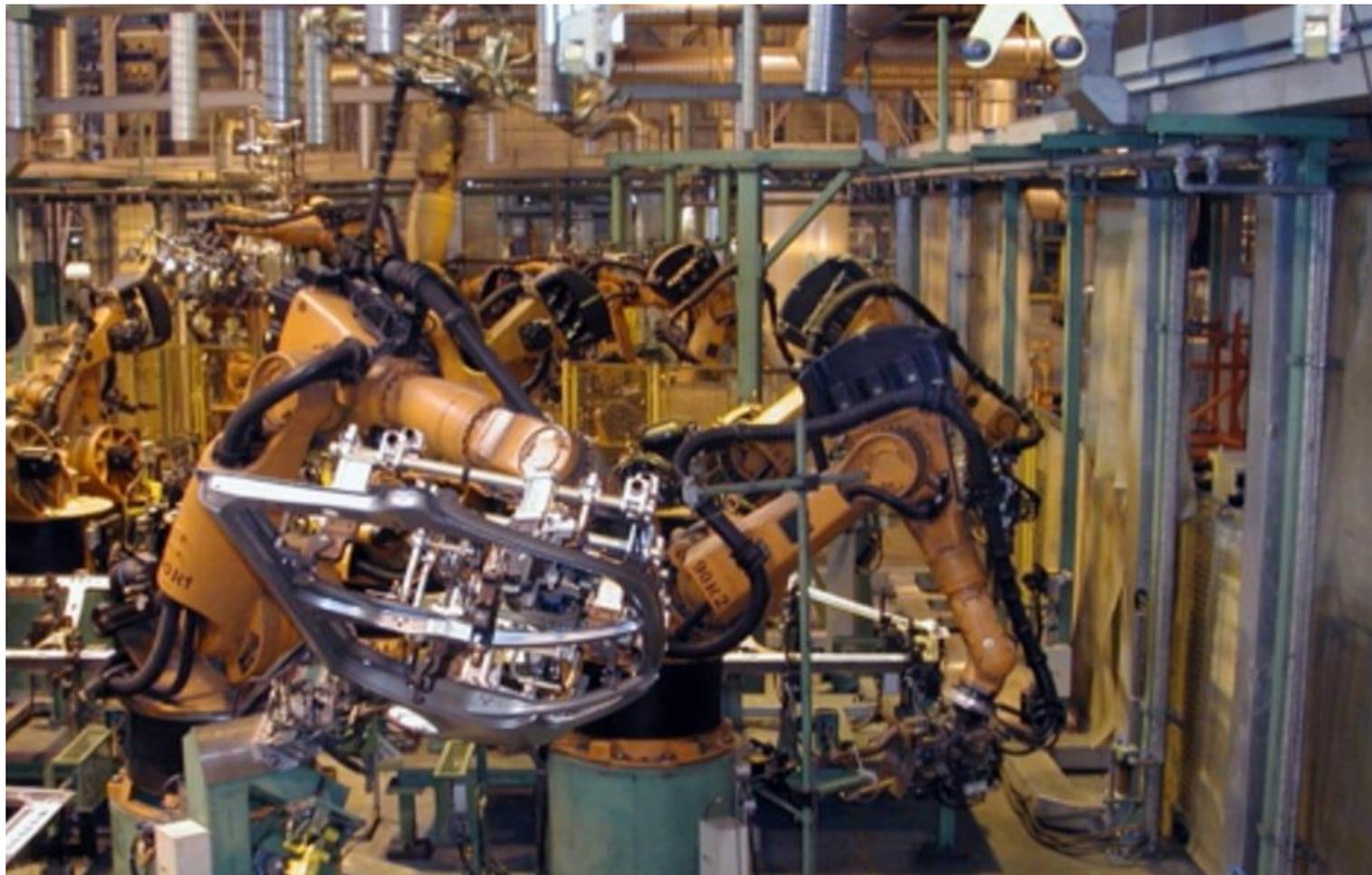
*Thinking is not the ability to manipulate **language**;  
it's the ability to manipulate **concepts**. Computer  
science **should be about concepts**, not languages.  
Лесли Лэмпорт*

## Лекция 1

# МОТИВАЦИОННАЯ

## **Р О Б О Т Ы и интеллектуальные СИСТЕМЫ**

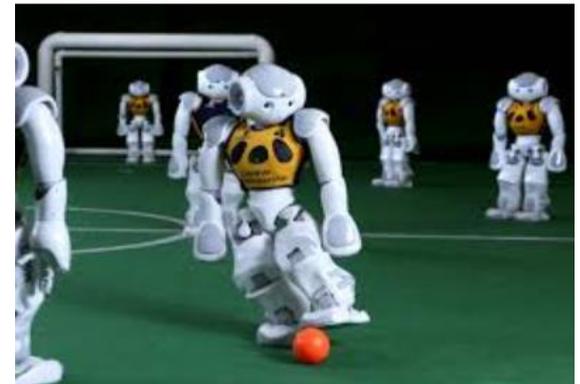
со страниц фантастических романов и фильмов  
становятся частью новой Реальности



Роботы – или киберфизические системы становятся частью промышленной инфраструктуры

Системы на базе технологии “Искусственный интеллект вытеснит :

- комбайнеров и трактористов. Курс на замещение всех видов водителей роботами”
- Моряков с Кораблей-контейнеровозов, оснащенных с автопилотами:  
“Через 5 лет профессия моряка отомрет”
- .....
- “Согласно прогнозам к 2030 году потеряют работу 800 млн чел.”, а
- на чемпионате мира 2050 г. команды автономных роботов обыграют футбольные команды из людей ”



- “Цифровизация” экономики;
- компьютерное управление критическими структурами;
- “умный дом”, “интернет вещей”;
- киберфизические системы;
- роботизированные производства;
- электронные правительства;
- электронное дистанционное обучение;
- роботы-хирурги;
- боевые роботы;
- ... ..



**очень скоро люди будут вынуждены**  
- впервые в истории цивилизации -  
**сосуществовать с чужим интеллектом с роботами, которых они**  
**сотворят сами!**



Но ..... облик будущего зависит  
и от нас -

**СОЗИДАТЕЛЕЙ НОВОГО МИРА**



Сотворение  
Адама  
Микеланджело  
Сикстинская капелла

- У нас есть уникальная возможность быть не только современниками и свидетелями этого процесса, но и непосредственными ТВОРЦАМИ новой искусственной жизни, которая будет сосуществовать и ЭВОЛЮЦИОНИРОВАТЬ вместе с нами на Земле и в космосе

By Peter Singer (профессор биоэтики Принстонского университета).

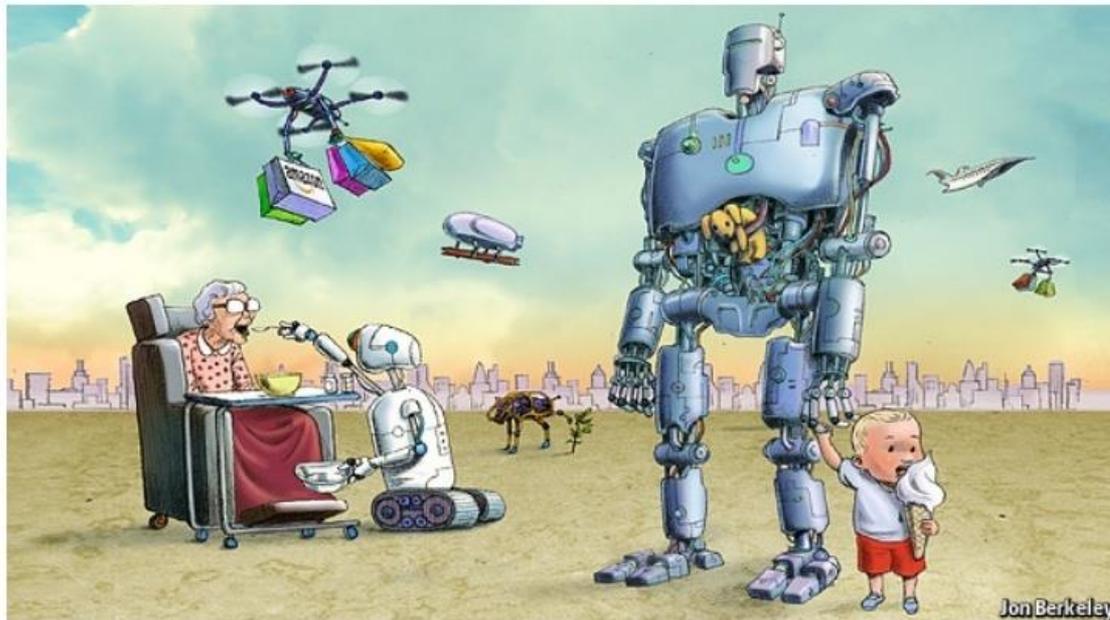
PRINCETON, NEW JERSEY – Last month, AlphaGo, a computer program specially designed to play the game GO, caused shock waves among people when it defeated Lee Sidol, one of the world’s top-ranked professional players, winning a five-game tournament by a score of 4-1.

Why, you may ask, is that news? Twenty years have passed since the IBM computer Deep Blue defeated world chess champion Garry Kasparov, and we all know computers have improved since then. But Deep Blue won through raw computing power, using its ability to calculate the outcomes of more moves to a deeper level than even a world champion can. Go is played on a far larger board (19 by 19 squares, compared to 8 by 8 for chess) and has **more possible moves than there are atoms in the universe**, so raw computing power was unlikely to beat a human with a strong intuitive sense of the best moves.

Instead, AlphaGo was designed to win by playing a huge number of games against other programs and adopting the strategies that proved successful. You could say that AlphaGo evolved to be the best go player in the world, **achieving in only two years what natural selection took millions of years to accomplish.**”

## Rise of the robots

*Prepare for a robot invasion. It will change the way people think about technology*



Майкрософт создал, что ИИ будет «умнее», чем любой человеческий мозг, практически в каждой области, ... . Такая система будет способна «перехитрить» любого человека, который захочет ее отключить.”

Оценки экспертов:

ИИ достигнет уровня мышления человека ~ к 2050 г.

ИИ превысит уровень мышления человека ~ к 2075 г.

- Стивен Хокинг:
  - "ИИ - возможный убийца человеческой цивилизации"
- Prof. Marek Perkowski. Dept of Electrical and Computer Engineering, PSU:
  - "Agent-based systems will dramatically change the world we live in".
- С. Рассел и П. Норвиг. Искусственный интеллект. //2002. Стр.1286:
  - “ ... системы ИИ могут создать угрозу самоопределению, свободе и даже выживанию людей. По этим причинам нельзя рассматривать исследования в области ИИ в отрыве от их этических последствий.”





Из Интернета:

“ ... сейчас ни о каком восстании машин речи идти не может — нынешние автоматы полностью подконтрольны своим операторам и не осознают себя мыслящими единицами.

Они способны лишь выполнять команды, которые дает им человек.

Те сбои, что происходят, не являются предвестниками какого-то "робокалипса", это **всего-навсего технические сбои**, вызванные либо несовершенством системы, либо **оплошностью** тех, кто следит за состоянием машины.”

**ЭТО НЕ ТАК !**

Это не оплошности и не сбои!

Это систематические проявления глубокой  
**концептуальной проблемы.**



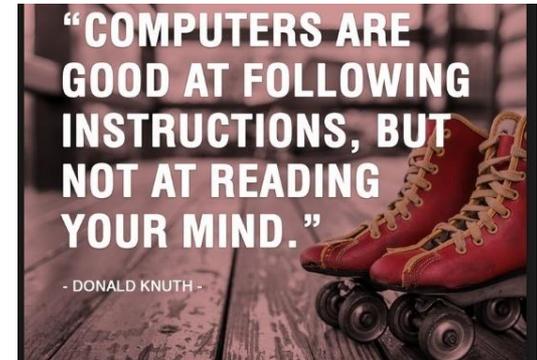
**Проблема в том, что**

системы, которые мы строим, ведут себя в точности так, как мы их запрограммировали,

**НО**

**совсем не так, как мы предполагали!**

Человеческий мозг не в состоянии справиться со сложностью при разработке программ объемом в сотни миллионов строк кода



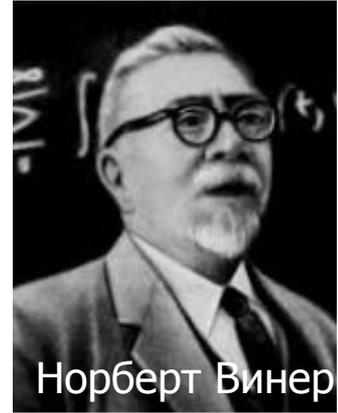
“Если магия вообще способна даровать что-либо, то она дарует именно то, что вы попросили, а не то, что вы подразумевали, но не сумели точно сформулировать.

Если вы ведете войну ..., то победа будет достигнута любой ценой, даже ценой уничтожения вашей собственной стороны, если только сохранение ее жизнеспособности не будет совершенно четко запрограммировано в числе условий победы”.

**Это нечто большее, чем невинный парадокс.**

“магическое исполнение заданного осуществляется машинами в высшей степени буквально.”

Винер видит выход в том, чтобы замкнуть принятие окончательных решений на человека - оператора



Норберт Винер

*«Компьютеры похожи на ветхозаветных богов: сплошные правила и ни капли жалости».*

( Дж. Кэмпбелл)

**Если компьютер решит, что вы умерли, то вы умерли ...**

В 2010 году ошибка управления педалью газа в автомобилях Toyota привела к гибели 89 человек, отказ компьютерной сети в нескольких американских аэропортах вызвал крушение трех авиалайнеров.

Программирование является единственной областью инженерной деятельности, где разработчик (**ПОКА**) не отвечает за качество своей работы.

Обычно ПО имеет 10-15 ошибок на 1000 строк кода,  
ПО высокого качества – 3 ошибки на 1000 строк кода.

Современное ПО, управляющее реальной технической системой,  
состоит из сотен миллионов строк кода.

Программы, управляющие техникой вокруг нас, наполнены ошибками.

SOFTWARE HORROR STORIES: <http://www.cs.tau.ac.il/~nachumd/horror.html>

Collection of Software Bugs: <http://www5.in.tum.de/~huckle/bugse.html>

Worst software defects in history:

<http://cristianpocovnicu.wordpress.com/2011/03/22/worst-software-defects-in-history/>

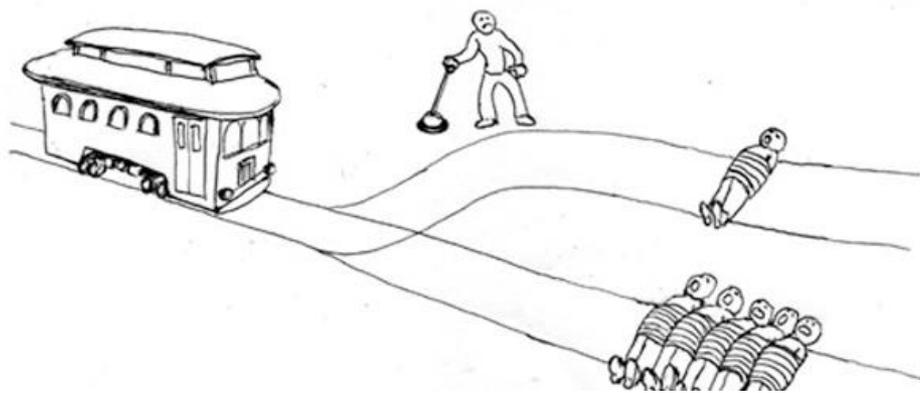
Википедия:

[http://en.wikipedia.org/wiki/List\\_of\\_software\\_bugs](http://en.wikipedia.org/wiki/List_of_software_bugs)

- Tesla начала продажи Model S с встроенным роботом, однако автономный **Tesla улетел под прицеп разворачивающегося грузовика**. Водитель погиб на месте. Причиной ДТП стало ослепление датчиков прямыми лучами солнца. *Компьютер не смог понять, что за объект вырос под носом.*
- Эксперименты Google и Jaguar Land Rover показали, что роботизированная машина, едущая исключительно по правилам, способна создать **еще большую опасность, чем лихачи**.
- Автопилот не видит разницы между картонной коробкой и бетонным блоком. **Следуя машинной логике, он никогда не покинет кругового перекрестка, ведь для этого требуется формальное нарушение правил.** ... На круге действуют джентельменские соглашения, не понятные электронному разуму. Компьютер не распознаёт вежливости попутчиков ... .

- Основной элемент интеллектуального робота – встроенный микропроцессор, управляющий его поведением, включая принятие решений и координацию нескольких роботов для коллективного решения задач.
- Новая область интеллектуальных систем называется “*многоагентными системами*”, МАС
- Метафора МАС – новый взгляд на мир, состоящий из агентов, в котором мы сами – только часть, биологические агенты.
- *Искусственные агенты – РОБОТЫ - скоро будут функционировать в обществе наряду с нами.*
- **Разработка многоагентных систем требует качественно новых подходов, теорий и технологий программирования.**

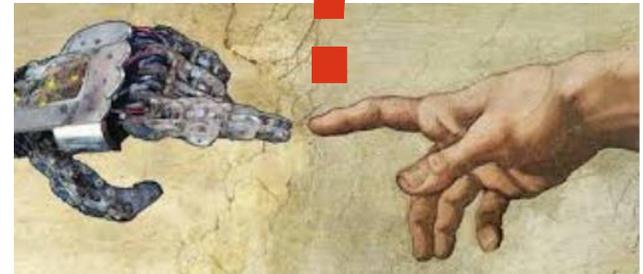
К 2025 году в мире ежегодно будет продаваться 15 млн автомобилей с автопилотами, которые будут обходиться без водителя



Какую альтернативу выберет автопилот, **если одна из них неизбежна**: задавит одного ребенка или пять мужиков?

**Этические и моральные проблемы для роботов выходят на первый план**

- ГЛОБАЛЬНЫЙ ВЫЗОВ ЧЕЛОВЕЧЕСКОЙ ЦИВИЛИЗАЦИИ
- всевѣдущий «универсальный разум», сотворил человека, передал ему моральные законы.
- Теперь человек сам создает искусственный цифровой мир и реализует в нем законы поведения искусственных агентов.



Сможем ли **МЫ** при создании роботов сделать их предсказуемыми и управляемыми, заложить в них **МОРАЛЬНЫЕ** законы и **НОРМЫ** социального поведения, которые предотвратят катастрофу?

*“It is an incredible privilege to be a software developer.  
You know, how cool is that?*

*At the same time, it's a deep responsibility,  
because what we do does change the world”.*

(Grady Booch, Chief Scientist of Rational Software)  
разработчик UML, *Object Oriented Analysis and Design*



... то, что мы делаем, **ДЕЙСТВИТЕЛЬНО** меняет  
мир

Задача №1 – разработка интеллектуальных агентов с **нужными** свойствами

- При чем здесь логика ?
- Как формализовать **нужные** законы *поведения* и *этики* и внедрить их в “мозги” (ПО) агентов?
- Как можно построить “**компьютерные мозги**”, которые будут взаимодействовать с ДРУГИМИ мозгами, машинными либо биологическими? Какие права и обязанности могут они иметь?

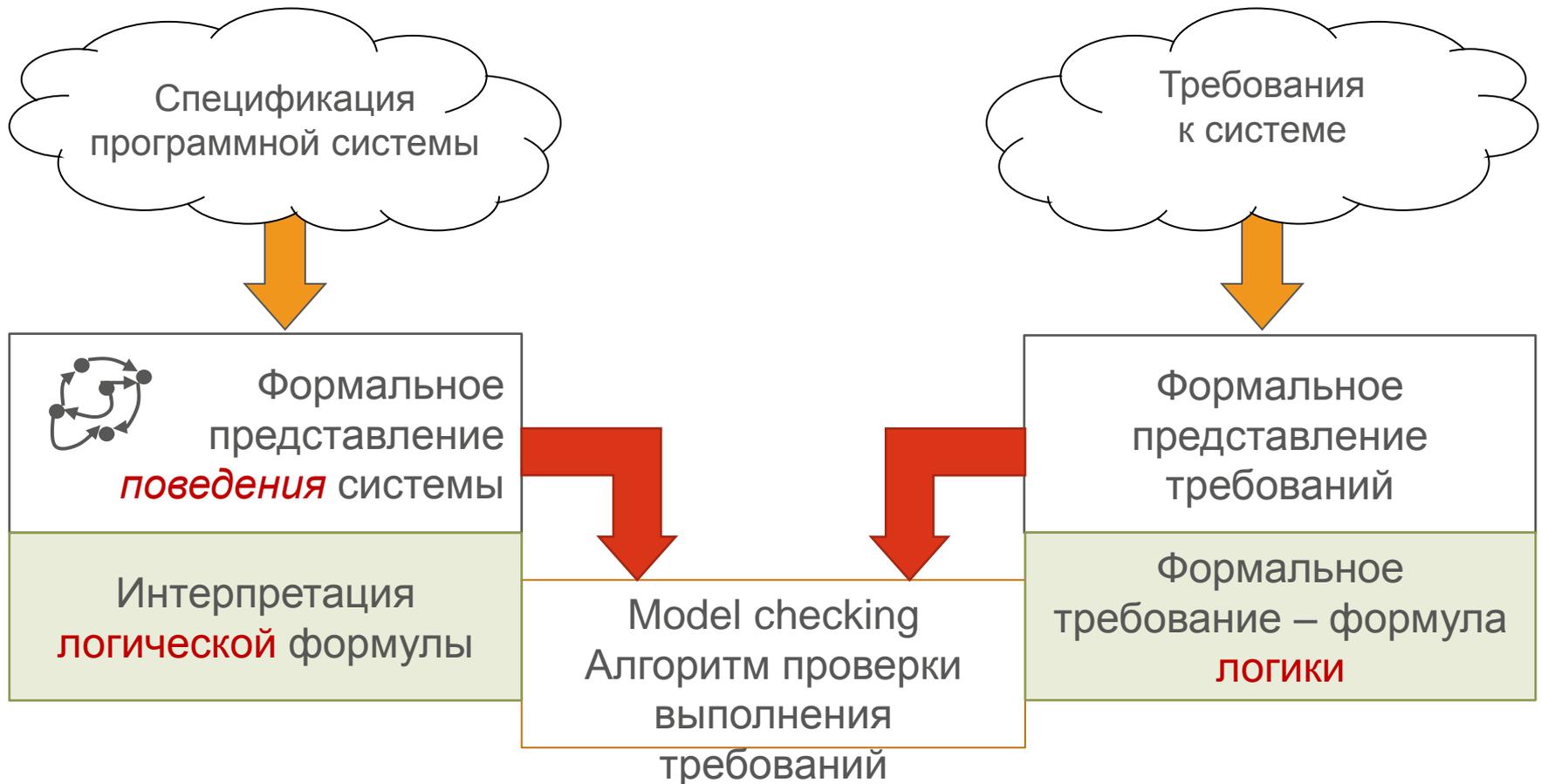
- Формальные методы сегодня почти не используются в практике программирования. Программисты боятся формальных методов, считают их ненужным приложением к хитростям программирования.
- Программисты строят свои программы, исходя из “здорового смысла”.

Разработать интеллектуальную многоагентную систему с предсказуемым и контролируемым поведением, исходя из “общих соображений”,

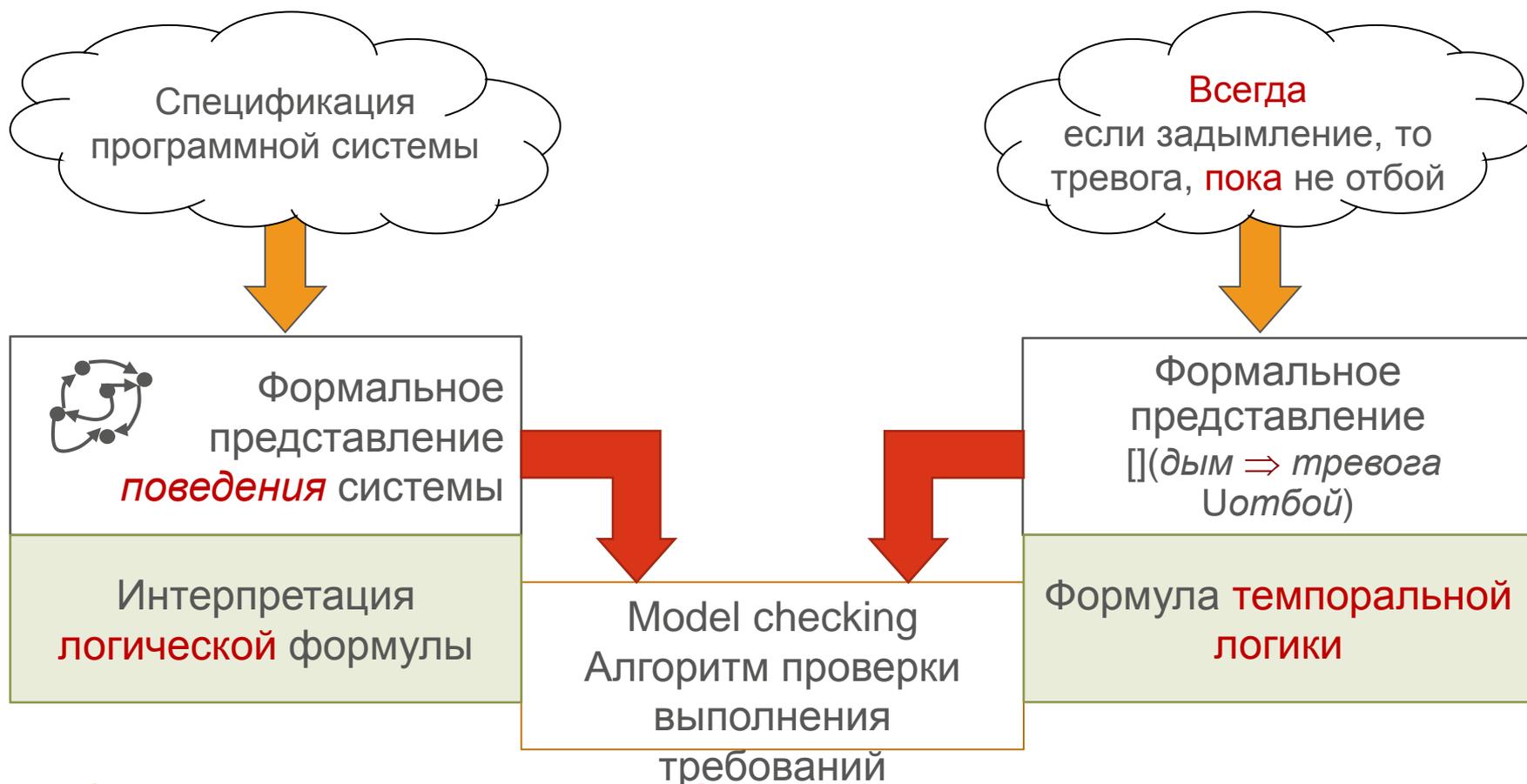
**НЕВОЗМОЖНО !**

Например, вопросы изменения знаний агентов, задания норм поведения агентов в социуме и множество подобных проблем требуют тонких формальных методов их спецификации и анализа.

# КАК МОЖНО ГАРАНТИРОВАТЬ, ЧТО ТРЕБОВАНИЯ БУДУТ ВЫПОЛНЯТЬСЯ ПРИ ПОВЕДЕНИИ ПРОГРАММНЫХ АГЕНТОВ?



- **Model checking** – алгоритм проверки того, что логическая формула на данной интерпретации выполняется - принимает значение “истина”



*ВСЕ алгоритмы управления должны проверяться относительно их БУДУЩЕГО поведения во времени*

- Модальность (от лат. *modus* – вид, способ, наклонение) – *оценочная категория, определяющая отношение высказывания к действительности. Модальная логика - любая формальная логическая система, в которой присутствуют модальные операторы.*

–  $\diamond p$  - “*p* возможно (может быть) истинно”  
 –  $\Box p$  - “*p* обязательно (должно быть) истинным”

} *Логика возможности/необходимости*

–  $\diamond p$  – “*когда-нибудь в будущем p будет истинным*”  
 –  $\Box p$  – “*p будет истинным всегда в будущем*”

} *Временная логика*

–  $\Box_A p$  – “*агент A знает, что p истинно*”  
 –  $\diamond_A p$  – “*агент A предполагает, что p истинно*”  
 $\Box_i$ : “*идет дождь*” – “*агент i знает, что идет дождь*”.

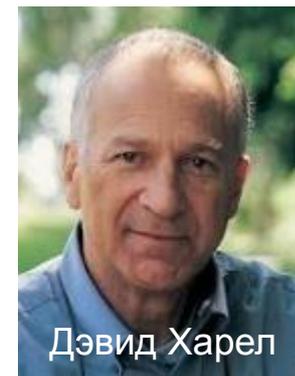
} *Логика знаний*

- МЛ – формализм для изучения *логически связанных пар модальностей* типа “необходимо” и “возможно”, которые могут стоять перед утверждениями (высказываниями) с возможными значениями “истина” или “ложь”.  
Формально эти модальности обозначаются операторами  $\Box$  и  $\Diamond$ .

Логика	Модальность необходимого $\Diamond p$	Модальность возможного $\Box p$
Алетическая alethic	$p$ <b>необходимо</b> (p is necessary)	$p$ <b>возможно</b> (p is possible)
Доксастическая doxastic	$p$ <b>обязательно</b> (p is believed)	$P$ <b>предположительно</b> (p is supposed)
Темпоральная temporal	<b>всегда в будущем</b> $p$ (p always)	<b>иногда в будущем</b> $p$ (p sometimes)
Деонтическая deontic	$p$ <b>должно быть</b> выполнено (p is obligatory)	$p$ <b>разрешается</b> выполнить (p is allowed)
Эпистемическая epistemic	<b>знаю, что</b> $p$ (p is known)	<b>думаю (полагаю), что</b> $p$ (p is believed)
... ..	... ..	... ..

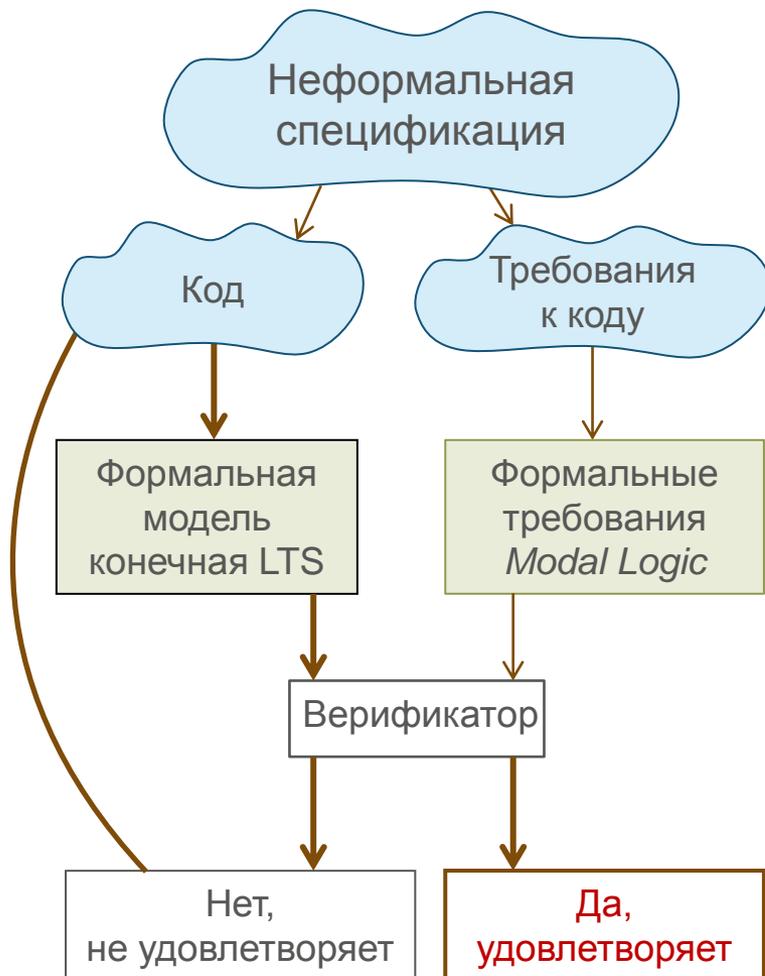
David Harel:

*“хотя единственного средства (серебряной пули) нет, но последовательное внедрение новых идей может внести существенный вклад в разработку сложных программных систем”.*



При разработке мультиагентных систем  
**полностью описать требования к ним невозможно!!!**  
получение новых теоретических результатов,  
разработка и внедрение практических технологий  
требует новых знаний

## Классический подход



## Синтез супервизоров



- Лекция 1. Мотивационная
- Лекция 3. Модальные логики
- Лекция 4. Структуры Крипке
- Лекция 5. Реляционная семантика модальных логик
- Лекция 6. Аксиоматические системы
- Лекция 7. Модальные аксиоматические системы
- Лекция 8. Аксиоматика эпистемической модальной логики
- Лекция 9. Мультимодальная логика знаний
- Лекция 10. Темпоральные модальные логики
- Лекция 11. Темпоральные логики и анализ дискретных динамических систем
- Лекция 12. Знание в многоагентных системах
- Лекция 13. Динамические эпистемические логики
- Лекция 14. Технологии разработки МАС





*Proc. of the 16th Int. Conf. AAMAS'17)*

## A Path in the Jungle of Logics for Multi-agent System: On the Relation between General Game-playing Logics and Seeing-to-it-that Logics

Emiliano Lorini  
IRIT-CNRS, Toulouse University, France  
Emiliano.Lorini@irit.fr

François Schwarzenruber  
ENS Rennes, France  
francois.schwarzenruber@ens-rennes.fr

### ABSTRACT

In the recent years, several concurrent logical systems for reasoning about agency and social interaction and for representing game properties have been proposed. The aim of the present paper is to

produces the result. For example, in this logic one can represent the fact that an agent 1 kills another agent 2 (agent 1 sees to it that agent 2 is dead) but one cannot represent the fact that agent 1 kills agent 2 by deciding to shoot. To overcome this limitation, DLA has been introduced by IJOL as a variant of STIT in which agents' choices

*“The aim of the present paper is to put some order in this ‘jungle’ of logics ...”*

*Halpern and Vardi consider **ninety six logics**. In all cases the complexity of the satisfiability problem ranges between the intractable exponential time and the undecidable ... (J.Y. Halpern and M.Y. Vardi. The complexity of reasoning about knowledge and time.//J. Comput. Syst. Sci.,*

Типичная фраза: *“The basic **logic** of belief, information acquisition, and trust (BIT Logic) is an extension of the traditional **logic**, which is in turn a multi-agent version of the normal modal **logic**... “*

