О понятии “кибернетической машины” у Лукаса.

Андрей Родин (ИФ РАН и СПбГУ)

Искусственный интеллект: парадоксы и противоречия (60 лет статье Лукаса 1961)

13 октября 2021

Утверждение:

Аргумент Лукаса ошибочен (хотя его заключение, по всей видимости, истинно), поскольку Лукас некорректно применяет понятие “механизма” к современным вычислительным устройствам и их математическим прототипам.

В противоположность тому, что утверждает Лукас, старая дискуссия о детерминизме и свободе воли не переводится (по крайней мере тем простым способом, который предлагает Лукас) в контекст формальных систем и вычислений.

Пример:

Классическое исчисление предикатов обладает свойством семантической **полноты** (всякая истинная формула выводима, т. е., “доказуема” в этом исчислении), но при этом **не разрешимо** (не существует универсального алгоритма, т.е., “механического” способа ответить на вопрос истинна ли данная произвольная (правильно построенная) формула этого исчисления, или ложна.

Компьютер можно научить выводить некоторые истинные формулы исчисления предикатов, но всегда найдется такая формула из этого класса, про которую компьютер (без дополнительной помощи человека) не сможет сказать, истина она или ложна. Свидетельствует ли это обстоятельство о превосходстве человеческого ума над компьютером?

Пример (продолжение):

Аргумент Лукаса “от неполноты” в данном случае нерелевантен: исчисление предикатов обладает свойством полноты.

Я думаю, что в той системе понятий и метафор, которую использует Лукас, на поставленный вопрос нельзя ответить, поскольку эта система понятий не позволяет аккуратно различить полноту и разрешимость.

Лукас 1961: основной тезис:

GODEL'S Theorem seems to me to prove that Mechanism is false, that is, that minds cannot be explained as machines.

Лукас 1961: мотивация:

If the proof of the falsity of mechanism is valid, it is of the greatest consequence for the whole of philosophy. Since the time of Newton, the bogey of mechanist determinism has obsessed philosophers.

If we were to be scientific, it seemed that we must look on human being as determined automata, and not as autonomous moral agents;

if we were to be moral, it seemed that we must deny science its due, set an arbitrary limit to its progress in understanding human physiology, and take refuge in obscurantist mysticism.

Кибернетическая машина

Лукас: Gödel's theorem must apply to cybernetical machines, it is of the essence of being a machine, that it should be a *concrete instantiation of a formal system*.

Это сомнительное утверждение. Аналогия формальной системы (к которой применимы теоремы Геделя о неполноте) это **язык** программирования, а не современный (после Konrad Zuse 1941) пере-программируемый компьютер (hardware), а также не компьютер, который выполняет некоторую фиксированную программу или набор таких программ.

Кибернетическая машина

Мне неясно, проводит ли Лукас различие между последними двумя понятиями. Иногда возникает впечатление, что под кибернетической машиной Лукас понимает устройство подобное автоматическим станкам первого поколения, которые выполняют фиксированный набор действий по жестко фиксированной программе, которую невозможно изменить.

Шахматная аналогия формальной дедуктивной системы: свод правил игры в шахматы (включающие начальную позицию и правила ходов), а не запись отдельной партии, и не программа для игры в шахматы.

Кибернетическая машина

We understand by a cybernetical machine an apparatus that performs a set of operations according to a definite set rules. Normally we “programme” a machine: that is, we give it instructions about what it is to do in each eventuality; and we feed in the initial "information"on which the machine is to perform its calculations.

Our idea of a machine is just this, that behaviour is completely determined by the way it is made and the incoming “stimuli”.

Замечание:

Аргументы Лукаса о космических лучах и других случайных внешних факторах, которые могут повлиять на работу вычислительного устройства непредсказуемым образом, нерелевантны.

Основной “непредсказуемый” (для внешнего наблюдателя) фактор, который ключевым образом влияет на работу компьютера — это программист, который создает и запускает на компьютере ту или иную программу (а также вводит те или иные исходные данные, которые обрабатываются этой программой).

Ответы Лукаса на аргументы Тьюринга 1950:

Тьюринг: Математический аргумент в духе Гёделя (пример геделевого предложения) можно формализовать и реализовать на компьютере.

Лукас (ответ и усиленная формулировка основного аргумента):

However complicated a machine we construct, it will, if it is a machine, correspond to a formal system, which in turn will be liable to the Gödel procedure for finding a formula unprovable-in-that-system. This formula the machine will be unable to produce as being true, although a mind can see that it is true. And so the machine will still not be an adequate model of the mind. We are trying to produce a model of the mind which is essentially "dead" but the mind, being in fact "alive", always go one better than any formal, ossified, dead system can. Thanks to Godel's theorem, the mind always has the last word.

Комментарий:

“Последнее слово всегда остается за [человеческим] умом” только постольку, поскольку человек контролирует компьютер (в частности, создавая и запуская различные вычислительные программы), а не наоборот. Такой контроль в некоторых аспектах оказывается проблематичным, но теоремы Гёделя не имеют прямого отношения к этому вопросу. Ср. ситуацию с исчислением, которое полно, но неразрешимо.

Ответы Лукаса на аргументы Тьюринга 1950:

Тьюринг: При высоком уровне сложности поведение вычислительного устройства становится непредсказуемым.

Лукас: В этом случае машины станут умными

Although it sounds implausible, above a certain level of complexity, a machine ceased to be predictable, even in principle, and started doing things on its own account or, to use a very revealing phrase, it might begin to have a mind of its own. It might begin to have a mind of to have a mind of its own when it was no longer entirely predictable and entirely docile, but was capable of doing things which we recognized as intelligent.

Комментарий 1:

Рассмотрим в качестве примера алгоритм вычисления десятичных знаков числа $\pi = 3,1415926\ldots$. Колмогоровская сложность этой последовательности относительно низкая (базовый алгоритм несложный и был известен уже Архимеду), однако процесс вычисления можно назвать полностью непредсказуемым в том смысле, что, например, нельзя заранее ответить на вопрос появится ли в будущем в результате данного вычисления последовательность 0123456789 (если она пока не появилась).

Другой класс подобных примеров: программные генераторы паролей и псевдо-случайных чисел.

Комментарий 2:

Ситуацию можно сравнить с детерминированным механическим хаосом (например, в проблеме трех тел), однако в данном случае речь не идет о проблеме устойчивости предсказаний к малым изменениям начальных данных. Просто устроенная “кибернетическая машина” может вести себя сложно и непредсказуемо. Механические интуиции Лукаса про простые и предсказуемые “мертвые механизмы” и сложные и непредсказуемые “живые умы” оказывается обманчивой. Сложные умы, в свою очередь, часто ведут себя вполне предсказуемо.

Заключение 1:

Понятие “кибернетической машины” у Лукаса представляет собой неправомочный перенос понятий из (популярной) классической механики в вычислительный контекст. Это плохо построенное понятие оказывается не чувствительным к различию между полнотой и алгоритмической разрешимостью формальной системы. Аргумент Лукаса 1961, который в качестве посылки использует первую теорему Гёделя о неполноте, не имеет силы.

Заключение 2:

Попытки строить вычислительные модели человеческого мышления (и мышления животных) неправомерно отбрасывать, используя аргумент Лукаса в качестве основания. Но нельзя исключить, что найдутся другие основания для отказа от таких моделей.

Заключение 3:

На мой взгляд, статья Лукаса 1961 — это пример неправомерного использования математического аргумента как основания для философского аргумента. Теоремы Гёделя о неполноте формальной арифметики не имеют тех далеко-идущих философских следствий, которые им приписывает Лукас и некоторые другие философы.

Сам Курт Гёдель не делал такого рода обобщающих философских заключений из своих теорем.

СПАСИБО!