

Знания и их представление в компьютерную эпоху¹

Участники: С.П. Ковалёв и А.В. Родин

Организатор проекта и ведущая – Ю.В. Синеокая

Ковалёв Сергей Протасович, доктор физико-математических наук, ведущий научный сотрудник Института проблем управления им. В.А. Трапезникова РАН, Россия, 117997, Москва ул. Профсоюзная, д. 65, email: kovalyov@nm.ru

Родин Андрей Вячеславович, кандидат философских наук, старший научный сотрудник Института философии РАН, Россия, 109240, г. Москва, ул. Гончарная, д.12, стр.1., email: andrei@philomatica.org

Синеокая Юлия Вадимовна – доктор философских наук, профессор РАН, заведующая сектором истории западной философии. Институт философии РАН. Российская Федерация, 109240, г. Москва, ул. Гончарная, д. 12, стр. 1; e-mail: sineokaya@iph.ras.ru

Аннотация:

Представление знаний (ПЗ) – это область исследований, которая находится на стыке компьютерной науки и философской эпистемологии. В техническом смысле ПЗ можно характеризовать как раздел исследований по искусственному интеллекту (ИИ), который связан с разработкой информационных и коммуникативных систем решающих задачу эффективного использования больших объемов данных. С другой стороны, знания и их представление – это одна из центральных тем философской рефлексии и анализа; это замечание относится как к традиционной философии, так и к аналитической философии 20-го века, которая существенно опирается на формальные логические подходы. 8 декабря 2016 года в рамках цикла философских бесед «Реплики», совместного проекта Института Философии РАН и библиотеки им. Ф.М. Достоевского «Анатомия философии: как работает текст» проблему представления знаний обсудили философ-эпистемолог А.В. Родин и представитель компьютерной науки С.П. Ковалёв. Участники этой беседы с разных точек зрения обсудили новые эпистемические практики использующие компьютерные технологии представления знаний.

Ключевые слова: представление знаний, искусственный интеллект, нейронные сети, онтологии, экспертное знание.

ЮС: Дорогие друзья, мы начинаем новую встречу в цикле “Реплики” в рамках совместного проекта Института философии Российской академии наук и библиотеки им. Ф.М. Достоевского «Анатомия философии: как работает текст». Мы собираемся на Чистопрудном бульваре в доме 23 каждый четверг и обсуждаем одну из насущных философских проблем. Уже больше года мы встречаемся в формате диалога. Сегодня у нас будет диалог двух людей – философа Андрея Родина и специалиста в области компьютерных наук Сергея Ковалева, которые представляют

¹ Эта работа подготовлена при поддержке гранта РГНФ 16-03-00364, в котором участвуют С.П. Ковалев и А.В. Родин

два института Российской академии наук: институт философии и институт проблем управления. Я передаю микрофон Андрею, мы начинаем.

AP: Знание - это общее и абстрактное понятие, которое является предметом исследования для философской дисциплины эпистемологии. Понятие *представления* знаний указывает на связанные со знаниями формальные и материальные аспекты. Как я сейчас покажу, компьютерная революция последних десятилетий затрагивает оба эти аспекта представления знаний и требует серьезного переосмысления привычных эпистемических практик. Я начну с анализа материального аспекта проблемы и потом немного скажу о формальном.

По крайней мере для людей моего поколения самым привычным материальным носителем знаний является книга и другие виды печатных текстов. Но идея о том, что печатные тексты – это своего рода сосуды, в которых содержится нематериальная субстанция, которую мы называем знанием, является, конечно, очень наивной. Книгами мы называем не просто материальные объекты, обладающие определенными физическими характеристиками, а объекты, которые мы определенным образом используем. Подобным образом устроены такие обиходные понятия как “молоток”, “посуда” и многие другие. Знание не “содержится” в книгах в буквальном смысле; скорее можно сказать, что книга - это инструмент вроде молотка, который используется в сложном социальном процессе, включающем в себя обучение, исследования, проверку, распространение, применение и другие эпистемические (то есть, связанные со знаниями) процедуры. Книги также используются других целей, которые не связаны напрямую со знаниями - в частности, для эстетических целей; такие способы использования книг мы в этом разговоре оставим в стороне.

Итак, чтобы понять, каким образом знания представлены в книгах, нужно прежде всего обратить внимание на то, как люди пользуются книгами. Говоря о способах использования книг, я имею в виду не только индивидуальные практики чтения и написания книг, но и социальные практики и институты, такие как издательства, редакционные коллегии, научные и учебные библиотеки, исследовательские институты. Хотя базовые характеристики всей этой сложной инфраструктуры определяются многими факторами, технологические факторы, связанные с книгопечатанием, вносят свой вклад и накладывают определенные ограничения на возможные инфраструктурные решения. Наши привычные образовательные, исследовательские и просветительские институты (в частности, публичные библиотеки) приспособлены именно для работы с книгами и другими печатными текстами. С книжной культурой, по-видимому, тесно связано и традиционное деление знаний на различные разделы и дисциплины, которое отражается в институциональных делениях между факультетами университетов и между различными академическими специальностями. Я, конечно, не утверждаю, что особенности технологии книгопечатания полностью определяют дисциплинарную и институциональную структуру традиционных образовательных и исследовательских институтов. Но я указываю на то, что в рамках такой традиционной институциональной структуры эпистемические задачи решаются именно с помощью книг и других печатных текстов, и что этот материальный аспект традиционной эпистемической практики нельзя игнорировать.

Важный вопрос состоит в том, в какой степени традиционные эпистемические институты позволяют нам реализовать новые возможности, которые предоставляют компьютерные технологии. Я говорю сейчас только о таких возможностях, реализация которых способствует достижению эпистемических целей. Вопрос о том, каковы именно эти цели, в свою очередь является открытым и заслуживает обсуждения в каждом конкретном случае. Я не думаю, что

всякая новая технология требует повсеместного внедрения только потому, что она новая. Тем не менее, нетрудно найти примеры того, как компьютерные технологии помогают более эффективно достигать таких общепризнанных эпистемических целей, как обеспечение массового открытого доступа к существующим знаниям и обеспечение обмена знаниями между экспертами разного профиля. Наверняка найдутся люди, которые будут оспаривать тезис о том, что обеспечение открытого доступа к знаниям любого рода является желательным, но я сейчас не буду начинать дискуссию по этому вопросу и приму этот тезис в качестве рабочей гипотезы.

Для ученого-гуманитария вроде меня самая очевидная польза компьютера и интернета состоит в том, что эти вещи позволят делать привычные нам вещи более эффективно, то есть с меньшими усилиями. Я имею в виду создание, редактирование, публикацию и распространение текстов и обмен текстами в электронной форме. Хотя за эти преимущества приходится платить временем и усилиями, потраченными на приобретение компьютерной грамотности, эти усилия очень быстро окупаются. На первый взгляд, в подобных случаях новые технологии не приносят ничего существенно нового и ничем не угрожают нашим привычным институтам и практикам: мы делаем то же, что делали и в докомпьютерную эпоху, но при этом тратим меньше времени на техническую рутину вроде переписывания и перепечатаывания черновых вариантов текста, выписывания цитат из книг в читальном зале библиотеки и проч. Однако это только на первый взгляд. Википедию нельзя назвать академическим проектом, но она на сегодняшний день является очень эффективным инструментом для представления и распространения знаний. Главным отличием Вики от традиционных бумажных энциклопедий является то, что у нее нет ни главного редактора, ни редакционной коллегии, ни специально выбранной группы авторов: роль авторов и редакторов выполняют активные пользователи. Администрация Вики минимальна (главная ее задача - борьба со злонамеренным вандализмом, а не контроль качества статей) и по большей части также состоит из волонтеров. Ничего подобного Вики невозможно было бы организовать, используя традиционную бумажную технологию представления знаний. Таким образом, Вики - это пример того, как развитие материальной базы технологий представления знаний приводит к возникновению новых социальных форм эпистемических практик. Другим таким примером является бурное развитие MOOC (Massive Open Online Courses), которое по всей видимости обещает очень существенно видоизменить существующие сегодня образовательные институты, в первую очередь университеты, обеспечив доступ к качественному образованию миллионам и миллиардам молодых людей из бедных стран, которые сегодня не имеют такого доступа.

Разумеется, новые электронные технологии представления и распространения не только открывают новые возможности, но и ставят новые проблемы. Я скажу несколько слов только об одной проблеме, которой, на мой взгляд, пока уделяется недостаточно внимания. Что будет с нашими электронными публикациями через 50, 100 и 500 лет? Технологический прогресс в наши дни происходит с такой скоростью, что предсказывать будущее очень трудно. Но мой вопрос - не из области футурологии. В хранилищах и читальных залах наших библиотеках есть книги и рукописи, появившиеся на свет в разное время. Ответственность библиотекарей перед будущими поколениями состоит в том, чтобы все эти тексты, включая самые старые, были по-прежнему доступны в последующие века и эпохи в неизменном виде. Важно сохранить и передать в пользование нашим потомкам не только сами тексты, перенесенные на новый носитель, но и материальные артефакты: старые книги и рукописи. Дело не только в романтических эмоциях, которые многие переживают прикасаясь к старым предметам, но и в том, что материальные артефакты представляют собой самые надежные исторические свидетельства времени, места и других обстоятельств собственного происхождения, которые в этом случае можно проверить с помощью радиоизотопного анализа и других объективных методов. Как подобная задача может и должна решаться в случае электронных текстов и других видов медиа (аудио- и видеозаписи), остается неясным. У человечества пока нет такого опыта. Существующие технологии позволяют хранить большие объемы информации в компактном виде, но вопрос о том, насколько такое хранение является надежным на

исторических временах порядка столетий, остается пока открытым². Под надежностью хранения информации в этом случае нужно понимать не только идентичность во времени последовательностей битов на подходящем для этих целей носителе, но и какие-то гарантии того, что ключи к расшифровке и правильной интерпретации этих последовательностей битов не будут утеряны в ходе последующих технологических революций и возможных исторических катаклизмов.

Возможность легко исправлять и заново переписывать тексты, которую предоставляют нам компьютерные технологии, не во всех ситуациях оказывается эпистемическим преимуществом. Я, тем не менее, думаю, что в будущем будут найдены технические решения, которые позволят сочетать простоту ревизии знаний с надежностью их сохранения и таким образом облегчат работу для будущих поколений историков.

До сих пор мы говорили о знаниях, представленных в традиционной текстовой форме на том или ином естественном языке, возможно расширенном специальным профессиональным жаргоном, символьными обозначениями, подобным тем, которые используются в математике, а также чертежами и другими изображениями. Однако значение современных информационных технологий для представления знаний состоит не только в том, что эти технологии позволяют эффективно создавать, копировать и распространять традиционные тексты. Компьютерные технологии дают также широкие возможности для развития новых форматов представления знаний, включая создание специально предназначенных для этой цели искусственных языков. Здесь мы касаемся темы представления знаний в более специальном и техническом смысле слова: в компьютерных науках так называют специальную область исследований по Искусственному Интеллекту (ИИ), в которой рассматривается теория программных средств для представления знаний. Такие средства уже реализованы и используются в некоторых областях техники. Не входя в вопросы технического характера, в которых Сергей более компетентен, я только замечу, что и в этой специальной сфере важную роль может играть эпистемологическая экспертиза.

Базовый элемент Представления Знаний (ПЗ) - это спецификация *онтологии* (это тоже термин используемый в компьютерных науках, а не только в философии!), то есть спецификации основных типов объектов данной области знаний, отношений между объектами, правил и аксиом регулирующих объекты и отношения и, возможно, других подобных принципов и элементов. Для задач ПЗ важно, чтобы онтология данной области знаний была сформулирована не просто в общих словах на естественном языке, но также формально с использованием символьного логико-математического формализма, который далее сможет быть реализован в программном коде. Слушатели, знакомые хотя бы в общих чертах с аналитической философией 20-го века, сразу заметят, что использование одного и того же термина “онтология” в философии и компьютерных науках не случайно: начиная по крайней мере с “Философии логического атомизма” Рассела (1918 г.) задачи формального представления знаний ставились и по-разному решались в рамках этой философской традиции. Сегодня формальная онтология – это та область, где философы, логики и специалисты в компьютерных науках активно сотрудничают; в числе пионеров такого сотрудничества со стороны философии можно назвать Барри Смита (Barry Smith) из университета Буффало.

Аналогично дело обстоит с формальной эпистемологией – более новой философской дисциплиной, которая пользуется формальными логическими методами для решения традиционных и новых эпистемологических проблем. Хотя связь между формальной философской логикой и работами по искусственному интеллекту и компьютерному

² Проект направленный на решение этой проблемы осуществляется в настоящее время под эгидой Российского фонда перспективных исследований в лаборатории лазерного наноструктурирования стекла Российского химико-технологического университета им. Менделеева под руководством профессора Владимира Сигаева. Сайт проекта: <http://fpi.gov.ru/activities/lab/rxtu>

представлению знаний кажется очевидной, на самом деле философски мотивированные логические схемы и конструкции очень редко можно применить на практике, реализуя их напрямую в виде программного кода. Дело в том, что философски мотивированная формальная логика в 20-м веке как правило не принимала в расчет ограниченность вычислительных ресурсов и ресурсов памяти, а ставила и решала свои проблемы “в принципе”, то есть с помощью сильных идеализаций. Исключением является конструктивное направление в логике, которое с самого начала было мотивировано идеей вычислительной реализации. Можно сказать, что в современных исследовательских работах в области ПЗ и в практических технологических проектах, относящихся к этой области, многие теоретические идеи и подходы, которые были ранее разработаны в рамках логически ориентированной аналитической онтологии и эпистемологии, проверяются на практике. Как обычно бывает в подобных случаях, практика оказывается во многих отношениях богаче теории и дает новый материал для теоретических построений и философских рефлексий. Таким образом, исследование формальных аспектов представления знаний оказывается той областью, где философы и специалисты в области компьютерных наук могут продуктивно сотрудничать. Наш общий с Сергеем Ковалевым проект “Логические и эпистемологические аспекты конструктивного знания”, поддержанный Российским государственным гуманитарным фондом (код проекта 16-03-00364), может служить примером такого сотрудничества.

В заключении этого введения я хочу еще раз подчеркнуть, что социальные формы, в которых существуют человеческие знания, тесно связаны с форматами представлений этих знаний. Я уже говорил о значении книжных форматов для традиционных образовательных и исследовательских институтов, таких как школы, академии и университеты. Контуры новых институтов, адекватные новым формам представления знаний, сегодня только начинаются прорисовываться. Эту метафору не нужно понимать так, что речь тут идет о каком-то естественном процессе, за которым мы наблюдаем со стороны. Речь идет о социальном, техническом и научном прогрессе, в котором мы все тем или иным образом участвуем. Будущее наших институтов и наших технологий в той или иной мере зависит от усилий каждого из нас: даже когда мы выступаем только в роли потребителей, мы выбираем между разными сервисами, и совокупный эффект таких потребительских выборов благодаря экономическим и другим социальным механизмам может быть очень значительным. Именно поэтому подобные вопросы важно, на мой взгляд, обсуждать публично. Я надеюсь, что наша сегодняшняя встреча поможет найти общее понимание эпистемологических проблем и сюжетов, связанных с развитием информационных технологий.

СК: Целый ряд вопросов, которые затронул Андрей, демонстрирует наличие одной крупной проблемы, которую я сейчас сформулирую. Проблема состоит в неконтролируемом росте количества знания. Современное состояние человечества отличается от исторического прошлого, в числе прочего, такими факторами, как увеличение количества людей, увеличение объема свободного времени, увеличение доступности образования, увеличение пропускной способности коммуникационных каналов. Все эти факторы способствуют повышению объема производства знаний. На протяжении всей истории люди считали, что чем больше у них знаний, тем лучше. Что мы имеем в итоге? Сегодня знания имеют тенденцию к размножению, к катастрофическому количественному росту, который также, надо признать, во многих случаях влечет за собой рост качества знаний. Но при этом отсутствуют механизмы обратной связи, которые позволяли бы контролировать этот процесс и делать его целенаправленным, делать знания полезными в широком смысле этого слова. Эту ситуацию можно сравнить с попаданием кроликов в Австралию. В Европе у кроликов есть естественные враги, такие как волки и лисы, благодаря которым европейская популяция кроликов остается ограниченной. Когда кроликов в середине 19 века завезли в Австралию, где у них не оказалось естественных врагов, они начали катастрофически размножаться: их стало намного больше, чем людей, которые могли бы их истреблять и ими питаться. Это привело к экологической катастрофе в масштабах целого континента. Есть другие примеры подобных экологических катастроф.

Ситуация во многих областях знания сегодня схожая. Имеется огромное количество знаний, которые мало кто способен “переварить”, оценить и тем более применить. Я приведу пример из математики. Речь идет о классификации всех простых конечных групп - результате, который был получен группой из примерно 100 математиков (часть из которых работала совместно, а часть независимо друг от друга) и который опубликован в виде нескольких сотен журнальных статей в период с 1950-х до начала 2000-х годов. Если сформулировать этот результат в виде одной теоремы, то ее доказательство займет около 15 000 печатных страниц математического текста³. Проблема состоит в том, что это доказательство могут понять едва ли несколько человек в мире, причем ни один из них не способен проверить все шаги этого доказательства. Несмотря на такие трудности, эти несколько человек договорились о том, чтобы признать данную классификацию конечных простых групп правильной. Можно согласиться с тем, что эта классификация - значительное достижение в области чистой математики. Но давайте спросим себя: зачем оно нужно? Если бы эту теорему можно было бы включить в обычный университетский учебник, то студенты, изучающие эту теорему, могли бы пытаться применить ее в новых компьютерных технологиях или где-то еще. Если бы эта теорема была доступна экспертам в разных областях знаний - будь то космология, археология или какая-то другая область - они могли бы пытаться ее использовать для решения задач в своих специальных областях. Но в данном случае это, очевидно, невозможно. Как и в случае кроликов в Австралии, причина этой проблемы состоит, на мой взгляд, в отсутствии обратной связи, которая бы регулировала количество и полезность знания в соответствии с той или иной шкалой ценностей. При отсутствии такой обратной связи неконтролируемое развитие может иметь гораздо больше отрицательных и вредных последствий, чем положительных и полезных.

AP: Я по старинке считаю, что “чем больше знаний, тем лучше”, и не считаю, что бурный рост количества знаний составляет какую-то проблему сам по себе. Другое дело, что такой быстрый рост приводит к тому, что существующие эпистемические институты и практики не справляются с этим ростом и нуждаются в модернизации и реформировании. Разумеется, ситуация, при которой математическая или любая другая научная теория доступна пониманию только для небольшой группы экспертов, тогда как все остальное научное сообщество и другая публика вынуждена доверять этим экспертам, не имея никакой возможности проверить их суждения и выводы, не является нормальной. Она возвращает нас к донаучным, псевдонаучным и просто ненаучным формам знания, таким как “сакральное знание”, которое предполагает, что знаниями во всей полноте владеют только члены специальной касты жрецов, тогда как все другие члены общества могут пользоваться этими знаниями только под непосредственным руководством жрецов при условии безусловного признания их авторитета. Открытость и публичность – это фундаментальные характеристики именно научного знания, которую подчеркивали многие философы прошлого, включая, например, Фрэнсиса Бэкона. Мой общий энтузиазм по поводу эпистемологического значения новых компьютерных технологий связан именно с этим моментом: поскольку эти технологии обеспечивают знаниям большую публичность и открытость и дают дополнительные возможности для коммуникации, их следует считать благом для науки, несмотря на сопутствующие новые проблемы, с которыми мы при этом сталкиваемся и которые нам приходится решать.

Я также разделяю более специальную озабоченность, о которой сказал Сергей в связи с тем же примером: для прогресса науки и технологии имеет значение не просто доступность новых знаний широкой публике, но и доступность этих знаний экспертам разного профиля, которые могут использовать эти знания в областях своей компетенции. Такое “перекрестное опыление” между разными областями знания - это важнейший механизм прогресса, который, конечно, перестает работать, когда какое-то знание оказывается доступным только узкой группе специалистов.

3) <http://mathworld.wolfram.com/ClassificationTheoremofFiniteGroups.html>

Я готов называть только что описанные проблемы “проблемами роста” или даже “проблемами взрывного роста” знаний, но я не считаю, что подобные проблемы нужно решать с помощью какого-то регулирования и ограничения этого роста или с помощью подчинения научного прогресса каким-то внешним целям, которые позволят нам отличить полезное знание от бесполезного, ограничить рост знания и направить этот рост в нужное нам русло.

Биологическая аналогия, которую использовал Сергей, на мой взгляд применима только внутри науки или даже только внутри каждой отдельной узкой научной дисциплины. Конкуренция научных теорий и инженерных решений действительно играет важную роль в развитии любой области знания: каждая научная теория и каждая инженерная идея проходят проверку временем, конкурируя с альтернативными теориями и идеями; в этом отношении аналогия с дарвиновским “выживанием самых приспособленных” представляется уместной. Однако Сергей использует эту аналогию иначе, говоря о том, что процесс получения новых знаний (то есть познание) должен быть подчинен каким-то высшим целям социального и этического характера. Знания тут играют роль кроликов, а социальные цели – роль хищников, которые питаются этими кроликами. Я допускаю, что социальные интересы и этические принципы могут входить в конфликт с эпистемическими интересами в определенных ситуациях, но на мой взгляд отсюда не следует, что вторые должны быть подчинены первым. Кроме того, я не знаю примеров из истории науки, когда конфликты такого рода были продуктивными. Таких конфликтов, на мой взгляд, лучше по возможности избегать.

Можно также дать прагматический аргумент в пользу автономии науки и против идеи регулировать развитие знаний с помощью каких бы то ни было внешних целей. Ситуация, когда знания, которые впоследствии применяются в технике и таким образом оказываются полезными, были получены с использованием только внутренних мотиваций (таких как ответа на открытый математический вопрос) - это скорее правило, чем исключение. С позиции научного реализма объяснить такое положение вещей нетрудно: чтобы как-либо использовать знания в собственных интересах, нужно сначала разобраться с тем, как устроен мир безотносительно к любым социальным и практическим интересам, включая общечеловеческие. Если бы наши знания о мире были всегда ограничены теми аспектами мира, которые непосредственно связаны с текущими практическими интересами исследователя или его социальной группы (в качестве упражнения я рекомендую слушателям попробовать составить список таких интересов для сообщества наших предков, живших в эпоху палеолита), то такие знания не могли бы в принципе обеспечить технический и социальный прогресс человечества в том виде, в котором он в действительности имел место на протяжении последних веков и тысячелетий. Я сознаю, что отказ от прогресса тоже может быть продуманной позицией, но хочу подчеркнуть, как много тут ставится на карту.

Что касается примера математического доказательства длиной в 15 000 страниц, то с ним также связана проблема, о которой Сергей не говорил. Если это доказательство способны понять несколько человек в мире, и никто не способен его аккуратно проверить, то откуда у нас может быть уверенность в том, что это доказательство не содержит ошибок? Почему мы должны считать эту теорему надежно установленным математическим результатом? Я думаю, что ключ к решению этой проблемы, а также многих из тех проблем, о которых говорил Сергей, лежит в математической теории и связанной с этой теорией компьютерной технологии. В истории математики есть примеры того, как ранее построенные теории удавалось существенно упростить и таким образом сделать более доступными для всех заинтересованных людей с помощью аксиоматизации, как это случилось с теорией топосов, которая сначала была построена в неформальном виде Александром Гротендиком в конце 1950-х, а затем аксиоматизирована Биллом Лавером в начале 1970-х. В наши дни профессор Принстонского института наш соотечественник Владимир Воеводский является лидером проекта, который он назвал “унивалентными основаниями математики”. Речь идет о создании стандартного формализма, который позволит записывать математические доказательства в виде

компьютерного кода и проверять корректность этих доказательств с помощью компьютера⁴. В настоящее время уже удалось формализовать таким образом доказательства ряда важных математических теорем, и вполне вероятно, что у 15 000-страничного доказательства, о котором упомянул Сергей, также вскоре будет формальная компьютерная версия, допускающая такую проверку. Такой путь решения проблемы длинных доказательств и других аналогичных эпистемологических проблем мне кажется и более эффективным, и более оправданным с эпистемологической точки зрения, чем регуляция науки социальными ценностями или другими внешними регуляторами.

СП: С последним утверждением я согласен: регулирование в экологических системах, на которое я указал в качестве аналогии, также является внутренним, а не внешним. Но сейчас я хотел бы выйти за пределы круга тех проблем, которые мы обсуждаем сейчас. В развитии технологий представления знания существует “развилка”, которая также имеет не только техническое, но и эпистемологическое значение. Речь идет о вопросе, который может быть сформулирован следующим образом: хотим ли мы с помощью компьютеров воспроизводить или имитировать только *результаты* человеческой деятельности по получению и обработке знаний, или же мы хотим воспроизводить или имитировать также и те когнитивные *механизмы*, которые лежат в основе этой деятельности? Например, традиционные программные средства доказательства теорем, о которых мы говорили, воспроизводят готовые математические доказательства, но не когнитивные механизмы, которые позволяют математикам находить эти доказательства.

Однако в последние годы активно развивается альтернативный подход к компьютерному формированию и обработке знаний, который основан на идее имитации когнитивных механизмов, с помощью которых человек формирует и обрабатывает знания. Речь идет о так называемых искусственных нейронных сетях. Идея нейронной сети, которая была впервые высказана уже более 70 лет назад, состоит в следующем. Нейрон человеческого мозга можно смоделировать на компьютере в виде простого элемента, который имитирует биологическую функцию раздражимости. На входы такого элемента поступают различные сигналы (скажем, электрические), и как только суммарный уровень поступающих сигналов переходит через определенный порог, этот элемент в свою очередь выдает сигнал на выходе, то есть “откликается на раздражение”. Далее такие элементы соединяются друг с другом так, что выходные сигналы одних элементов подаются на входы других. Оказывается, что достаточно многокомпонентные и подходящим образом организованные системы такого рода могут имитировать сложные функции мозга, такие как распознавание образов. Допустим, у нас есть картинки, на некоторых из которых изображены кошки, а на некоторых – собаки. Если человека соответствующим образом проинструктировать, он может не реагировать на изображения кошек, а при предъявлении изображения собаки, например, поднимать руку. Такое поведение можно имитировать с помощью искусственной нейронной сети, состоящей из нескольких десятков или сотен нейронов, если ее соответствующим образом обучить. На входные нейроны сети нужно подавать цифровое представление изображения, а с выходных нейронов снимать сигнал ответа. Сеть в подавляющем большинстве случаев будет выдавать выходной сигнал при предъявлении изображения собаки и не выдавать такого сигнала при предъявлении изображения кошки. Процесс обучения нейронной сети выглядит следующим образом. Сначала мы разрешаем сети работать произвольно, то есть реагировать на предъявляемые изображения случайным образом. Далее собирается статистика, которая учитывает доли “правильных” и “ошибочных” реакций, и данные этой статистики используются для корректировки удельных весов соответствующих сигналов в связях между нейронами. После этого процесс повторяется. Оказывается, что при правильно подобранных алгоритмах корректировки весов доля ошибок при последующих итерациях обучения будет

4 Об аксиоматизации теории топосов и программе построения унивалентных оснований математики см. Andrei Rodin, *Axiomatic Method and Category Theory* (Synthese Library vol. 364), Springer 2014.

снижаться. При достаточном числе итераций этого процесса сеть начинает отличать кошек от собак с высокой степенью надежности. Все это кажется похожим на то, что происходит в более обычной ситуации, когда роль обучаемого играет не машина, а человек, которого хвалят за правильную реакцию и ругают за неправильную.

Первые лабораторные эксперименты с нейронными сетями были проведены более 50 лет назад. С помощью нейронных сетей исследователи пытаются, в частности, решать инженерные задачи, такие как распознавание свойств композиционных материалов на основе их структуры. Опытный машиностроитель может предсказать “на слух”, сколько прослужит та или иная конструкция, постучав по конструкции молотком. Конечно, явно сформулировать какие-то формальные основания для своего суждения такой машиностроитель не может. В начале 2000-х годов проводились эксперименты, в которых подобные компетенции удавалось до некоторой степени развить у компьютеров с помощью нейронных сетей. Однако прогресс тормозился ввиду недостатка компьютерных ресурсов – вычислительной мощности и памяти. Прорыв в этой области произошел два года назад, когда стали доступными графические ускорители и возникла возможность конструировать большие нейронные сети буквально на одном кристалле. Значительный прогресс в последние годы также имел место в области программного обеспечения, используемого при конструировании нейронных сетей. Сегодня нейронные сети уже умеют порождать осмысленные тексты на естественных и искусственных языках (включая языки программирования), распознавать тексты и речь, сочинять музыку и т. д. Такого рода примеров можно найти очень много, однако, важно ответить, что весь этот прогресс до сих пор не служит решению инженерных задач. Самая свежая работа, в которой предлагается использовать нейронные сети для предсказания свойств композиционных материалов, как ни странно, датируется примерно 2004 годом. Некоторые специалисты объясняют такого рода факты тем, что более новые работы в этой области ведутся в секретном режиме в интересах военных ведомств, однако мне такое объяснение не представляется достаточным. Я вижу проблему в том, что, как правильно сказал Андрей, существует разрыв между традиционными математическими подходами в науке вообще и в материаловедении в частности, и новыми подходами, в данном примере основанными на нейронных сетях. До сих пор в материаловедении доминирует старомодная идея о том, что свойства материалов можно наилучшим образом описать с помощью математических уравнений. Вычислительные подходы используются в первую очередь для того, чтобы решать (или в некоторых случаях порождать) эти уравнения. Таким образом, в данном случае речь идет о том, что компьютер имитирует только результаты человеческой когнитивной деятельности. Идея же о том, что компьютерные технологии могут позволить также имитировать и сами когнитивные механизмы человека, как это уже делается в компьютерных нейронных сетях, до сих пор встречает значительное сопротивление.

AP: Я хотел бы сделать замечание, которое касается идеи о том, что техника имитирует какие-то естественные процессы и их результаты, включая когнитивную деятельность человека и результаты такой деятельности. Эта идея играет важную эвристическую роль в развитии техники, но ее значение все же не нужно преувеличивать. Такой технический инструмент, как молоток, ничего не имитирует, а скорее расширяет возможности человека по преобразованию окружающей среды. С компьютерной техникой дело обстоит аналогично. Существенный вопрос, на мой взгляд, состоит не в том, в какой степени искусственные нейронные сети имитируют результаты когнитивной деятельности человека и естественные когнитивные механизмы, а в том, в какой степени и каким именно образом эти когнитивные возможности расширяются за счет использования нейронных сетей и других новых компьютерных технологий.

СП: Я с этим согласен и могу привести пример ограниченности основанного на имитации подхода к нейронным сетям. Если в деталях знать алгоритмы, по которым происходит обучение нейронной сети, то натренированную сеть можно обмануть с помощью очень небольших и незаметных для человеческого глаза модификаций предъявляемых изображения, так что эта

сеть будет принимать, например, кошку за страуса⁵. Это показывает, что системы, основанные на имитации нейронных связей, работают неустойчиво. Когда изображение анализирует человек, то наряду с биологическим механизмом, который, как можно предположить, функционирует подобным образом, важную роль играет понимание структуры изображения, расположения и связей частей изображенного объекта. Так что я согласен с тем, что попытки имитировать какую-то определенную биологическую функцию очень быстро себя исчерпывают, а также с тем, что самые интересные технические решения обычно ничего не имитируют, а имеют самостоятельную ценность. Вместе с тем, я хочу отметить, что в области представления знаний таких решений на сегодняшний день нет, так что они остаются делом будущего.

ЮС: Спасибо. Теперь переходим к вопросам и репликам слушателей.

Вопрос из зала: Можно ли сказать, что в компьютерных системах представления знаний знания всегда представляются в виде текстовых нарративов?

АР: Я думаю, что нет. Тот способ представления знаний, о котором Вы говорите, связан скорее с книжной эпохой, с идеей о том, что книга и текст на естественном языке - это универсальные инструменты для фиксации, сохранения и передачи любого рода знаний. На практике далеко не все виды знаний передавались таким способом: многие практические знания передавались и до сих пор передаются из рук в руки от более опытных людей к менее опытным. Письменные инструкции могут играть при этом какую-то роль, но не во всех случаях эта роль является главной. В качестве примера можно привести знание того, как управлять автомобилем. Если же теперь посмотреть, что происходит со знаниями в электронных коммуникациях, то даже если оставить в стороне специальные вопросы, связанные с использованием искусственных программных языков для представления знаний, можно заметить существенные отличия от стандартного книжного формата. Как я уже говорил, хотя Википедия для пассивного пользователя выглядит как обычная бумажная энциклопедия, на самом деле она устроена совершенно иначе. Главное отличие, на которое я обращаю внимание, состоит в том, что Википедия - это постоянно обновляемая динамическая структура, которая задает иную динамику формирования, распространения и ревизии знаний по сравнению с обычными бумажными энциклопедиями. Если распечатать все страницы Википедии и опубликовать их в виде одного многотомного бумажного издания, то есть в виде большого нарратива, то такой бумажный вариант Википедии просто не сможет функционировать в прежнем виде, то есть существенные характеристики архитектуры Википедии будут при этом утеряны. Если же принять во внимание более технические аспекты современных систем компьютерного представления знания, о которых говорил Сергей, то таких отличий от традиционного книжного формата можно найти еще больше.

Вопрос из зала: Не считаете ли Вы, что современный информационный коллапс связан с тем, что инженеры не обращают внимания на слова классиков, включая академика Колмогорова, Винера и Шеннона, которые обращали внимание на то, что знания и информация - это не одно и то же. Возможно, что проблема состоит в том, что это важное различие не учитывается в приложениях теории информации.

СП: Я могу Вас успокоить, обратив внимание на тот факт, что в современных компьютерных технологиях теория информации почти не используется. Поэтому даже если эта теория является некорректной, это не оказывает существенного влияния на разработку новых компьютерных технологий. Проблема скорее состоит в том, что у нас, разработчиков компьютерных систем, на сегодняшний день для теории информации просто нет подходящей замены. Я имею в виду даже не обязательно математическую, но также и философскую теорию.

5 <https://arxiv.org/abs/1312.6199>

Классическая наука индустриальной эпохи до Второй Мировой войны вполне довольствовалась тем математическим и философским базисом, который заложили Ньютон, Лейбниц и другие классики. Это позволяло успешно развивать технику, используя результаты фундаментальных наук, прежде всего физики и химии. Однако начиная со Второй Мировой войны нам пришлось иметь дело с более сложными техническими системами, включая информационные системы и сложные многокомпонентные материалы, свойства и характеристики которых не описываются стандартными математическими средствами такими как системы дифференциальных уравнений. Что касается технической стороны этой проблемы, то ее удается в какой-то степени решать за счет использования новых и или во всяком случае непривычных для техники математических теорий, таких как теория графов или теория категорий. Более существенно то, что в настоящее время не существует никакого общего философского представления о том, каким образом связаны между собой современные фундаментальные научные знания и их технические приложения. Такой философской теории нам, практикам в области создания информационных систем, катастрофически не хватает. Примерно в то же время, когда такой запрос был явно сформулирован, то есть в середине 20 века, философия по каким-то причинам сфокусировала свои усилия на решении проблем совсем другого рода - с некоторыми исключениями, которые подтверждают это правило. Во всяком случае, мои попытки найти в современной философской литературе ответы на интересующие меня философские вопросы, не дали никакого результата. На мой взгляд, это гораздо более серьезная проблема, чем возможные ошибки в стандартной теории информации.

Вопрос из зала: Можно провести аналогию между развитием знания и биологической эволюцией. Некоторые знания устаревают, появляются новые знания и так далее. Вопрос состоит в том, кто властен над этим процессом, как им можно управлять и как определить приоритетные направления развития. Или же этот процесс нужно предоставить самому себе.

AP: Я согласен с этой биологической аналогией и думаю, что изучение эпистемических проблем в контексте биологии вообще и эволюционной теории в частности может быть очень интересным и продуктивным. Если опираться на эту аналогию, то можно сказать, что знания, как и жизнь на нашей планете, регулируют сами себя и не нуждаются ни в каком внешнем контроле.

СП: Мой тезис о важности экологии знаний также опирается на эту биологическую аналогию.

AP: Давайте попробуем немного развить эту аналогию, чтобы применить ее к вопросу о контроле и регулировании научных исследований и вообще процесса получения новых знаний. Поскольку мы, люди, являемся биологическим видом наряду с другими биологическими видами, населяющими нашу планету, мы, разумеется, влияем на ход биологической эволюции уже постольку, поскольку мы в ней участвуем: мы боремся за выживание и процветание нашего вида, как это делают и представители любых других биологических видов, включая тех из них, которые в эволюционном отношении находятся очень далеко от человека. Наши знания о мире и основанные на этих знаниях технологии включая, в частности, генную инженерию, ставят нас в особое положение и позволяют влиять на развитие жизни на планете более глубоко и более эффективно, чем это могли делать наши предки. Однако те же самые научные знания о нашей планете, ее биосфере и наших собственных обществах, дают нам и серьезные поводы опасаться, что непродуманное использование этой технологической мощи может легко привести наш вид, а может быть и всю жизнь на планете, к катастрофе. Изучение такого рода обстоятельств и контекстов и является предметом экологии. Понятия регулирования и контроля играют в практических аспектах экологии важную роль, но важно отдавать себе отчет, что в таких случаях речь идет не о полном контроле за ситуацией, а только об осмысленном принятии коллективных решений, последствия которых могут затрагивать состояние биосферы, а значит и коллективных человеческих интересов (формулировка которых также остается предметом дебатов).

Если теперь перейти к вопросу о контроле и регулировании научных исследований, то нужно прежде всего уточнить, что мы рассматриваем в качестве субъекта контроля, а что в качестве объекта контроля. Как я уже сказал, вопрос о приоритетных направлениях исследования решается, во-первых, внутри самого научного сообщества. Это по-разному происходит в разных странах этого сообщества с использованием разных институциональных механизмов. Подобного рода решения могут иметь различную степень обязательности: часто речь идет просто о фактических приоритетах в выборе научных программ. Во-вторых, нужно, конечно, иметь в виду, что современная наука стоит больших денег, источником которых во многих случаях являются государственные и даже межгосударственные (как в случае Европейского Научного Фонда) бюджеты. Это важная, но не единственная причина по которой научное сообщество не является полностью автономным и зависит от поддержки государств и других общественно-политических объединений. Если речь идет о государственном финансировании научных исследований, то научное сообщество, конечно, вынуждено каким-то образом объяснять свои цели всем налогоплательщикам. Впрочем, сказанное относится только к демократическим государствам, тогда как в государствах авторитарного типа для получения необходимого финансирования может быть достаточно убедить одного человека или небольшую группу людей. Как и в случае управления биологическими процессами, при этом нужно иметь в виду, что ни общество в целом, ни отдельный просвещенный диктатор не может в полном смысле слова “быть властным” над процессом получения новых знаний. Главным регулятором развития науки служит сама Природа, которая заставляет ученых отбрасывать некорректные теории, принимать более корректные и развивать с их помощью более эффективные технологии. Хотя вопрос о целях, для достижения которых нам служат наши технологии, относится уже к прагматическим и этическим аспектам техники и ее философии, Природа в таких случаях также играет роль регулятора: если мы, коллективное человечество, сильно промахнется в постановке наших глобальных целей - или даже если нам удастся поставить цели разумно, но мы не сможем их реализовать - то наш биологический вид может вскоре просто уйти с мировой сцены. Чтобы избежать этого трагического сценария или по крайней мере отодвинуть его на неопределенный срок в будущее, нам нужно, на мой взгляд, не пытаться поставить эволюцию жизни на планете и человеческую историю под чей-то безусловный контроль, а научиться ставить практические цели в соответствии с правильно сформулированным общими человеческими интересами и с человеческими возможностями, и достигать этих целей используя все доступные технологические ресурсы. Условием для такого разумного и эффективного отношения к природе и человеку является понимание того, как устроен окружающий нас природный мир без учета специальных человеческих целей. Именно в этом состоит, на мой взгляд, фундаментальное значение естественных наук для человечества.

ЮС. Я благодарю наших блестящих докладчиков и до следующего четверга!

Список литературы:

Luciano Floridi, *The Forth Revolution*, Oxford University Press 2014

Ruth Hagengruber and Uwe V. Riss (eds.), *Philosophy, Computing and Information Science*, Pickering & Chatto 2014

Andrei Rodin, *Axiomatic Method and Category Theory* (Synthese Library vol. 364), Springer 2014.

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow

and Rob Fergus, Intriguing properties of neural networks, /arxiv.org/abs/1312.6199 (last revised 19 Feb 2014)

Knowledge and its Representation in the Computer Era

Participants: S.P. Kovalyov and A.V. Rodin

Project's organizer and moderator: J. V. Sineokaya

Sergei P. Kovalyov, Institute of Control Sciences, Russian Academy of Sciences, 65 Profsoyuznaya st. Moscow, 117997, Russia, email: kovalyov@nm.ru

Andrei V. Rodin, Institute of Philosophy, Russian Academy of Sciences, 12/1 Goncharnaya st., Moscow 109240, Russia, email: andrei@philomatica.or

Julia V. Sineokaya, Institute of Philosophy, Russian Academy of Sciences. 12/1 Goncharnaya Str., Moscow, 109240, Russia; e-mail: sineokaya@iph.ras.ru

Abstract:

Knowledge Representation (KR) is a research area at the borderline between Computer Science and Philosophical Epistemology. Technically KR can be described as a division of Artificial Intelligence focused on prospective information and communication technologies aiming at the efficient use of large data sets. On the other hand, knowledge and its representation is a central theme of philosophical reflexion and analysis; this remark concerns not only the traditional philosophy but also the Analytic philosophy of the 20th century, which essentially uses formal logical methods. On December 8, 2016 the problem of Knowledge Representation has been discussed by epistemologist A.V. Rodin and computer scientist S.P. Kovalyov; this discussion has been organized within the Replicas series making part of joint project of Institute of Philosophy RAS and Dostoevsky Library “Anatomy of Philosophy: working of the text”. The participants discussed new epistemic practices made possible by new computer-based technologies of knowledge representation.

Keywords: knowledge representation, artificial intelligence, neural networks, ontologies, expert knowledge

References:

Luciano Floridi, *The Forth Revolution*, Oxford University Press 2014

Ruth Hagenhuber and Uwe V. Riss (eds.), *Philosophy, Computing and Information Science*, Pickering & Chatto 2014

Andrei Rodin, *Axiomatic Method and Category Theory* (Synthese Library vol. 364), Springer 2014.

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow and Rob Fergus, Intriguing properties of neural networks, [/arxiv.org/abs/1312.6199](https://arxiv.org/abs/1312.6199) (last revised 19 Feb 2014)