

---

---

## МАШИННОЕ ОБУЧЕНИЕ И ОНТОЛОГИИ КАК ДВА ПОДХОДА К ПОСТРОЕНИЮ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

**Андрей Михайлюк, Николай Петренко**

***Аннотация:** В статье описаны основные принципы и особенности двух наиболее популярных подходов к построению интеллектуальных систем: машинного обучения и онтологий. Через раскрытие их особенностей и сравнительный анализ достоинств и недостатков очерчены наиболее вероятные области применения каждого из данных подходов: машинное обучение как наиболее распространенный инструмент «интеллектуализации» крупных систем промышленного уровня и онтологии как подход к созданию преимущественно исследовательских или узко направленных продуктов и подсистем. Рассматривается возможность и предпочтительность совместного применение машинного обучения и онтологий, в поддержку этого приводятся примеры такого применения во всемирно-известных интеллектуальных системах, разрабатываемых глобальными корпорациями. Работа так же раскрывает вероятную причину недостаточного внимания к онтологиям в промышленном и бизнес-ориентированном сообществе по сравнению с машинным обучением.*

***Ключевые слова:** Онтология, Машинное обучение, Интеллектуальные информационные системы.*

***ITHEA Keywords:** 1.2.4 Knowledge Representation Formalisms and Methods.*

---

### **Введение**

---

Согласно множеству прогнозов (например, [Brooks, 2018]), будущее развитие технологий тесно связано с повышением уровня интеллектуальности компьютерных систем. Активное развития отрасли подтверждается появлением государственных программ [USA, 2019],

стремительным ростом количества патентов (рисунок 1), стартапов и инвестиций в данной области (прогноз - более чем 169 млрд. долларов США до 2025 года [Allied Market Research, 2018]), а также огромным количеством научных сообществ, профильных конференций и журналов.

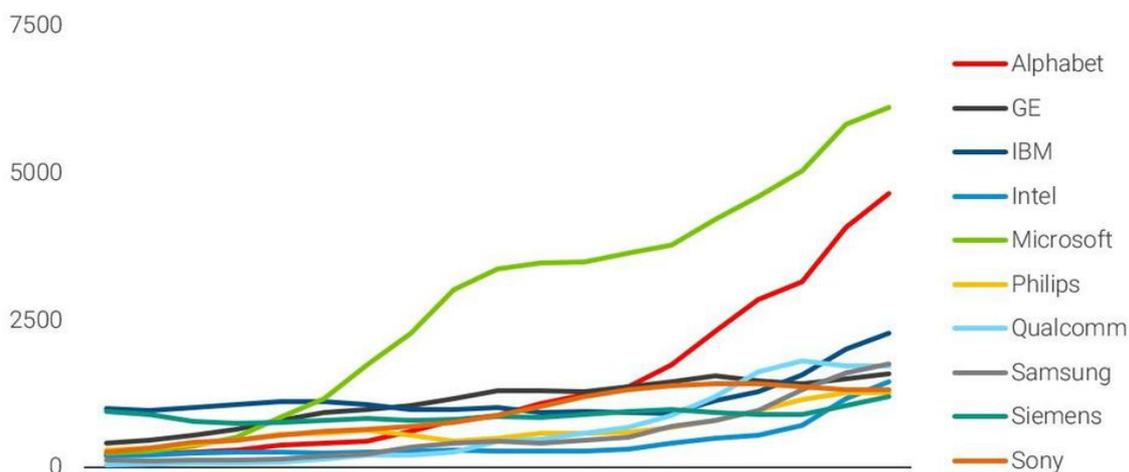


Рисунок 1. Рост количества AI патентов с 2000 по 2018 год [Columbus, 2018]

Интересен тот факт, что подавляющее большинство источников, выделяя основные направления в области искусственного интеллекта (AI), указывают на ряд аппаратно-вычислительных направлений (квантовые вычисления, робототехника, интернет вещей), а также не связанных напрямую с аппаратной составляющей (обработка естественного языка (NLP), распознавание и генерация речи, обработка изображений, машинное обучение (ML)) [Brooks, 2018, Allied Market Research, 2018, Columbus, 2018, Teich, 2018]. В то же время, в научном мире явно выделяется еще одна область исследований и инженерии - онтологии, теоретическими и прикладными аспектами которых занимается множество международных групп - как научных, так и практических [Protégé, Ontology Community, IAQA и др.]. Таким образом, возникает ряд справедливых вопросов:

- 
- Почему промышленное и бизнес-ориентированное сообщество не выделяет онтологические исследования и инженерию как отдельную область или технологию?
  - Как связаны онтологии с другими отраслями AI: ML, NLP и т.д.?
  - Используются ли онтологии в крупных интеллектуальных системах, имеющих множество пользователей, применяемых в различных отраслях и разрабатываемых всемирно-известными компаниями?

В данной статье предпринята попытка ответить на указанные вопросы, описать место онтологий в мире интеллектуальных систем, указать особенности, преимущества и недостатки онтологий как технологии, по сравнению с машинным обучением, привести реальные примеры интеллектуальных систем с явным и неявным использованием онтологий.

---

### **Машинное обучение**

---

В широком смысле, машинное обучение можно описать как любое изменение внутренней структуры, алгоритмов или данных системы, если это изменение позволяет системе лучше выполнять поставленные задачи [Nilsson, 1998]. Данное определение, фактически, описывает понятие обучения без привязки к обучаемой сущности - оно может быть в равной степени применено к определению обучения не только формальной системы (машины), но и, например, человека. Очевидно, что при всей полноте и справедливости такого определения, оно не раскрывает специфику именно машинного обучения, хотя далее автор все-же перечисляет типы моделей машинного обучения, где явно обозначается формальная природа обучаемого:

1. статистические модели;
2. модели, построенные по аналогии со строением мозга (нейронные системы);
3. эволюционирующие модели, основанные на идее изменения системы в процессе ее функционирования (например, за счет реализации эволюционных алгоритмов);

4. искусственный интеллект (системы, целью которых является имитация интеллектуальной деятельности человека, как например, построение модели по описанию, дерево решений, логический вывод в экспертных системах и т.п.);
5. психологические модели, базирующиеся на исследованиях о психологии процесса обучения (семантические графы, проблемная ориентация обучения, поощрение и наказание как механизм обучения, обучение с подкреплением и др.).

В то же время, множество других (как правило, более поздних) источников [Smola, 2008, Shalev-Shwartz and Ben-David, 2014 и др.], рассматривая машинное обучение в целом, подразумевают исключительно обучение на основе статистических методов или нейронных сетей. В [MachineLearning.ru] машинное обучение разделяется на индуктивное и дедуктивное, где экспертные системы и графы знаний составляют множество дедуктивных моделей и методов, но далее термин “машинное обучение” отождествляется исключительно с индуктивными методами: *“Дедуктивное обучение принято относить к области экспертных систем, поэтому термины машинное обучение и обучение по прецедентам можно считать синонимами”*. Таким образом, можно заключить, что машинное обучение, как правило, ассоциируется и отождествляется с индуктивными моделями, основанными на математической статистике (и теории вероятностей) и нейронных сетях (включая глубокое обучение). Далее в статье термин “машинное обучение” понимается в этом смысле.

Отметим, что принято разделять машинное обучение на представленное при помощи математического аппарата (статистика, численные методы) и на искусственные нейронные сети. Однако, т.к. нейронные сети возможно формализовать в математических терминах (полиномы, матрицы и т.д.) [Hassoun, 1995], в дальнейшем описании не будем разделять эти две области.

Общую задачу машинного обучения можно описать как нахождение функции (строго говоря, отображения)  $f: X \rightarrow Y$  по подмножествам заведомо известных пар  $(x \in X_l, y \in Y_l)$ , где  $X$  - множество наблюдаемых

---

событий или объектов,  $Y$  - множество исходов, позволяющих системе лучше выполнять поставленные задачи, множество  $X_l \subset X$  - события или объекты, для которых известны соответствующие исходы  $Y_l \subset Y$ . Пары  $(x, y)$  называют прецедентами, а множества  $X_l$  и  $Y_l$  называют учебной выборкой, так как они являются исходными данными для нахождения целевой функции  $f$  - "учебным материалом" для машины. Однако, в указанной постановке задача машинного обучения не решается на практике, так как имеет ряд особенностей и ограничений:

1. Предположение о корреляции наблюдаемых событий/объектов с исходами. Для того, чтобы найти отображение  $X$  на  $Y$ , элементы данных множеств должны коррелировать, что, очевидно, не всегда справедливо. Более того, известно, что даже формально коррелирующие значения совершенно не гарантируют семантическую корреляцию, т.е. если возможно найти формальную зависимость между множествами величин, описывающих некие события или объекты, это не всегда означает, что данные события или объекты на самом деле зависимы (например, [Spurious correlations]). К сожалению, данная гипотеза в практических задачах не имеет гарантированного разрешения - решение о существовании корреляции принимается исключительно человеком и представляет собой некоторого рода искусство [Воронцов].
2. Гипотеза о решающей функции. В поставленную задачу нахождения функции  $f$ , очевидно, следует добавить оговорку о том, что если корреляция между множествами событий/объектов и исходов существует, то  $f$  должна быть представима аналитически кусочно-заданным отображением с небольшим количеством интервалов - иначе, в вырожденном виде, где каждый интервал отображения  $f$  содержит лишь одно значение аргумента, значения  $f$  на не-учебных данных будет неизвестным. На данный момент не существует эффективных методов автоматического поиска решающего отображения - алгоритм обучения или специалист по данным просто перебирает известные функции, пытаясь найти наиболее приближенную. Т.е., начальная постановка задачи преобразуется в

задачу нахождения не функции  $f$ , а некоего его приближения  $f'$ , которое на учебной выборке давало бы результат, наиболее близкий к наблюдаемым исходам. Отображение  $f'$  также называют гипотезой о функции  $f$ , а метод поиска  $f'$  называют агностическим алгоритмом [Shalev-Shwartz and Ben-David, 2014]. Иногда метод поиска  $f'$  называют методом обучения  $\mu: X_l \times Y_l \times F' \rightarrow f' \in F'$ , где  $F'$  - множество известных функций, среди которых метод и пытается выбрать наиболее близкую к  $f$ . На практике, поиск данной функции является одним из ключевых этапов решения задачи машинного обучения.

3. Требование точности, обобщенности и простоты искомого отображения. Так как машинное обучение предполагает алгоритмическую реализацию применения “обученной” модели на новых данных, найденное отображение  $f'$  должно быть эффективно вычислимо средствами ЭВМ. Кроме того, оно, с одной стороны, должно максимально соответствовать значениям  $f$  на учебной выборке, но при этом должно быть достаточно обобщенным для получения ожидаемых результатов на новых данных. Ситуацию, когда  $f'$  теряет общность, т.е. максимально настроена на получение ожидаемых значений на учебной выборке, но из-за этого не применима на новых данных, называют переобучением. Данные три требования формируют общий критерий предпочтительности выбора одних гипотез  $f'$  над другими и зависят от задачи (например, простота вычисления важнее точности). На практике применяются различные методы для повышения точность и минимизации эффекта переобучения [Smola, 2008, Shalev-Shwartz and Ben-David, 2014, Воронцов].
4. Оптимальность выбранной гипотезы. Для удовлетворения требования точности приближения  $f'$  относительно  $f$  существует ряд подходов и рекомендаций (подробно описанных в [Nilsson, 1998, Smola, 2008, Shalev-Shwartz and Ben-David, 2014, Воронцов]). К сожалению, ни один из них не является универсальным, и возлагается на специалиста по машинному обучению. Отметим

лишь, что точность приближения  $f'$  относительно  $f$  является таким же параметром результирующей модели, как и сама решающая функция  $f'$ . Т.е., задачей является нахождение  $f'$  такой, что  $Q(f', f) \rightarrow \min$  на  $X_l \times Y_l$  и (что немаловажно) всего лишь предположительно на  $X \times Y$ . Иными словами, если признак  $Q$  качества выбора гипотезы оптимален на учебной выборке, это не означает, что он оптимален на неизвестных данных. Отметим также, что так как поиск оптимальной гипотезы  $f'$  также является алгоритмом, а значит и частью реализации модели машинного обучения, вместо оценки самой гипотезы можно вводить оценку алгоритма поиска гипотезы  $Q(\mu, f) \rightarrow \min$  [Воронцов].

5. Параметры событий/объектов и исходов. Формализация исходных данных – событий/объектов и исходов - требует разложения их на некоторые измеряемые и вычисляемые (в широком смысле) параметры. Например, для задачи прогнозирования возврата кредита объектом является заемщик, а интересующим исходом – возврат кредита. Однако, измерение и формализация заемщика как такового невозможны - вместо этого, понятие объекта заменяется на набор его признаков: возраст, средний доход, род деятельности, стаж и т.д. Тогда каждый элемент  $x \in X$  следует заменить на  $x = \{p^x\}$ , где  $p^x \in P = \{p_i^x\}_{i=1}^N$  – множество из  $N$  параметров события/объекта  $x$ . Аналогично следует сделать с элементами множества  $Y$  исходов. Однако, при таком усложнении задачи машинного обучения возникает необходимость определить, какие параметры следует брать во внимание, какие возможно измерить и как привести измерения разнородных параметров к единой шкале. Например, в упомянутой задаче оценки рискованности заемщика, вероятно, цвет волос заемщика не имеет значения (не коррелирует с интересующим исходом); другие (важные) параметры, такие как стаж работы и город проживания, должны быть учтены в модели, но какой из них более важен, насколько важен и как, вообще говоря, можно сравнивать города между собой? Для приведения значений параметров в сравнимый вид модель еще более усложняется

вводом некой функции  $\pi: P \rightarrow P'$ , где любой элемент множества приведенных признаков сравним с любым элементом другого приведенного множества. Подобное усложнение модели добавляет еще две задачи, которые необходимо решить специалисту по данным при построении модели обучения: выбор значимых параметров и нахождение их функций приведения.

Как можно увидеть, модель машинного обучения полностью формализуется при помощи математического аппарата, за исключением предположений (1) о корреляции и (4) оптимальности выбранной гипотезы. С одной стороны, формализация позволяет построить эффективные алгоритмы использования модели, а наличие недоказуемых предположений позволяет применять модель на практически любых данных (и система всегда выдаст какой-то результат), с другой же стороны, подобранные функции  $f'$ ,  $Q$ ,  $\pi$  и др. семантически не связаны с предметной областью (далее - ПдО), поэтому оценка правильности и адекватности данных функций, возможности их применения для решения задачи, объяснение результатов работы системы не всегда возможны, а значит и последующее улучшение построенной модели значительно усложняется.

---

## **Онтологии**

---

Термин "онтология" происходит из древнегреческой философии и в последующем использовался для описания разного рода понятий: разделов наук, структур данных, инструментов информационных систем и др. [Палагін, 2007, Hennig, 2019]. Далее ограничимся рассмотрением формальных онтологий в соответствии со стандартами W3C, т.к. именно они по своим задачам и использованию сравнимы с системами машинного обучения. Формальная онтология (далее просто "онтология") представляет собой формальную структуру, содержащую концепты (понятия, классы, фреймы) и признаки, которые определяют эти концепты (глобально либо относительно других концептов в рамках онтологии). Например, такими признаками могут быть:

- 
- отношения с другими концептами: родо-видовые, количественные, специфичные для ПдО;
  - элементарные признаки (атрибуты), пригодные для прямой интерпретации в рамках некой системы, использующей онтологию (например, время, длина, цвет и т.п.);
  - ограничения (множество или диапазон значений, выражения математической логики и др.);
  - явное перечисление экземпляров, в совокупности представляющих данный концепт.

Иногда частью онтологии считают функции, которые может выполнять онто-система над данным концептом и которые являются полезными в рамках целевых задач. Также, распространенной практикой является включение в онтологию самих экземпляров (объектов, сущностей, элементов класса), в совокупности представляющие соответствующие концепты [Михайлюк, 2014]. Таким образом, онтология является структурированным формальным представлением некой ПдО (универсума в случае онтологий верхнего уровня [Палагін, 2006]), пригодным для автоматического использования программной системой. При этом, ключевыми особенностями онтологий являются:

- Естественное представление. Онтология представляет ПдО через концепты и их признаки, и такое же представление присуще человеку - именно такими категориями как понятия, признаки и экземпляры оперирует человек при изучении и применении полученных знаний. Таким образом, представленная в виде онтологии ПдО естественно понятна человеку.
- Простота интерпретации и валидации результатов, полученных при помощи онтологии, ровно как и самой онтологии. Очевидно, что при естественном и интуитивно понятном представлении ПдО, все задачи, решаемые системой (например, логический вывод или прогнозирование) также могут быть при надобности “вручную” перепроверены экспертом и, если система допустила ошибку,

внести соответствующие изменения либо в систему, либо в онтологическое представление ПдО.

- Эффективное автоматическое использование. Данная особенность является причиной появления именно формальных онтологий и достигается использованием известного математического аппарата для их представления: логика предикатов, графы, теория множеств и алгебра отношений и др. (как правило, несколькими аппаратами совместно).
- Полная эксплицитность представления ПдО. Данная особенность является одновременно достоинством и недостатком онтологии как подхода к построению интеллектуальных систем, т.к. гарантирует простоту интерпретации, но требует максимальной точности при построении онтологии (точность построения онтологии определяет качество и диапазон ее использования в интеллектуальной системе).
- Возможность мета- и само-представления достигается благодаря универсальности способа представления ПдО: онтология может представить произвольную ПдО, включая саму область онтологических систем. Это, в свою очередь, позволяет системе делать логические выводы о самой себе, тем самым эволюционируя.

Перечисленные особенности онтологий как подхода к построению интеллектуальных систем определяют области возможного их применения. Прежде всего, объем знаний, представляемых онтологиями достаточно мал по сравнению с, например, объемом информации, представленной в индексе поисковой системы, например, Google: тысячи понятий [Lists of ontologies] против сотен триллионов статистически проиндексированных веб-страниц [Schwartz, 2016]. Как указывалось выше, причиной относительно небольшого объема онтологий является преимущественно ручной способ из создания или проверки - в противном случае, низкое качество онтологии не позволяло бы производить точные логические выводы или адекватный семантический поиск. Как следствие, на данный момент не существует онтологии, охватывающей все ПдО, -

лишь доменные онтологии и несколько онтологий верхнего уровня, используемые для отображения и интеграции доменных онтологий [Lists of ontologies] (стоит отметить, что задачи отображения и интеграции онтологий до сих пор являются актуальными задачами онтологического инжиниринга [Палагин, 2010]). В рамках ПдО онтологии являются идеальным подходом к объяснению, поиску или логическому выводу знаний, требующих точность и доказательность: поиск возможных исходов судебного дела по нормативным актам, представленным в виде онтологии; постановка диагноза по симптомам и онтологически-представленным знаниям и протоколам в области медицины; перенос данных между двумя гетерогенными системами с онтологией описания форматов и структуры данных в обеих системах и т.п. Еще одной, возможно, наиболее популярной областью применения онтологий являются NLP системы: поскольку естественный язык - это природный способ представления и передачи информации и знаний между людьми, то онтологии, будучи формальным представлением структуры знаний, наиболее подходят для аналогичных задач в формальных системах. Кроме того, структура самого естественно-языкового представления (морфология, синтаксис, семантика) так же идеально формализуются посредством онтологии. Иными словами, онтологии позволяют представить как язык, так и знания, передаваемые языком.

### **Сравнение двух технологий и области их применения**

Описав основные принципы и особенности машинного обучения и онтологий как инструментов и подходов к построению интеллектуальных систем, сравним их по общим критериям, соответствующим основным этапам разработки и использования таких систем:

	<b>Машинное обучение</b>	<b>Онтологии</b>
Выбор или построение модели	Одна из основных задач, длительная и трудоемкая, выполняется итеративно по мере добавления данных	Разработка модели представления онтологии скорее научная задача, на практике обычно просто

	для обучения	выбирается одна из существующих моделей
Добавление начальных данных	Преобразование элементов обучающей выборки в вектора значимых параметров производится в рамках построения модели - само добавление данных полностью автоматическое	Добавление данных в онтологию - сложный процесс, выполняемый экспертом, т. к. требует решения неоднозначностей, противоречий или поиска соответствия параметров новых и уже добавленных в онтологию концептов [Палагин, 2010]
Реализация системы, использующей готовую модель	Вычислительная часть, как правило, крайне проста и сводится к вычислению арифметических выражений; данные независимы, поэтому вычисления легко параллелизируются (и как следствие, система легко масштабируется)	Выражения логики предикатов, операции над множествами и графовые алгоритмы легко реализуются, однако структура данных зачастую связанная, поэтому параллелизация вычислений не всегда возможна
Применение системы	Применение системы, построенной на основе модели машинного обучения, крайне простое: данные вводятся в систему, а модель дает ответ аналогично процессу обучения	Применение системы не представляет сложностей, однако зачастую данные перед применением необходимо строго структурировать (например, обработка естественно-языковых текстов требует предварительного синтаксического разбора)
Объяснение и проверка	Модель машинного обучения представляет	Способ представления знаний в онтологии

качества результатов	семантику решаемой задачи в виде математических функций и числовых коэффициентов, поэтому для человека, как правило, крайне сложно соотнести это с ПдО и, как следствие, объяснить, почему система выдает именно такой ответ, проверить, насколько ответ правильный [Holzinger, 2018]	нагляден и естественен для человека, поэтому любые результаты, полученные при помощи онтологий, легко объяснимы и позволяют “пошагово отследить” процесс получения результата работы системы и оценить правильность как каждого этапа, так и всего результата в целом
Корректировка системы в процессе работы	При выявлении несоответствия модели определенным прецедентам (ситуациям) ПдО, модель необходимо “подстроить”: в случае корректировки количественных параметров модели, “подстройка” выполняется автоматически, однако в случае несоответствия самой модели, ее необходимо полностью перестроить, т.е. по-сути, начать решать задачу сначала	Несоответствие онтологии некоторой ситуации или элементу знаний ПдО решается структурным изменением онтологии (родовидовой иерархии, семантических отношений и т.п.), что не составляет труда для эксперта и, в некоторых случаях, возможно полностью автоматизировать; изменять формальную модель представления онтологии, как правило, не требуется

Приведенные в таблице данные позволяют сделать следующие выводы:

- Для применения машинного обучения к решению некой задачи необходимо прежде всего подобрать подходящую модель и натренировать ее на имеющемся наборе данных. Именно этот этап является наиболее трудоемким и сложным, именно поэтому роль специалиста по машинному обучению наиболее важна. Кроме того,

для построения модели необходимы качественные (полные, отображающие ПдО) данные большого объема. Инвестировав в специалиста и достаточный объем качественных данных, построенная модель будет эффективно применяться в любых масштабах, причем даже при недостатке входящей информации выдача результата, пусть и менее точного, системой гарантирована. Такой подход наиболее свойственен промышленному, коммерческому использованию: значительные начальные инвестиции для простого, широкого, потенциально бесконечного применения с гарантированным получением результата (например, система рекомендации отеля по предпочтениям пользователя даже при недостатке информации о пользователе должна все-равно что-либо предложить – иначе, отсутствие результата воспринимается как неработоспособность системы).

- Решение задач при помощи онтологий требует небольших начальных усилий (выбор формализма и среды работы с онтологией, наполнение ее данными ПдО), которые зачастую требуют наличия лишь небольшого объема энциклопедических знаний и специалиста в ПдО (специалист в области инженерии онтологий не требуется). Построенная система легко применяется в заданной ПдО (хотя ее масштабирование для массового применения - довольно сложная инженерная задача), практически всегда выдает формально правильный и объяснимый результат, а в случае сомнительности результатов - легко проверяется и адаптируется. Очевидно, что такой подход более присущ исследовательским задачам или узким задачам в рамках больших промышленных систем.

---

### **Машинное обучение и онтологии в промышленных системах**

---

Машинное обучение используется повсеместно - сложно привести более или менее исчерпывающий список известных систем с применением машинного обучения - все глобальные поисковые системы, такие как Google Search, Bing, Yandex и др., используют машинное обучение для

---

---

поиска по тексту и изображениям, определению объектов и людей на изображениях, анализа тональности и т.д. [Rowe, 2018]; Google, Tesla и другие используют машинное обучение в системах автономного вождения [Bandom, 2018]. Применение же онтологий, как и описывалось выше, носит гораздо менее масштабный характер, хотя и здесь существует довольно много известных систем:

- IBM Watson - первая программа, победившая в интеллектуальной викторине Jeopardy!, также основана на онтологическом представлении знаний, полученных из энциклопедий и сторонних онтологий, в частности DBPedia, WordNet и Yago [IBM Watson, Sinha, 2016];
- HelathNavigator [HelathNavigator], определяющий возможные состояния здоровья на основе данных симптомов с использованием онтологии болезней и используемый в недавно представленном продукте Microsoft Healthcare Bot [Microsoft Healthcare Bot] - интегрированном сервисе для определения возможных начальных состояний здоровья и ведения процесса лечения посредством текстового чата человека с ботом;
- программы-ассистенты Microsoft Cortana, Apple Siri, S-Voice и Google Now используют онтологии для “понимания” естественного языка, поиска в базе знаний и выбора желаемой реакции (ответ, действие, дополнительные вопросы) [Barbosa, 2017]. Более того, одним из основателей компании, создавшей Siri, согласно [Stuart, 2016], является Том Грубер - автор наиболее известного определения онтологии [Gruber, 1993].

---

## **Выводы**

---

Очевидно, что определить однозначное преимущество какого-либо из подходов невозможно – целесообразность выбора зависит от задачи. А если быть более точным, то оба подхода применимы к решению задач интеллектуальных систем, просто каждый из них применим на разных этапах решения этих задач. Так, все перечисленные выше системы, на самом деле, в той или иной мере успешно используют оба подхода:

например, используя базовые онтологии синтаксиса языка, система обучается распознавать роль слов в предложении; основываясь на онтологии верхнего уровня, получает семантическую структуру текста; семантические признаки составляют исходные данные для ML-системы поиска, которая, обучившись на предыдущих запросах, определяет релевантный ответ, найденный в категории, определенной с помощью онтологий ПдО. Т.е., успешное применение машинного обучения и онтологий представляет собой не выбор одного из подходов, а скорее некий симбиоз этих подходов: онтологии позволяют формализовать данные для машинного обучения, а система машинного обучения позволяет автоматически наполнять или адаптировать онтологию для ее дальнейшего, более точного применения к формализации обучающих данных и т.д.

Такое «скрытое» использование онтологий в крупных интеллектуальных системах не придает достаточной видимости данной области, формируя восприятие роли онтологий как вспомогательных ресурсов для ML-систем. Системы же, основанные на исключительно онтологиях, ввиду описанных выше особенностей, не являются достаточно крупными и широко применяемыми. Вероятно, такая «внутренняя» роль онтологий в известных ML-системах и слабая известность чисто-онтологических систем является причиной низкого внимания к онтологиям в промышленных и бизнес-ориентированных сообществах.

---

### **Дальнейшие направления работы**

---

Очертив области применения ML и онтологий и раскрыв потенциал их совместного применения, следующим шагом может быть разработка обобщенной схемы применения этих двух подходов при построении интеллектуальных систем определенной направленности. Например, при построении поисковой системы, возможно дать рекомендации о том, на каких этапах целесообразно использовать онтологии, что решать при помощи машинного обучения, как связывать результаты этих двух подходов, как оценивать качество таких «гибридных» систем.

---

Другим очевидным развитием вопроса совместного использования ML и онтологий являются исследования в каждой из двух областей. Например, как при помощи ML улучшить качество онтологической интеграции знаний; или же, при построении модели ML, для выбора оптимальной решающей функции, вместо простого перебора множества известных функций, изучить возможность построения и использования онтологий решающих функций.

Еще одним направлением дальнейшей (хотя и не столько научной) работы, можно считать популяризацию онтологий как подхода к построению интеллектуальных систем через более явное представление их роли в крупных промышленных продуктах и посредством разработки чисто-онтологических систем более общего назначения.

---

## Литература

---

[Allied Market Research, 2018] Artificial Intelligence (AI) Market by Technology (Machine Learning, Natural Language Processing, Image Processing, Speech Recognition), and Industry Vertical (Media & Advertising, BFSI, IT & Telecom, Retail, Healthcare, Automotive & Transportation, and Others) - Global Opportunity Analysis and Industry Forecast, 2018-2025. Allied Market Research, 2018. <https://www.alliedmarketresearch.com/artificial-intelligence-market>

[Barbosa, 2017] Barbosa, Simone Diniz Junqueira, Breitman, Karin. Conversations Around Semiotic Engineering. Springer, 2017, 113 pages. ISBN 3319562916, 9783319562919

[Brandom, 2018] Brandom, Russell. Self-driving cars are headed toward an AI roadblock. The Verge, 2018. <https://www.theverge.com/2018/7/3/17530232/self-driving-ai-winter-full-autonomy-waymo-tesla-uber>

[Brooks, 2018] Brooks, Chuck. Four Emerging Technology Areas That Will Help Define Our World In 2019. Forbes, 2018. <https://www.forbes.com/sites/cognitiveworld/2018/12/24/four-emerging-technology-areas-that-will-help-define-our-world-in-2019/#71f1f49358dd>

- [Columbus, 2018] Columbus, Louis. 25 Machine Learning Startups to Watch In 2018. Forbes, 2018. <https://www.forbes.com/sites/louiscolombus/2018/08/26/25-machine-learning-startups-to-watch-in-2018/#15af72cd6f99>
- [Gruber, 1993] Gruber, Tom. A Translation Approach to Portable Ontologies. Knowledge Acquisition, 5(2): 199-220, Stanford University, 1993.
- [Hassoun, 1995] Hassoun, Mohamad H. Fundamentals of Artificial Neural Networks. MIT Press, USA, 1995. 511 pages, ISBN 0-262-08239-X
- [HelathNavigator] HelathNavigator, <https://healthnavigator.com/differential-diagnosis-engine>
- [Hennig, 2019] Hennig, Boris. What is Formal Ontology? 2019. [https://www.researchgate.net/publication/27518522\\_What\\_is\\_Formal\\_Ontology](https://www.researchgate.net/publication/27518522_What_is_Formal_Ontology)
- [Holzinger, 2018] Holzinger A. From Machine Learning to Explainable AI. In 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), IEEE, 2018. pp. 55-66. <https://ieeexplore.ieee.org/abstract/document/8490530>
- [IAOA] IAOA – International Association for Ontology and its Applications <https://iaoa.org>
- [IBM Watson] IBM Watson, [https://en.wikipedia.org/wiki/Watson\\_\(computer\)](https://en.wikipedia.org/wiki/Watson_(computer))
- [Lists of ontologies] Lists of ontologies, W3C. [https://www.w3.org/wiki/Lists\\_of\\_ontologies](https://www.w3.org/wiki/Lists_of_ontologies)
- [MachineLearning.ru] MachineLearning.ru - Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. [http://www.machinelearning.ru/wiki/index.php?title=Машинное\\_обучение](http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение)
- [Microsoft Healthcare Bot] Microsoft Healthcare Bot <https://www.microsoft.com/en-us/research/project/health-bot>
- [Nilsson, 1998] Nilsson, Nils J. Introduction to Machine Learning. Robotics Laboratory, Department of Computer Science, Stanford University, Stanford, 1998
- [Ontology Community] Ontology Community. <http://ontologforum.org/index.php/WikiHomePage>

- 
- [Protégé] Protégé - Stanford Center for Biomedical Informatics Research.  
<https://protege.stanford.edu>
- [Rowe, 2018] Rowe, Kevin. How Search Engines Use Machine Learning: 9 Things We Know for Sure. 2018. <https://www.searchenginejournal.com/how-search-engines-use-machine-learning/224451/#close>
- [Schwartz, 2016] Schwartz, Barry. Google's search knows about over 130 trillion pages. 2016. <https://searchengineland.com/googles-search-indexes-hits-130-trillion-pages-documents-263378>
- [Shalev-Shwartz and Ben-David, 2014] Shalev-Shwartz, Shai, Ben-David, Shai. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, New York, USA, 2014. ISBN 978-1-107-05713-5 (print version). <http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning>
- [Sinha, 2016] Sinha, Tanmay, Boyd, Andrew. IBM Watson Knowledge Studio – Teach Watson about your domain. IBM, 2016. <https://www.ibm.com/blogs/watson/2016/06/alchemy-knowledge-studio>
- [Smola, 2008] Smola, Alex, Vishwanathan, S.V.N. Introduction to Machine Learning. Cambridge University Press, Cambridge, United Kingdom, 2008. ISBN 0-521-82583-0 (print version)
- [Spurious correlations] Spurious correlations, <http://www.tylervigen.com/spurious-correlations>
- [Stuart, 2016] Stuart, David. Practical Ontologies for Information Professionals. Language Arts & Disciplines, 2016, 224 pages. ISBN 1783300620, 9781783300624
- [Teich, 2018] Teich, David A. Machine Learning and Artificial Intelligence in Business: Year in Review, 2018. Forbes, 2018. <https://www.forbes.com/sites/davidteich/2018/12/26/machine-learning-and-artificial-intelligence-in-business-year-in-review-2018/#6cb281b02041>
- [USA, 2019] Accelerating America's Leadership in Artificial Intelligence, 2019. <https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence>
- [Воронцов] Воронцов, К., Математические методы обучения по прецедентам - курс лекций.

[http://www.machinelearning.ru/wiki/index.php?title=Машинное\\_обучение\\_\(курс\\_лекций,\\_К.В.\\_Воронцов\)](http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_(курс_лекций,_К.В._Воронцов))

[Михайлюк, 2014] Михайлюк, Андрей. OWL как стандартная модель представления трансдисциплинарных знаний в Semantic Web. International Journal "Information Content and Processing", Volume 1, Number 3, 2014. <http://www.foibg.com/ijicp/vol01/ijicp01-03-p04.pdf>

[Палагин, 2010] Палагин, Александр, Михайлюк, Андрей, Величко, Виталий, Петренко, Николай. К интеграции онтологий предметных областей. Information Models of Knowledge, Kiev, Ukraine – Sofia, Bulgaria, 2010, с. 69-85. ISBN 978-954-16-0048-1. [http://foibg.com/ibs\\_isc/ibs-19/ibs-19-p08.pdf](http://foibg.com/ibs_isc/ibs-19/ibs-19-p08.pdf)

[Палагин, 2006] Палагин Олександр, Петренко Микола. Модель категоріального рівня мовно-онтологічної картини світу. Математичні машини і системи, №3, 2006, с. 91-104

[Палагин, 2007] Палагин Олександр, Петренко Микола, Михайлюк Андрій. Розвиток та порівняльні характеристики логіко-онтологічних формальних теорій. Математичні машини та системи, №2, 2007. с. 3-18. <http://www.aduis.com.ua/books/6.pdf>

---

#### Информация об авторах

---



**Andrii Mykhailiuk** – Microsoft; Senior Software Engineer. Nad uzlabinou 708/7, Prague 10800, Czech Republic, e-mail: [amykhailiuk@gmail.com](mailto:amykhailiuk@gmail.com)

**Основные области научных исследований:** онтологии, формальные модели представления знаний, системы обработки знаний



**Петренко Николай Григорьевич** – Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: [petrng@ukr.net](mailto:petrng@ukr.net)

**Основные области научных исследований:** методология и инструментальные средства автоматизированного проектирования онтологий предметных областей, системная интеграция трансдисциплинарных научных знаний

## **Machine Learning and Ontologies as Two Approaches for Building Intellectual Informational Systems**

**Andrii Mykhailiuk, Mykola Petrenko**

**Abstract:** *The article describes major principles and features of two the most popular approaches to development of intellectual systems: machine learning and ontologies. Exposition of the approaches' details and comparative analysis lead to definition of the most suitable application domains: machine learning is the most commonly used instrument for "intellectualization" of large production-level systems and ontologies as an approach for creation of mostly research or highly-specific products and sub-systems. Machine learning and ontologies are considered and advised to be leveraged together; to support this suggestion examples of such combined usage are given for well-known intellectual systems developed by global corporations. The paper also reveals probable reason of low attention to ontologies in production and business-oriented circles comparing to the current popularity of machine learning.*

**Keywords:** *Ontology, Machine Learning, Intellectual Informational Systems.*