



ОСНОВНИ ПОНЯТИЯ В ТЕОРИЯТА НА ДЪЛБОКО СТРУКТУРИРАНОТО ОБУЧЕНИЕ

1. Обща теория.

През последните години ускореното развитие на компютърните технологии, по-конкретно нарастващите показатели на капацитета на компютърното апаратно осигуряване, включително за персонално ползване, възможностите за съхраняване, натрупване и достъп до широкомащабни данни и задълбоченото разбиране и познаване на теорията, подпомагат добавянето на нови техники и методи към традиционните области на изкуствения интелект. Изкуственият интелект (Artificial Intelligence - AI) е едно от най-бързо развиващи се научни направления.

След „Третата индустриална революция“, започнала през 60-те години на миналия век и наричана още „Компютърна революция“ или „Цифрова революция“, човечеството е навлязло в „Четвъртата индустриална революция“, наричана още и „Индустрия 4.0“. Реалността на дигиталната ера се характеризира с високи технологични иновации и постижения в области като изкуствен интелект, нанотехнологии, самоуправляващи се автомобили, виртуални асистенти, софтуер за превод, триизмерно принтиране, съхраняване на енергия, квантови изчисления, и много др. Напредъкът на цифровите технологии, глобализацията, високо-технологичната индустрия и технологиите базирани на изкуствен интелект, вече променят живота на много хора от развитите страни и на част от развиващите се. Тези промени са осезателни в редица области като наука, индустрия, бизнес, медицина, фармация, военно дело и редица други. В персоналните „умни“ устройства – телефони, гривни, часовници, телевизори и всякаква домакинска техника свързана с Интернет на нещата (Internet of Things – IoT), са внедрени алгоритми на изкуствения интелект. Редица производители на „умни“ устройства добавят изчислителни блокове за невронни мрежи, като се възползват от хардуерните им възможности за ускоряване на изчисленията при задачи като машинно обучение и компютърно зрение. Свързването на хора, процеси, данни и устройства чрез Интернет на



всичко (Internet of Everything - IoE) надгражда IoT, подобрявайки резултатите на науката, бизнеса и индустрията.

Това е тясно свързано и с концепцията за “всичко като услуга” (Everything as a Service, Anything as a Service (EaaS, XaaS, *aaS)), която се отнася до възможността за многократно използване на софтуерни компоненти в мрежата, с помощта на облачни изчисления. Трите основни „стълба“ от множеството предлагани законни версии на услугите са – „софтуер като услуга“ (Software as a Service - SaaS), „платформа като услуга“ (Platform as a service - PaaS) и „инфраструктура като услуга“ (Infrastructure as a Service - IaaS).

Предоставянето на зловреден софтуер на крайните потребители за целенасочено използване на различни видове кибератаки, т. нар. „зловреден софтуер като услуга“ (Malware as a Service – MaaS) е криминалната версия на този вид услуга. Същият подход последните години се прилага и към ново поле - „Изкуствен интелект като услуга“ (Artificial Intelligence as a Service - AIaaS). С помощта на компютърни системи и вградени алгоритми на AI се решават задачи като изучаване на големи обеми от данни, правене на обобщения и статистически оценки от тях. Прилагането на тези алгоритми в изчислителни, програмируеми устройства доближават системите с изкуствен интелект до човешките действия. Способността на приложните системи с изкуствен интелект да демонстрират интелигентност в една или друга област, е свързано с т. нар. „тесен“ или „слаб“ AI (Artificial Narrow Intelligence - ANI), като част от тези системи са изключително успешни, в това което правят и оказват пряко въздействие върху хората.

В контекста на съвременното разбиране за Изкуствения общ интелект (Artificial General Intelligence – AGI) и първоначалния фокус и ентусиазъм на изследователите на AI „към изграждането на софтуерна програма, която може да решава различни сложни проблеми в различни области и която се контролира автономно, със собствените си мисли, тревоги, чувства, силни, слаби страни и предразположения“, е изместен от мисията за “изграждане на безопасен AGI”. Редица правителства, неправителствени организации и компании създадоха за служителите си „Етичен кодекс за изкуствен интелект“, включващ селекция от водещи принципи и активни позиции в публичния дебат за AI. Други важни въпроси, които са поставени за решаване от международната академична



общност са свързани с „етиката в AI” (AI Ethics), „пристрастията в AI” и „демократизацията на AI” (AI Democratization).

Следваща насока, в която се работи е за създаване на оперативна съвместимост в използването на инструментите за AI и извеждане на разработчиците от затворената рамка или екосистема, в която работят, чрез отворен стандарт за оперативна съвместимост на машинното обучение (Open Neural Network Exchange – ONNX), приобщавайки ги да бъдат част от активно и отворено общество за AI. Според последните прогнози AI ще продължава да се развива с бързи темпове и не предстои следваща „Зима на AI”.

Понятието изкуствен интелект няма точна и универсално приета дефиниция. Едно от определенията за изкуствен интелект е „тази дейност, посветена на превръщането на машините в интелигентни, а интелигентността е онова качество, което позволява на обекта да функционира правилно и предвидимо в заобикалящата го среда“. Друга по ранна дефиниция е дадена от - „Предметът, известен под името изкуствен интелект, изяснява подходите, с помощта на които можем да направим поведението на машините интелигентно“. Други автори са се опитали да проучат и опишат подробно и изчерпателно областта изкуствен интелект, като определят за основна обединяваща тема идеята за интелигентния агент, давайки следната дефиниция за изкуствен интелект - „изследване на агенти, които получават възприятия от околната среда и извършват действия“.

В Декларация за позицията на Институтът на инженерите по електротехника и електроника (Institute of Electrical and Electronics Engineers - IEEE) относно AI, са посочени няколко дефиниции като една от тях е „изкуственият интелект включва изчислителни технологии, които са вдъхновени – но обикновено работят по различен начин – от начина, по който хората и другите биологични организми усещат, учат, разсъждават и предприемат действия“. IEEE – САЩ дефинира AI като „теорията и развитието на компютърни системи, които са в състояние да изпълняват задачи, които обикновено изискват човешка интелигентност, като визуално възприятие, разпознаване на реч, учене, вземане на решения и обработка на естествен език “. В Съобщение на Европейската Комисия до Европейския парламент, Съвета, Европейския икономически и социален комитет и Комитета на регионите —



„Изкуствен интелект за Европа“ използва следната дефиниция за понятието AI и за базираните на AI системи:

- „Наименованието изкуствен интелект се използва за системи, които показват интелигентно поведение, като анализират своята среда и — с известна степен на самостоятелност — предприемат действия за постигане на конкретни цели“.

- „Базираните на изкуствен интелект системи могат да бъдат изцяло софтуерни — действащи във виртуалния свят (например гласови асистенти, софтуер за анализ на изображения, търсачки, системи за разпознаване на глас и лица), а могат и да бъдат внедрени в хардуерни устройства (напр. усъвършенствани роботи, автономни автомобили, дроневи или приложения за „интернет на нещата“).

На национално ниво е изготвена „Концепция за развитието на изкуственият интелект в България до 2030 г., която е разработена от екип на Българската академия на науките (БАН) и външни експерти, и се позовава на 61 документа, включващи български и международни доклади, стратегии, програми, директиви, регламенти и др. Документът е съобразен с документите на Европейската комисия, които разглеждат изкуствения интелект и се основава на „Рамка за Национална стратегия за развитието на изкуствения интелект в България“ и последващата я „Стратегия за развитието на изкуствения интелект в България до 2030 г. (предварителна визия)“.

В исторически план редица ключови научни изследвания и идеи от различни научни направления — философия, математика, икономика, невронауки, психология, компютърни науки, кибернетика, лингвистика, етика и др., предопределят появата на изкуствения интелект като научно-изследователска област още от древността до наши дни.



2. Области на изследване в теорията на изкуствения интелект.

Областите на изследване в теорията на изкуствения интелект включват следните научни области:

Машинно обучение (Machine Learning - ML):

- видове обучение:
 - контролирано обучение, обучение с учител или надзиравано обучение (Supervised Learning - SL);
 - неконтролирано обучение, обучение без учител, ненадзиравано обучение (Unsupervised Learning - UL), самообучение (Self-Learning) и самоконтролирано обучение (Self-Supervised Learning - SSL);
 - учене с подсилване, обучение с утвърждение или подпомагано обучение (Reinforcement Learning - RL).
- хибридни подходи на обучение:
 - смесено обучение, полуконтролирано обучение или комбинирано обучение (Semi-Supervised Learning);
 - многостепенно обучение (Multi-Instance Learning);
- други видове, различни от основната категоризация:
 - дълбоко структурирано обучение или дълбоко обучение (Deep Learning - DL);
 - мета обучение (Meta Learning);
 - автоматизирано машинно обучение (AutoML).

Решаване на проблеми (Problem-Solving):

- решаване на проблеми чрез търсене (Solving Problems by Searching);
- търсене в сложни среди (Search in Complex Environments);
- състезателно търсене и игри (Adversarial Search and Games);
- проблеми за удовлетворение на ограниченията (Constraint Satisfaction Problems - CSPs).

Представяне на знания (Knowledge Representation):

- представяне на знания включващи предикатна логика (Knowledge Representation using Predicate Logic);



- представяне на знания включващи семантични мрежи (Knowledge Representation using Semantic Network);
- представяне на знания с помощта на рамки (Knowledge Representation using Frames);
- представяне на знания използвайки скриптове (Knowledge Representation using Scripts) и др.

Автоматизирани разсъждения (Automated Reasoning):

- логически агенти (Logical Agent) – съждителна логика (Propositional Logic), доказване на съждителна теорема (Propositional Theorem Proving), ефективна проверка на съждителен модел (Effective Propositional Model Checking), агенти базирани на съждителна логика (Agents Based on Propositional Logic);
- логика от първи ред (First-Order Logic);
- извод в логиката от първи ред (Inference in First-Order Logic), право верижно свързване (Forward Chaining), обратно верижно свързване (Backward Chaining) и др.

Несигурни знания и разсъждения (Uncertain knowledge and reasoning):

- количествена несигурност (Quantifying Uncertainty);
- вероятностни разсъждения (Probabilistic Reasoning);
- вероятностни разсъждения във времето (Probabilistic Reasoning over Time);
- вероятно програмиране (Probabilistic Programming);
- вземане на прости решения (Making Simple Decisions);
- вземане на комплексни решения (Making Complex Decisions);
- вземане на решения от многоагентна система (Multiagent Decision Making).

Автоматизирано планиране (Automated Planning):

- класическо планиране (Classical Planning);
- евристика за планиране (Heuristics for Planning);
- йерархично планиране (Hierarchical Planning);
- планиране и действие в непредвидими области (Planning and Acting in Nondeterministic Domains);



- планиране на време, графици и ресурси (Time, Schedules, and Resources);
- анализи на подходи за планиране (Analysis of Planning Approaches).

Машинно общуване (Machine Communicating):

- обработка на естествен език (Natural Language Processing);
- дълбоко обучение за обработка на естествен език (Deep Learning for Natural Language Processing);
- разбиране или интерпретиране на естествен език (Natural Language Understanding – NLU or Interpretation – NLI);
- генериране на естествен език (Natural Language Generation – NLG);
- аудио обработка - разпознаване на реч (Speech Recognition, Voice Recognition);

Машинни възприятия (Machine Perceptions (Perceiving)):

- компютърно зрение (Computer Vision – CV);
- машинен слух (Machine Hearing, Machine Listening);
- машинно докосване (Machine Touch);
- машинно обоняние (Machine Olfaction);

Машинно действие (Machine Acting) – роботика (Robotics):

- видове роботи според външния им вид - антропоморфен (Anthropomorphic robot); манипулаторен (Manipulator); мобилни роботи - автономни автомобили (Autonomous cars), роувъри (Rovers), роботи крака (Legged Robots); дронове квадрокоптери от вид безпилотни летателни апарати (Quadcopter drones a type of Unmanned Aerial Vehicle – UAV); автономни подводни превозни средства (Autonomous underwater vehicles – AUVs,);
- видове роботи според тяхната функционалност – индустриални; сервизни или обслужващи роботи: домашни роботи, роботи за сигурност и наблюдение; за изследвания – лабораторни роботи, работи за проучвания на околна среда или космически апарати; военни роботи – зооморфни; медицински работи – протези, нанороботи, работи извършващи хирургически операции; образователни работи; работи играчки и др.;



- видове работи според сензорите – работи с пасивни сензори като камери; работи с активни сензори като активни далекомери, сонари, структурирани светлинни проектори, сканиращи далекомери-лидари (Light Detection and Ranging -LiDAR), радари, тактилни сензори, сензори за локация, глобални системи за позициониране (Global Positioning Systems - GPS), диференциални GPS, инерциални сензори, проприоцептивни сензори, одометрични, енкодери с вал, сензори за сила, сензори за въртящ момент и др.
- видове работи според задвижващият механизъм – работи с електрически задвижващ механизъм, с хидравличен задвижващ механизъм и с пневматичен;
- работи с различни видове захвати; различни видове работи според задачите, които решават, ниво на планиране и други.

Изкуственият интелигентен агент взаимодейства с околния свят като извлича информация за околната среда чрез машинните възприятия, включващи всички форми на сензорен вход и интерпретира получените данни подобно на начина, по който хората използват сетивата си. С помощта на методите на машинно обучение, системите с изкуствен интелект, правят прогнози извличайки информация от данните. Представянето на знания и автоматизираните разсъждения помагат на системата с изкуствен интелект да се доближи до човешката интелигентност. Последните две концепции се фокусират върху начина на представяне на информацията (фактите) за околния свят във форма, която компютърната система може да използва за решаване на сложни задачи.

Системите с изкуствен интелект търсят решения за изграждане на последователност от действия, които да формират път към постигане на целеви състояния. Алгоритмите за търсене решават различни задачи, извършвайки поредица от действия, които трансформират началното състояние в целево състояние. Те работят в две направления - дефиниране на проблема и търсене на решения в пространството. Интелигентните агенти могат да се определят като динамични и автономни системи, които могат да взаимодействат и с други агенти, включително и хора, като тези взаимодействия изискват комбиниране на

www.eufunds.bg

Проект BG05M2OP001-2.016-0003 „Модернизация на Национален военен университет "Васил Левски"- гр. Велико Търново и Софийски университет "Св. Климент Охридски" - гр. София, в професионално направление 5.3 Компютърна и комуникационна техника“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.



подходи от различни нива на абстракция и различна степен на поведенческа и когнитивна сложност. На фигура 1 се предлага диаграма на областите на изследвания на науката изкуствен интелект, като с обединение на всички подходи се постига т. нар. интелигентен агент или система с изкуствен интелект.



Фиг.1. Диаграма на областите на изследвания на научното направление изкуствен интелект.

Теорията на изкуствения интелект включва различни *методи и техники*, които са форми на статистически и математически модели, предоставящи инструменти за решаване и изчисляване на задачи.

Представени са най-широко разпространените, без да се претендира за изчерпателност в изброяването:

- *евристичен метод на търсене (Heuristic Search in AI)* – търсене в дърво на решенията (Search in the Decision Tree);



- *алгоритми за търсене (Searching Algorithms);*
- *алгоритми за неинформирано търсене (Uninformed Searching Algorithms) – алгоритъм за обхождане в ширина (Breadth-First Search), алгоритъм за обхождане в дълбочина (Depth First Search), алгоритъм на Дейктрас, известен като алгоритъм за търсене на еднакви разходи (Uniform Cost Search), итеративно задълбочаване на първоначалното търсене в дълбочина (Iterative Deepening Depth First Search), алгоритъм за двупосочно търсене (Bidirectional Search);*
- *дървета на решенията (Decision Trees) – подсилени дървета на решенията и дървета на решенията в пакет (Boosted and Bagged Decision Tree);*
- *експертни системи (Expert Systems);*
- *учебни автомати (Learning Automats);*
- *многоагентни системи (Multi-Agent Systems);*
- *изкуствени невронни мрежи (Artificial Neural Networks - ANNs);*
- *еволюционни алгоритми (Evolutionary Algorithms) и генетични алгоритми (Genetic Algorithms);*
- *размити експертни системи (Fuzzy Expert Systems - FES);*
- *обработка на естествен език (Natural Language Process - NLP);*
- *Скрит модел на Марков (Hidden Markov Model - HMM), модел на Гаусовата смес (Gaussian mixture model), к-медоиди (k-medoids), к-средните величини (k-means);*
- *Алгоритъм за най-близкия съсед (k-Nearest Neighbour – kNN), линейни класификатори (Linear classifiers), Наивен Байесов класификатор/алгоритъм (Naïve Bayes Classifier/algorithm);*
- *Мрежи на Байес, статистика на Байес (Bayesian networks, Bayesian statistics);*
- *методи на опорните вектори (Support Vector Machine);*
- *частично и йерархично групиране (Partitional Clustering, Hierarchical Clustering);*
- *условни произволни полета (Conditional Random Fields);*
- *изучаване правилата на асоциации (Association Rule Learning);*



- *учене, базирано на примери (Instance-Based Learning or Memory-Based Learning) и, учене базирано на модели (Model-Based Learning);*
- *регресионен анализ (Regression Analysis) – линейна и логистична регресия (Linear Regression and Logistic Regression), регресия на Гаусов процес (Gaussian Process Regression);*
- *Q-обучение (Q-Learning) с Q-стойност (Q-Value);*
- *процесът на вземане на решение на Марков (Markov Decision process);*
- *разсъждения, базирани на казуси (Case-based reasoning);*
- *дискриминантен анализ (Discriminant Analysis) – линейен или квадратичен (Linear or Quadratic Discriminant) и др.*

Съществуват **два основни модела**, които категоризират системите с изкуствен интелект според способностите и според функционалностите им.

Според **първия модел**, базиран на способности, системата с изкуствен интелект може да бъде разделена на три групи:

- *тесен AI (ANI) - може да възпроизведе или да надхвърли човешките възможности относно конкретна задача. Тези системи използват методите на машинно обучение и дълбоко структурирано обучение;*
- *изкуствен общ интелект (AGI) е това, към което се стремят изследователите по AI- интелигентните агенти да успяват да решават много и различни задачи по модел на човек;*
- *силен или свръх AI (Artificial Strong Intelligence or Artificial Super Intelligence - ASI) е хипотетична система с изкуствен интелект, която далеч ще надминава човешките способности и интелект, ще стане достатъчно самоосъзната, ще може да се самоусъвършенства, да мисли абстрактно по начин, по който хората не могат дори да си представят.*

Според **втория модел**, базирана на функционалностите, системите с изкуствен интелект се делят на:

- *реагиращи машини с AI (Reactive Machine AI) - могат да възприемат настоящата ситуация и да имитират способността на човек да реагира на различни видове стимули, без да използват спомени за натрупан опит. Тези машини нямат функционалност базирана на памет;*



- *машини с ограничена памет с AI (Limited Memory AI)* – могат да съхраняват данни за определен период от време и да анализират тези знания, използвайки ги за вземане на решение. Това са системи обучени на базата на дълбоко структурирано обучение, които съхраняват в паметта си големи количества данни за решаване и подобряване точността на бъдещи задачи;
- *теория на разума в AI (Theory of Mind AI)* - машини, които ще могат да разбират другите с техните чувства, емоции и вярвания, и при взаимодействие ще могат да предизвикат емоции;
- *самоосъзнат AI (Self-Aware AI or Self-Knowledge AI)* – изключително интелигентни машини с AI. Не само ще могат да разбират и да предизвикват чувства и емоции, но ще имат собствени такива. Интелигентен агент с представа за себе си, със собствени нужди и вярвания, с възможност за самонаблюдение.

Както при определяне на обща и изчерпателна дефиниция за това какво е изкуствен интелект, така и в определяне на неговите направления, концепции и подобласти съществува пъстра палитра от възгледи, убеждения и разбирания на различни автори. Много от по-горе посочените допринасящи дисциплини или клонове на AI, могат да използват една или друга техника или няколко. Използването на термина „изкуствен интелект“ често е обединяваща или обобщаваща за всяка форма на технология, която включва интелигентни аспекти. Употребата му като равнозначен на отделна техника или област на AI, напр. машинното обучение или генетични алгоритми, невронни мрежи и други, е неправилна и цели да постигне маркетингово или рекламно въздействие. Постигането на изкуствен общ интелект, при който една машина достига върхова самостоятелна когнитивна способност, като у човека, и предприема самостоятелно рационални действия, все още остава недостижима и далечна за инженерите и изследователите по изкуствен интелект.

От друга страна, напредъка и успехите постигнати в отделни допринасящи клонове и техники в науката за изкуствен интелект са забележителни в наши дни и трябва да им се отдаде нужното внимание.



Бързото развитие на системите с изкуствен интелект и технологичния напредък в различни сектори, обуславя необходимостта от разработването на приложения в много сфери от съвременния живот.

Основните области на приложения на AI са:

- производствен сектор – „умна фабрика“, индустриални работи;
- банков и застрахователен сектор - технически приложения за оптимизация, мониторинг и прогнозиране на финансови пазари;
- приложения в логистика и транспорт – анализ и прогнозиране на предстоящи превози, оптимално организиране на обемите на пратки, предпазване от риск при транспортни компании, намаляване на разходи и въглеродни емисии при превози на товари и стоки;
- автомобилен сектор - автономни автомобили и функции за безопасност;
- военни приложения – автономни оръжия, разпознаване на цели, военна логистика и транспорт, разработване на военни системи за киберсигурност, здравеопазване на военно поле (роботизирани хирургични системи, роботизирани наземни платформи, осигуряващи отдалечена хирургична помощ и дейности по евакуация), бойна симулация и обучение, мониторинг на заплахите и ситуационна осведоменост;
- приложения в авиация и космонавтика – идентификация на пътници и багаж в авиокомпаниите, удовлетвореност на клиенти на авиокомпаниите, безопасност и поддръжка на въздухоплавателно средство, дистанционно наблюдение, откриване и класификация на летателни обекти, обработка на изображения в орбита, следене състоянието на космически кораби и космически обекти, и заобикалящата ги среда;
- приложения в аграрния сектор – „умно земеделие“ и „умни ферми“;
- приложения в сектор образование – персонализирана учебна програма, умно оценяване;
- здравеопазване и фармация – биомедицински изследвания: геномна диагностика, диагностициране на различни синдроми, генетични заболявания, класифициране на злокачествени заболявания, клинична практика, медицински анализ на изображения, дизайн на лекарства;
- маркетингови приложения;

www.eufunds.bg

Проект BG05M2OP001-2.016-0003 „Модернизация на Национален военен университет "Васил Левски"- гр. Велико Търново и Софийски университет "Св. Климент Охридски" - гр. София, в професионално направление 5.3 Компютърна и комуникационна техника“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.



ЕВРОПЕЙСКИ СЪЮЗ
ЕВРОПЕЙСКИ
СОЦИАЛЕН ФОНД



ОПЕРАТИВНА ПРОГРАМА
НАУКА И ОБРАЗОВАНИЕ ЗА
ИНТЕЛИГЕНТЕН РАСТЕЖ

- редица приложения използвани в социалните медии;
- приложения за забавление чрез обработка на изображения - промяна на външен вид и анимиране на образи, намиране на човек от снимка;
- приложения в гейминг индустрията и много други.

www.eufunds.bg

Проект BG05M2OP001-2.016-0003 „Модернизация на Национален военен университет "Васил Левски"- гр. Велико Търново и Софийски университет "Св. Климент Охридски" - гр. София, в професионално направление 5.3 Компютърна и комуникационна техника“, финансиран от Оперативна програма „Наука и образование за интелигентен растеж“, съфинансирана от Европейския съюз чрез Европейските структурни и инвестиционни фондове.