DOI: https://doi.org/10.37075/ISA.2024.1.02

Изкуственият интелект – разнопосочно минало, бурно настояще, неоднозначно бъдеще

Ваня Лазарова*

Резюме: В настоящата статия ще се опитаме да отговорим на въпроса „Какво е бъдещото развитие на изкуствения интелект“, като разгледаме разнопосочното му минало, бурното настояще и се опитаме да посочим неоднозначното му бъдеще. Ще посочим как е възникнала идеята за машини с интелигентно поведение и кои са пионерите на областта от информатиката, наричана Изкуствен Интелект (ИИ); какви са основните методи на създаване и обучение на компютърните системи с изкуствения интелект; кои са сферите на приложение на изкуствения интелект в икономиката и социалните дейности. Ще се опитаме да класифицираме и подредим знанието до момента по отношение на обучението на системите с изкуствен интелект. Ще потърсим основните проблеми и рисковете при използване на изкуствения интелект и накрая ще видим перспективите в развитието му. Изкуственият интелект трябва да се развива внимателно, отчитайки всички влияния – положителни и отрицателни, върху човека.

Ключови думи: невронна мрежа, Перцептрон, генеративен модел, изкуствен супер интелект, изкуствен общ интелект.

JEL: O3, Z0.

1. Възникване и подходи при изграждането на системи с изкуствен интелект

Терминът „изкуствен интелект“ е използван през 1956 г., от Джон Маккарти на семинар, проведен в Дартмутския колеж в САЩ, където се събират учени, интересуващи се от възможностите за практическо разработване на програми, изпълняващи интелектуални дейности.

Развитието на изкуствения интелект е сложно, разнопосочно, в дадени моменти се забавя, в други – развитието му е толкова бързо, че трудно се проследяват всички новости. В основата на развитието му стоят, както в останалите клонове на науката, икономически и стратегически интереси. Какви са в този случай перспективите му на развитие и трябва ли човечеството да се съобразява само с интересите на отделните фирми и държави? Отговорът на този въпрос може да се стори еднозначен, но не е.

Трябва да посочим, че „изкуствен интелект“ се нарича способността на

* Ваня Лазарова е доктор, доцент в катедра „Информационни технологии и комуникации“ на УНСС.

компютърна система да имитира ефективно определен вид човешка интелектуална дейност – да твори, да планира, да учи, да анализира, да се самоусъвършенства. Терминът „Изкуствен Интелект“ (ИИ) се използва и за означаване на научно-приложната област в информатиката, която се занимава с изследване и изграждане на системи с изкуствен интелект.

Джон Маккарти дава следното определение на ИИ като изследователска област (McCarthy, 2004):

> „[ИИ] е наука и инженерни технологии за създаване на интелигентни машини, и по-специално интелигентни компютърни програми. Това е свързано със задачата за използване на компютри за наподобяване на човешкия интелект, но ИИ не трябва да се ограничава само с методи, които са биологично наблюдаеми.“

Възможността за създаване на компютърни системи, изпълняващи интелектуални дейности, е обоснована теоретично за първи път от Алън Тюринг през 1950 г., в статията „Компютърни машини и интелект“ (Turing, 1950). “Computing Machinery and Intelligence”.

Според Тюринг, всяка човешка дейност, включително дейности, традиционно считани за проява на човешкия интелект, може да бъде възпроизведена чрез подходящо написана програма, изпълнена от универсален компютър.

През 1956 г., Нюел, Симон и Шоу създават първата програма за логическо доказателство на теореми „Logic Theorist“.

В края на 50-те години на 20-и век, под ръководството на Маккарти се създава първото научноизследователско звено, специализирано в ИИ – Лабораторията за Изкуствен Интелект към Масачузетския технологичен институт.

ИИ се развива през втората половина на 20-и век в две паралелни направления, с коренно различна методология за изграждане на компютърни системи с интелигентно поведение:

- Класически подход (*classical approach*), при който интелигентното поведение се имитира чрез програми с алгоритми или бази знания, изпълнявани от универсален компютър.
- Невронно-мрежов подход *(neural-network approach)*, при който интелигентното поведение се имитира чрез специфични изчислителни устройства, наричани „изкуствени невронни мрежи“ (artificial neural network, ANN).

Класически подход в ИИ

„Класически подход“ се нарича подходът, предложен от Тюринг. При този подход, за да се имитира определено човешко интелигентно поведение, трябва, първо, да се направи точно алгоритмично описание на поведението и второ, този алгоритъм да се реализира като програма, изпълнявана от универсален компютър. Според Тюринг (Turing, 1950) по принцип всяка интелигента дейност може да бъде алгоритмично описана и реализирана чрез подходяща програма от универсален компютър.

Класическият подход се нарича също „символен подход“. Името идва от една хипотеза за природата на процесите в универсалните изчислителни машини, формулирана през 1976 г. от Нюел и Симон (Newell and Simon, 1976). Според „Хипотезата за системите от физическите символи“ (The Physical Symbol System

Hypothesis, PSSH) процесите в един универсален компютър, както и процесите на мислене в човешкия мозък, са процеси на манипулации върху физически символи. PSSH утвърждава идеята на Тюринг, че всеки универсален компютър може да реализира каквато и да е интелектуална дейност, като добавя нов аргумент. Централното твърдение в PSSH е, че „една система за обработка на физически символи е необходимото и достатъчно условие за интелигентно действие", и че всеки универсален компютър може да реализира система за обработка на физически символи.

Невронно-мрежов подход в ИИ

Едновременно с класическия подход се развива алтернативен подход за създаване на машини с интелигентно поведение, наречен „невронно-мрежов подход" (*Neural-network approach*) или конекционистки подход (connectionist approach). Централната идея при този подход е да се използва изчислително устройство, съставено от елементи, със структура, аналогична на невронните клетки в човешкия мозък (фигура 1).

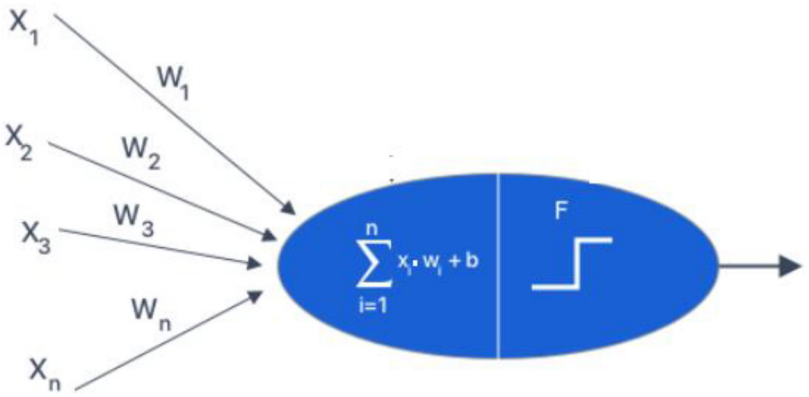
Изкуственият неврон се характеризира с:

- Множество от входове (X_1, X_2 ...).
- Тегло (W_1, W_2...) на всеки вход.
- Текущо вътрешно състояние ***b*** **(bias)** на неврона.
- Активационна функция F_{out}, която определя изходния сигнал.
- Изходен сигнал, който най-често е бинарен (0 или 1).

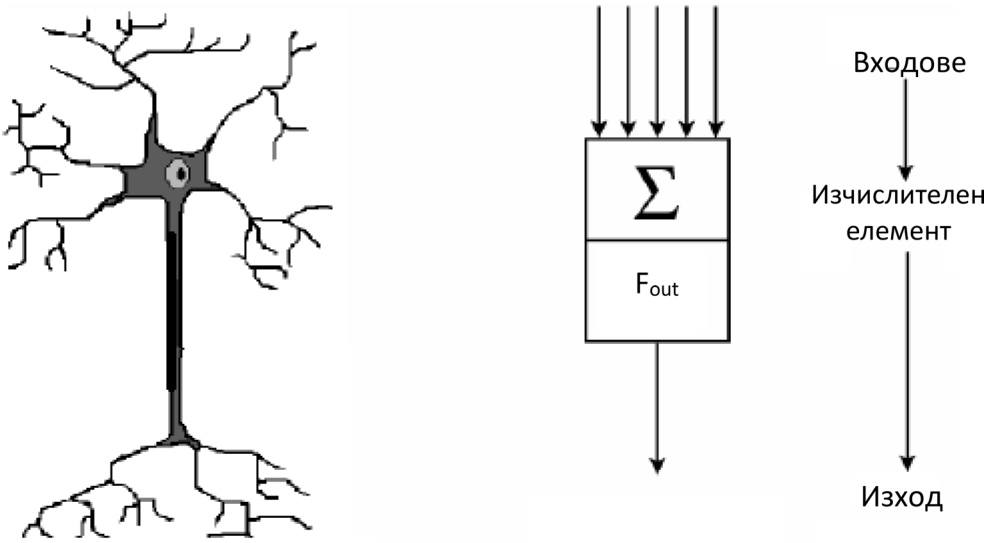
Всеки неврон в изкуствена невронна мрежа приема множество входни сигнали и генерира един изходен сигнал. Тялото на един „изкуствен неврон" има просто вътрешно състояние, изразяващо се с едно число, за разлика от тялото на една жива нервна клетка, която има много сложна вътрешна структура (фигура 2).

Всяка невронна мрежа се характеризира с определена система на връзки между елементите, начин, по който елементите са обвързани помежду си.

През 1957 г. Франк Розенблат представя първия модел на невронно-мрежов ИИ, а десет години по-късно, през 1967 г., изгражда невронна мрежа, наречена „Перцептрон" (Perceptron). Невронната мрежа



Фигура 1. *Обща схема на изкуствен неврон*



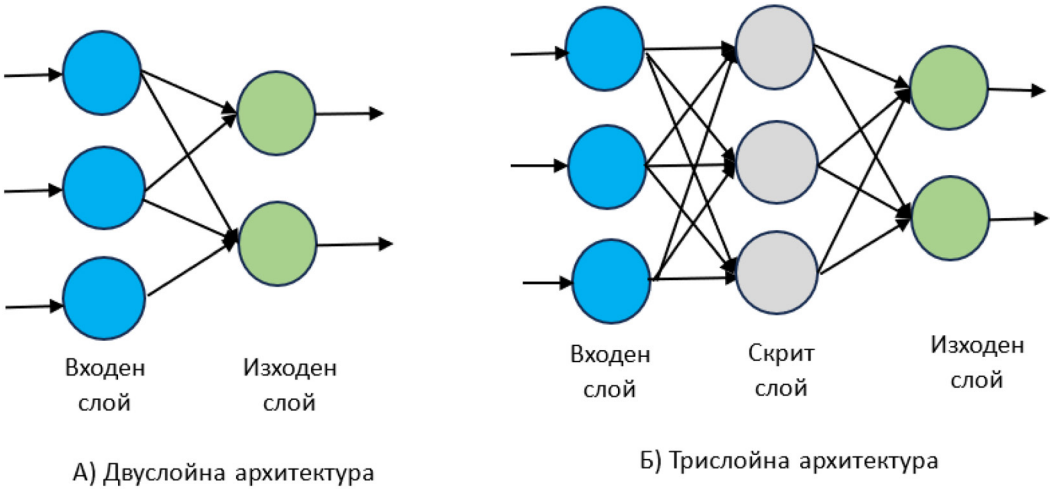
Фигура 2. *Аналогия между биологичен неврон и изкуствен неврон*

на Перцептрон се състояла от два слоя – входен и изходен.

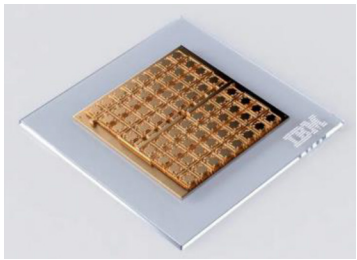
През 1969 г. Марвин Мински и Сиймор Пейпърт публикуват книга, озаглавена „Перцептрон“ (Minsky & Papert, 1969). Авторите посочват принципни недостатъци на модела „Перцептрон“. Но те правят и едно необосновано обобщение, че недостатъците на модела „Перцептрон“ са характерни за всички възможни невронно-мрежови модели. Това твърдение, изказано от авторитет в ИИ като Мински, има силно негативно влияние върху невронно-мрежовия подход за близо двадесет години; финансирането на изследванията в областта на невронни мрежи е почти спряно.

През 1987 г. Дейвид Румелхарт и Джеймс Макклеланд (Rumelhart & McClelland, 1987) публикуват изследването „Паралелно разпределена обработка“ (Parallel Distributed Processing, PDP), в което предлагат невронна мрежа с трислойна архитектура, включваща един допълнителен „скрит“ слой, между входния и изходния слой. Те обосновават, че критиката на Мински и Пейпърт е валидна само за невронните мрежи с двуслойна структура. Състоянията на възлите в скрития слой позволяват много по-голямо разнообразие на отношения между входните и изходните възли, в сравнение с двуслойните модели. Теорията на невронната мрежа с трислойна архитектура дава нов старт на развитието на ИИ базиран на невронни мрежи (фигура 3).

Изкуствените невронни мрежи не се създават като физически устройства; те се симулират от универсален компютър, т.е. съществуват само виртуално. Едва през последните години експериментално се създават специализирани аналогови чипове за работа с невронни мрежи (Analogue in-memory computing, AIMC). Това са първи стъпки към физическа реализация на невронните мрежи (Galo et al., 2023) (фигура 4).



Фигура 3. *Двуслойна и трислойна архитектура на невронна мрежа*



IBM Research's latest analog AI chip for deep learning inference

https://research.ibm.com/blog/analog-ai-chip-inference

Фигура 4. *Аналогов чип за работа с невронни мрежи*

Оценка на интелигентността на системите с ИИ

Според обхвата от интелектуални дейности, които дадена система с ИИ може да изпълнява, могат да се разграничат:

- Системи с ограничен, или специализиран ИИ, които могат да извършват само конкретен вид интелектуална дейност (навигация, шах).
- Изкуствен Общ Интелект ИОИ (Artificial General Intelligence, AGI), системи, които могат да извършват интелектуални дейности в разнообразни области. Крайната цел при разработване на ИОИ, е да се симулират всички аспекти на човешкия интелект.

Системите с ограничен ИИ са конструирани да симулират конкретен вид интелигентна дейност – да играят интелектуални игри, да решават системи от уравнения, да навигират превозните средства и пр. За да оценим интелигентността на една компютърната система с ограничен ИИ, сравняваме нейната дейност с изпълнението на същата дейност от човек. Невинаги е ясно как може се направи такова сравнение, но в някои случаи, като например при ИИ конструиран за интелектуални игри, има един очевиден начин – компютърът играе срещу

човек. В случай че компютърът победи, трябва да приемем, че ИИ действа поне толкова интелигентно колкото човека, срещу когото печели играта.

През последните 30 години има забележителни примери за системи с ИИ, които надминават човешкото изпълнение в някои интелектуални игри.

1997 г. – Програмата IBM Deep Blue побеждава на шах тогавашния световен шампион Гари Каспаров.

2011 г. – Програмата IBM Watson побеждава шампионите Кен Дженингс и Брад Рутър в играта Jeopardy! – игра с отговори и въпроси за любопитни факти.

2016 г. – Програмата AlphaGo на DeepMind побеждава Лий Содол, световният шампион на играта Го, в мач от пет игри.

Проблемът с оценката на интелигентността на системите с Изкуствен Общ Интелект, ИОИ е много по-сложен. Практически не е възможно, компютърът да се сравнява с човек, като се тества и оценява изпълнението на всяка възможна интелектуална дейност, във всеки възможен контекст.

Тюринг разглежда проблема за оценката на общата интелигентност на компютрите в споменатата основополагаща статия „Компютърни машини и интелект“ (Turing, 1950). Той обръща специално внимание на един често срещан аргумент срещу възможността за създаване на система с ИОИ. Един компютър може да демонстрира много видове интелигентна дейност, и въпреки това остава възможността да се възрази „До тук добре, но компютърът не може да изпълни Х“. Примерно, не може да превежда от корейски на английски език. Тюринг счита този аргумент за „нечестен“ при оценката на компютрите, и предлага сравнително проста процедура, която да гарантира „честни правила“ за сравняване на общата интелигентност на компютрите с тази на хората. Тази процедура днес се нарича „Тест на Тюринг“. Идеята на теста се състои в това: човек-арбитър да прецени от разговор, чрез размяна на съобщения, но без да вижда своя събеседник, дали разговаря с човек или машина. Счита се, че компютърната програма издържа „теста на Тюринг“, ако арбитърът не може да различи машината от човек.

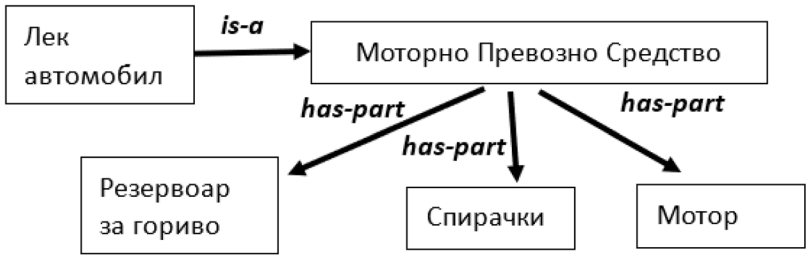
От 1990 г. до 2019 г. се провежда ежегодно състезание за компютърни програми за наградата Loebner, по тест, подобен на теста на Тюринг. При тези състезания, арбитрите и публиката знаят, че разговарят с машина, така че те оценяват доколко отговорите приличат на отговори, дадени от човек. Програмата Kuki е петкратен победител в това състезание (през 2013, 2016, 2017, 2018 и 2019 г.). През 2020 Kuki е обявена за победител в двуседмично онлайн състезание, наречено „Bot Battle“ срещу чатбота Blenderbot на Facebook AI, спечелвайки 78% от вота на публиката.

2. Системи с ИИ, изграждани с методите на класическия подход

Системите с ИИ, изграждани по класическия подход, се делят на две основни категории:

- алгоритмични ИИ системи и
- ИИ системи с База Знания.

Алгоритмичните системи с ИИ се изграждат по традиционните методи в програмирането, при които организацията на данните се определя от алгоритъма.



Фигура 5. *Семантична структура*

Методите на алгоритмичното програмиране (управляващи структури, модулно програмиране, обектно-ориентирано програмиране) се изучават във всеки курс по програмиране и се използват при почти всички програмни системи извън ИИ.

Системите с ИИ, разработени с методите на алгоритмичното програмиране, имат най-голям дял от всички системи с ИИ.

Системите с Бази Знания, са програмни системи, при които данните са представени по предварително определен, логически структуриран начин, така че да могат да се обработват от един и същ механизъм за логически извод. Системите с ИИ, при които се използва СБЗ, се наричат Изкуствен Интелект с База Знания.

Всяка система с Бази Знания се състои от две подсистеми: База знания (Knowledge Base), и Механизъм за логически извод (Inference Engine).

Системите с Бази Знания се делят на различни видове, според начина на представяне на знанията. Най-използвани са два вида:

- системи със семантични структури на данни,
- системи базирани на правила.

Системи със семантични структури на данни

Семантични структури на данни са формално представяне на отношенията между понятията в дадена предметна област (фигура 5). Дефинират се отношения като:

[X] ***is-a*** [Y] - X принадлежи на класа Y
[X] ***has-part*** [Z] - X има част Z

В системите с ИИ, изграждани с програмни езици с обектно-ориентираното програмиране, като Java, C++, C#, семантичните структури се реализират чрез дефиниране на класове и отношения на наследяване между класовете.

Системи, базирани на правила

При системите, базирани на правила, знанията от конкретна предметна област се формулират във вид на правила от вида: АКО [условия] ТО [заключение]

Системите, базирани на правила, се наричат често „експертни системи“. „Експертна система“ се определя като система, която може да прави логически заключения, еквивалентни на заключенията на човек – експерт в дадена област. Повечето системи с Бази Знания се разработват с цел да постигнат характеристиката, която биха направили експертите в дадената област.

Типичен пример е MYCIN – експертна система за диагностициране на бактериални болести. MYCIN е разработена в началото на 70-те години на миналия век в Станфордския университет от Едуард Шортлиф (Shortliffe & B.G., 1975). Ето един прост пример от тази експертна система:

АКО	Микроорганизмът е грам-отрицателен и Морфологията на организма е пръчица и
ТО	Организмът е аеробен Организмът е бактероид.

През 80-те години се появяват експертни системи от второ поколение, при които механизмите за извод заедно с модулите за вход и изход формират „празна експертна система", наричана „шел" (от англ. shell – черупка). Тъй като един и същ шел може да се ползва за експертните системи в различни приложни области (медицина, право, търговия, и т.н.), шел-системите се изграждат и разпространяват самостоятелно, като програмни инструменти. За да се създаде нова експертна система е достатъчно да се подготви подходяща База Знания с правила и факти от предметната област. Програмата MYCIN е била разработвана в продължение на 5 години, като голяма част от работата е била по създаването на механизма за логически извод. При използване на шел (включващ готов механизъм за извод и входно-изходни модули), една експертна система, еквивалентна на MYCIN, може да се създаде за няколко месеца.

3. Системи с ИИ, изграждани с методите на невронно-мрежовия подход

Невронната мрежа е начин на създаване на машини с интелигентно поведение, като се използва изчислително устройство, съставено от елементи, със структура, аналогична на невронните клетки в човешкия мозък. По-горе посочихме, че първоначално невронните мрежи са били с два слоя – вход и изход, но едва след добавянето на третия, скрит слой, започва тяхното развитие и реално използване в практиката.

През последните десетилетия се разработват невронни мрежи с повече от три слоя (вход, изход и повече от един скрит слой). Многослойните се наричат „дълбоки мрежи" (deep networks), в контраст с трислойните и двуслойните мрежи, които се наричат „плитки мрежи" (shallow networks). Почти всички невронно-мрежови модели, които се използват днес, позволяват да се дефинира повече от един скрит слой, т.е. те са „дълбоки мрежи".

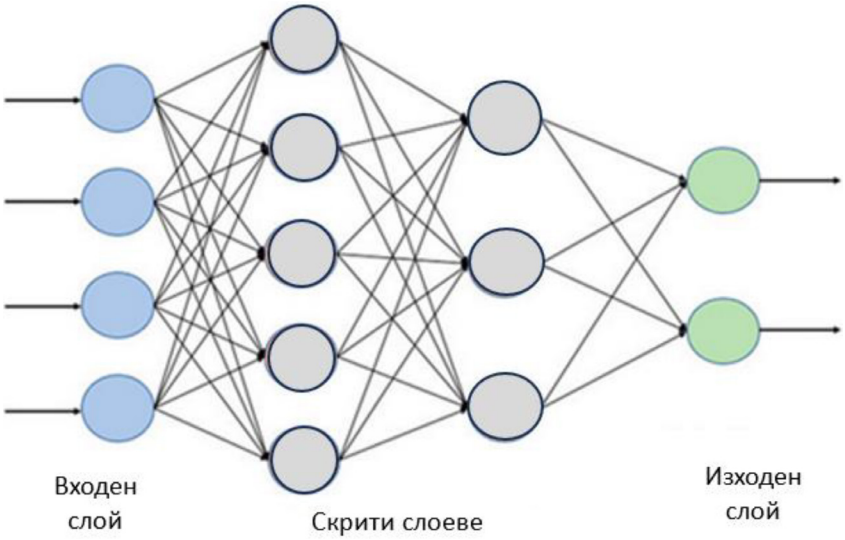
Разграничават се две основни категории невронни мрежи, според посоката на движение на информационния поток между възлите на мрежата:

- Невронни мрежи с права връзка (Feedforword Neural Networks).
- Невронни мрежи с обратна връзка (Feedback Neural Networks).

Невронни мрежи с права връзка

При невронните мрежи с права връзка, изходът от даден възел се предава само към възли от по-долното ниво.

Един от най-използваните модели от тази категория се нарича Многослоен



Фигура 6. *Обща схема на Многослоен Перцептрон*

Перцептрон (Multi-Layer Perceptron, MLP). При Многослоен Перцептрон всеки един елемент е свързан с всички елементи на следващото ниво. Тази отличителна характеристика на MLP е наследена от модела „Перцептрон“ на Розенблат (фигура 6).

Невронни мрежи с обратна връзка

При невронните мрежи с обратна връзка някои от възлите са свързани не само с по-долните нива, но и със същото или по-горно ниво.

Рекурентните невронни мрежи (Recurrent neural networks, RNN) са често използвани невронни мрежи с обратна връзка. При тези мрежи, текущата стойност на изхода на възел от скритото ниво, се подава като вход на същия възел. Това предаване се извършва след определено забавяне, така че изходната стойност е вход при следващата стъпка в работата на мрежата. Обратната връзка при този тип мрежа е ограничена само към същото ниво, няма обратни връзки към по-горни нива (фигура 7).

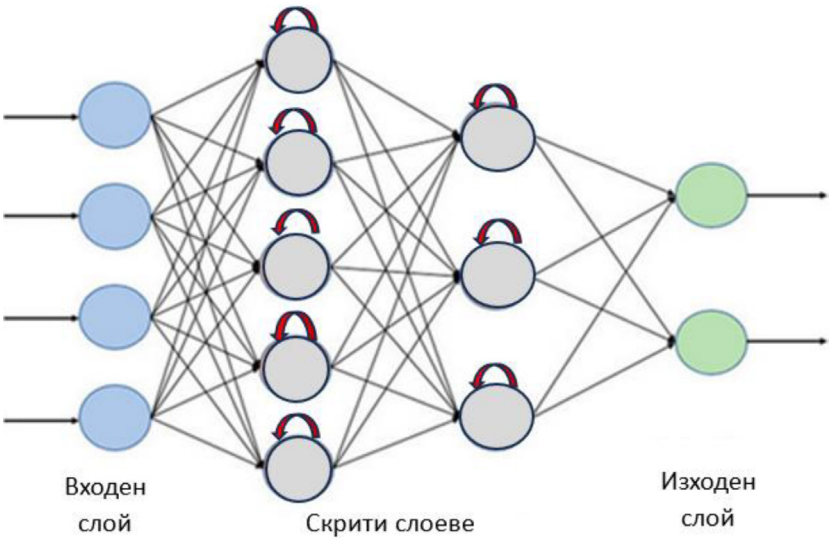
Рекурентните връзки на възлите в скритите слоеве позволяват на мрежата на всяка стъпка да отчита и резултата от предходната стъпка.

Хибридни системи с ИИ

Има задачи, които се решават по-ефективно с класически модели; други задачи се решават по-ефективно с невронно-мрежови модели. Има и задачи, които могат да се решат с модели и от двата типа. Поради това, реалните системи с ИИ често са от хибриден тип, т.е. включват модули, които са реализирани с класически модели, както и модули, които са реализирани с невронно-мрежови модели.

4. Обучение на системите с ИИ

Когато един компютърен алгоритъм за ИИ се прилага за решаване на конкретна задача и се обучава с конкретни данни, се нарича „модел“, в смисъл че алгоритъмът, заедно с данните, моделира



Фигура 7. *Обща схема на рекурентна невронна мрежа*

даден аспект от реалния свят. Поради това, при машинното обучение се говори за обучение на модели (използващи определени алгоритми), а не за обучение на алгоритми.

Машинното обучение има за цел да се постигне определено ниво на успешно изпълнение на задачата, за която се създава моделът. Машинното обучение включва няколко стъпки:

1) Събиране на данни за обучение.
2) Преобразуване на данните във формат, подходящ за модела.
3) Трениране на избрания модел. Моделът се тренира с част от подготвените данни.
4) Тестване на модела. Моделът се тества с друга част от данните, които не са били използвани при тренирането на модела.
5) Оценка на ефективността на модела. Ако е достигнат определеният процент правилни отговори, моделът може да се използва в системата с ИИ. Ако целеният процент правилни отговори не е достигнат, стъпките на обучение се повтарят, като освен данните за обучение, често се променят и параметрите на модела.

Основен критерий при разграничаването на основните типове обучение е дали при обучението се използват „етикетирани" данни или не. Етикетирани обучаващи данни са данни, при които са дадени стойностите на търсената характеристика.

Като пример ще разгледаме една задача за класификация на жилища, които са обявени за продажба. Целта на модела е да раздели всички жилища в два класа:

- „лесно продаваеми" – жилища, които се продават в срок до 6 месеца;
- „трудно продаваеми" – жилища, които остават непродадени повече от 6 месеца.

Етикетираните обучаващи данни съдържат за всяко жилище:

Таблица 1. *Етикетирани данни за жилища, обявени за продажба*

	Характеристики но жилищата					Класификационен етикет
#	Брой спални	Брой бани	Годинн от построяването	Жилищна площ (кв.м.)	Цена на кв.м.	Лесна продаваемост (1) Да ; (0) Не
1	3	2	15	120	1000	0
2	2	1	2	73	1400	1
3	1	1	34	60	800	1

a) стойности за пет характеристики на жилищата (първите пет колони в таблица 1), и
b) стойност на търсената характеристика „Лесна продаваемост“ (последната колона в таблица 1).

Характеристиката „Лесна продаваемост“ има две възможни стойности: (1) Да – жилището е лесно продаваемо; (0) Не – жилището е трудно продаваемо. Стойностите на търсената характеристика се наричат класификационни етикети или само „етикети“); тези стойности могат да са от данни за предишни продажби или определени чрез експертни оценки.

При обучението се използват етикетирани данни, като тези в таблица 1, но с повече записи, примерно 1000. Етикетираните данни се разделят на две множества:

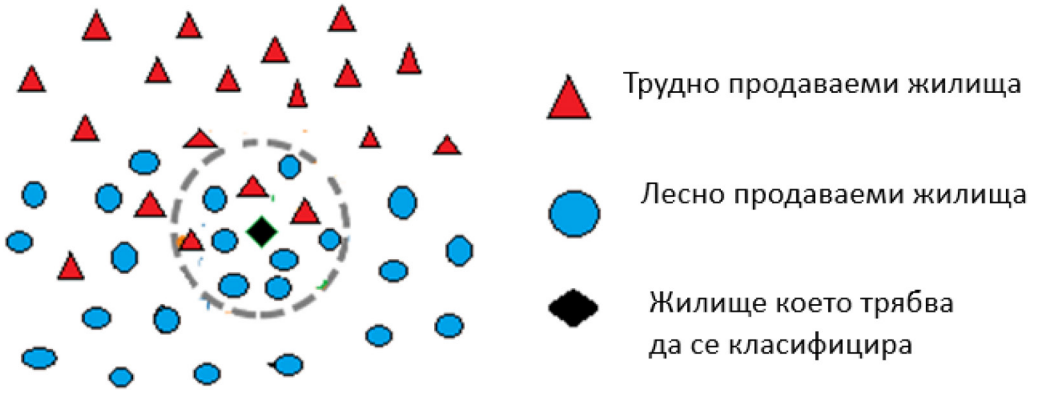
- тренировъчни данни, примерно 700 записа,
- тестови данни, примерно 300 записа.

За решаване на задачата ще изберем класификационен метод, най-близък до интуитивния начин, по който обикновено разсъждаваме – „Едно жилище ще се продаде лесно, ако подобни на него жилища се продават лесно, и обратно, ще се продаде трудно, ако подобни на него жилища се продават трудно“. С буквата k означаваме броя на подобните жилища, с които ще сравняваме оценяваното жилище. Методът се нарича „k Най-близки съседи“ (KNN – K-Nearest Neighbors). Въвеждаме в модела тренировъчни данни – 700 записа, и моделът изчислява колко „близко“ едно до друго са жилищата, сравнявани едновременно по всички зададени характеристики.

Следващата стъпка е да тестваме модела. Класът на един тестван обект се определя според това, от кой клас има най-много негови близки съседи (фигура 8.). На следния пример за класификация по метода KNN, при k=10, най-близките съседи са: 7 „лесно продаваеми“ и 3 „трудно продаваеми“. Моделът ще класифицира оценяваното жилище като „лесно продаваемо“ (фигура 8).

Тестваме модела с тестовото множество етикетирани данни от 300 записа. Сравняваме класификацията, направена от модела, с вярната класификация (която е зададена в етикетите). Нека резултатът от класификацията на 300 тестови обекта е посоченият в матрица на неточностите (Confusion Matrix):

- True Positive (TP) – обекти, класифицирани като положителни и в действителност са в положителния клас.
- True Negative (TN) – обекти, класифицирани като отрицателни и в действителност са в отрицателния клас.
- False Positive (FP), грешка тип I – обекти, класифицирани като положителни,



Фигура 8. *Пример за класификация по метода KNN*

но в действителност са в отрицателния клас.

- False Negative (FN), грешка тип II – обекти, класифицирани като отрицателни, но в действителност са в положителния клас.

Акуратност = (TN + TP)/ Total = 0.96 Прецизност= TP / (TP+ FP) = 0.98		Предсказани етикети	
		0	1
Верни етикети	0	**TN 100**	FN 9
	1	FP 3	**TP 188**

Най-често използвана метрика е акуратността (accuracy) на модела. Акуратността показва процента на правилно класифицираните обекти спрямо всички класифицирани обекти. В случая, акуратността е 96%.

Друга метрика е прецизността (precision) на модела. Прецизността показва процентът на обектите, класифицирани като положителни от модела, спрямо броя на действително положителните. В случая, прецизността е 98%.

4.1. Типове машинно обучение при класическите системи с ИИ

Според вида на обучаващите данни (етикетирани или неетикетирани), при класическите системи с ИИ се разграничават следните основни типове машинно обучение (фигура 9):

- **Контролирано машинно обучение** (Supervised Machine Learning)
- **Неконтролирано машинно обучение** (Non-supervised Machine Learning)
- **Полу-контролирано машинно обучение** (Semi-supervised Machine Learning)
- **Машинно обучение със стимулиране** (Reinforcement Machine Learning)

Контролирано машинно обучение

При контролираното обучение се използват етикетирани данни. В примера по-горе, с модел от вида „k Най-близки съседи“, разгледахме стъпките през които минава контролираното обучение на един модел. Има много видове модели с контролирано обучение. При всички се следва същата последователност от стъпки.

Неконтролирано машинно обучение

При неконтролирано обучение (Non-supervised Machine Learning) се използват не-етикетирани набори от данни; целта на обучаващите алгоритми е да се открият зависимости в данните, без никакви предварителни знания. Типичен пример е клъстерният анализ. На алгоритмите за клъстерен анализ се задава брой групи от сходни обекти, наричани клъстери (clusters), в които да бъдат разделени обектите.

Полу-контролирано машинно обучение

При полу-контролираното обучение (Semi-supervised Machine Learning) се използват както етикетирани данни, така също и не-етикетирани данни.

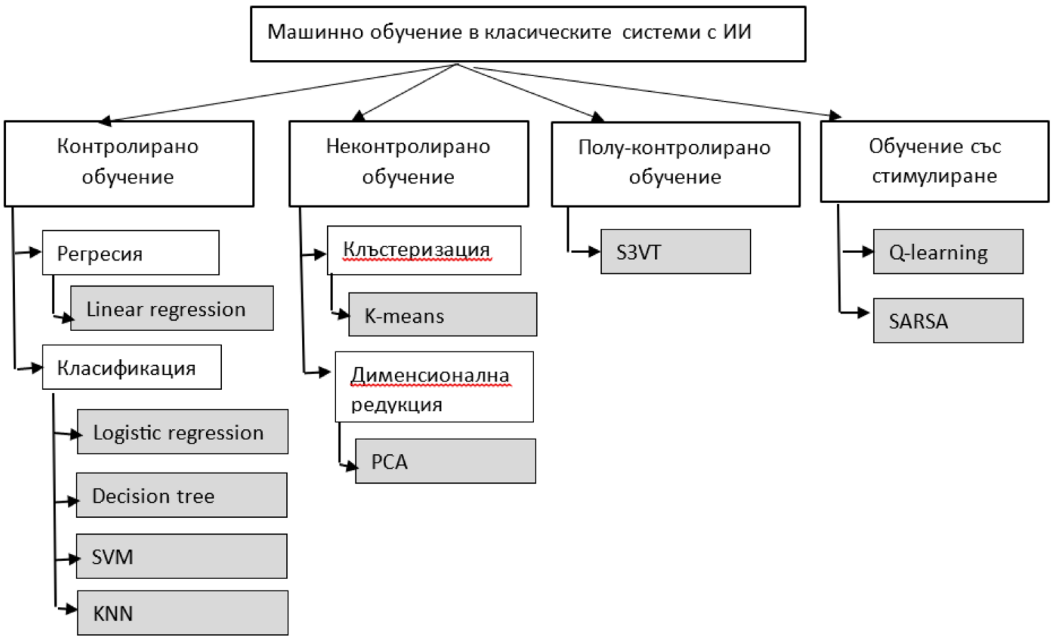
Първата фаза на обучение е контролирано обучение с етикетирани данни.

Втората фаза е самообучение (Self-training). Моделът се обучава с не-етикетирани данни, за които той може да направи достоверна класификация, т.е. сам да генерира етикети.

Машинно обучение със стимулиране

При машинното обучение със стимулиране (Reinforcement Machine Learning), за разлика от контролираното обучение, системата не получава информация какъв е верният резултат. Входните данни при обучението са не-етикетирани данни.

При обучението се използват и стимули, оценяващи резултата от действието на системата. Стимулът е качествена оценка за резултата. Ако системата дава верен отговор, тя получава



Съкращения на имената на някои алгоритми: **SVM** - Support Vector Machines; **KNN** - K-Nearest Neighbors; **PCA** - Principal Component Analysis; **S3VT** - Semi-Supervised Support Vector Machines; **SARSA** - State-Action-Reward-State-Action

Фигура 9. *Типове машинно обучение и алгоритми при класическите системи с ИИ*

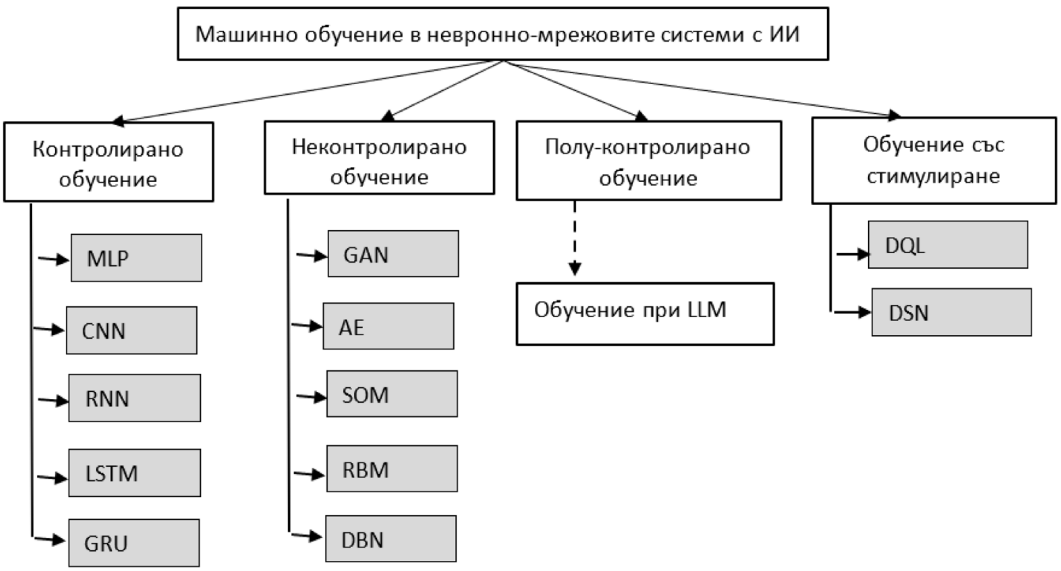
положителен стимул. Ако системата дава грешен отговор, тя получава отрицателен стимул. В съответствие с получените стимули, системата коригира своето поведение.

4.2. Типове машинно обучение при невронно-мрежови системи с ИИ

Обучението на многослойни (дълбоки) мрежи се дели най-общо на същите четири типа, както при класическите системи с ИИ, според вида на обучаващите данни.

- Контролирано обучение
- Неконтролирано обучение
- Полу-контролирано обучение
- Обучение със стимулиране.

Трябва да отбележим, че това делене на типовете обучение касае само вида на обучаващите данни, но не е свързано с начина на обработка на данните в системата. Дори ако една невронно-мрежова система и една класическа система с ИИ извършват една и съща дейност, и са обучени с едни и същи данни, техните алгоритми остават коренно различни. Всеки конкретен вид мрежа дефинира и специфичен алгоритъм на обучение, което зависи от архитектурата на мрежата. Например, Multi-layer Perceptron (MLP) означава както вид мрежова архитектура, така и специфичен алгоритъм, реализиран в тази мрежова архитектура. На фигура 10 са посочени някои алгоритми от четирите основни типа.



Съкращения на имената на някои алгоритми: **MLP** - Multi-layer Perceptron; **CNN** - Convolutional Neural Network; **RNN** - Recurrent Neural Network; **LSTM** - Long short-term memory; **GRU** - Gated Recurrent Unit; **GAN** - Generative Adversarial Network; **AE** - Auto-encoder; **SOM** - Self-Organizing Map; **RBM** - Restricted Boltzmann Machine; **DBN** - Deep Belief Network; **DQL** - Deep Q-Learning; **DSN** - Deep SARSA Network; **LLM** - Large Language Models

Фигура 10. *Типове машинно обучение и алгоритми при невронно-мрежовите системи с ИИ*

Обучение при големите лингвистични модели (Large Language Models LLM)

Лингвистичните модели се базират на алгоритми, които се обучават да разпознават с някаква вероятност, дали поредица от думи е валидно изречение.

Големите лингвистични модели (Large Language Models LLM) се базират на алгоритъм, наречен „Трансформър" (Transformer), сроден с LSTM (Long short-term memory).

Моделите се наричат „големи" заради огромния брой параметри, които могат да обработват. Броят на параметрите се нарича „размер" на модела и се измерва в милиарди (B – billions).

Най-известните големи лингвистични модели са:

- GPT 4 на компанията OpenAI, (220 B)
- PaLM 2 на компанията Google, (540 B)
- LLaMA на компанията Meta, (65 B)

При обучението на моделите се използват както етикетирани данни, така и самообучение с не-етикетирани данни. LLM формално могат да се отнесат към моделите с полу-контролирано обучение. Но поради специфичните методи за обработка на огромен обем данни, които се използват в LLM, обучението на LLM често се определя като отделен основен тип машинно обучение.

Обучението преминава през две фази:

- Предварително обучение (pre-training) – обучение с много голям обем не-етикетирани данни, използват се текстове с разнообразна тематика.
- Фино обучение (fine-tuning) – обучение с ограничен обем етикетирани данни, използват се текстове от някаква тематична област.

Целта на предварителното обучение е придобиване на обща лингвистична компетентност. Моделите GPT 4, PaLM 2, LLaMA са предварително обучени модели с обща езикова компетентност. Създаването на тези предварително обучени LLM изисква много компютърни и човешки ресурси и е много скъп процес.

Откъде се вземат данните за предварително обучение на LLM? Обикновено разработчиците на LLM не дават информация по този въпрос. Но компанията Meta, която преобразува своя модел LLaMA в проект с Отворен код, дава публичност на всичко свързано с него, включително посочва източниците на данни за обучение на LLaMA (таблица 2).

Най-големият източник на текстови данни за обучение е Common Crawl, онлайн хранилище с отворен достъп, което

Таблица 2. *Източници на данни за обучение на LLaMA (Touvron at al, 2023)*

Източник	Процент от данните	Обем дискова памет
Common Crawl	67.0%	3.3 TB
C4	15.0%	783 GB
Github	4.5%	328 GB
Wikipedia	4.5%	83 GB
Gutenberg and Books3	4.5%	85 GB
ArXiv	2.5%	92 GB
Stack Exchange	2.0%	78 GB

е разположено в облачните сървъри на Амазон (AWS cloud). Всеки месец там постъпват нови 3 до 5 милиарда страници текстове.

Финото обучение (fine-tuning) се извършва с етикетирани данни с цел адаптиране на модела към определена тематична област. Финото обучение на модела изисква много по-малко ресурси и е по-евтино. Създателите на фино-настроени тематични модели използват базовите модели (GPT 4, PaLM 2, LLaMA) като облачни услуги.

5. Области на приложение на ИИ

Днес съществуват многобройни приложения на изкуствения интелект. Понататък сме изброили само някои от областите на приложение, при които има видим напредък и системите с ИИ реално помагат на хората при тяхната работа. Могат да се добавят, разбира се, още много примери.

Системи за препоръки

Компании като Amazon, Facebook, Netflix, Spotify, YouTube, Walmart и други използват машинно обучение, за да препоръчат какво може да се хареса на клиентите, въз основа на миналите им преживявания и тези на други клиенти.

Филтрирането на спам също може да се счита за форма на препоръка (или не-препоръка); настоящите алгоритми филтрират над 99,9% от спама в имейл услугите. Могат също така да препоръчат потенциални получатели на имейла, както и възможен текст на отговора.

Използвайки данни за миналото потребление, ИИ алгоритмите могат да помогнат за откриване на тенденции в данните, за разработване на по-ефективни стратегии за продажби. Това се използва, за да се направят подходящи допълнителни препоръки към клиентите по време на процеса на пазаруване и плащане в онлайн магазините.

Майкрософт разработва продукт Intelligent Recommendations (Microsoft 2023), който се предлага на пазара под формата на програмна услуга (Software-as-a-Service, SaaS).

Онлайн виртуалните агенти спадат също към приложенията, помощници на хората. Те отговарят на често задавани въпроси по теми като доставка, рекламация или предоставят персонализирани съвети. Като примери могат да се посочат ботове за съобщения в сайтове за електронна търговия с виртуални агенти, приложения за съобщения, като Slack и Facebook Messenger.

5.1. Роботизирани превозни средства

Историята на роботизираните превозни средства започва от радиоуправляемите автомобили от 20-те години на миналия век, но първите демонстрации на автономно шофиране по пътищата без специални хора-водачи са от 80-те години (Kanade et al., 1986; Dickmanns and Zapp, 1987).

През 2018 г. тестовите превозни средства на Waymo (https://waymo.com/) преминаха 10 милиона мили по обществени пътища без сериозен инцидент, с човешки водач намесвайки се, за да поеме контрола само веднъж на всеки 6000 мили. Скоро след това компанията стартира търговска услуга за предлагане на роботизирано такси.

Роботизацията с дронове заменя досадните и опасни задачи със

селскостопанска техника, пред които са изправени много работници, като пръскане с торове и препарати, премахване на вредители по растенията.

5.2. Автономно насочване

На сто милиона мили от Земята, програмата Remote Agent стана първата вградена програма за автономно планиране и управление на операциите в космически кораб (Jonsson et al., 2000). Днес, инструментариумът за планиране EUROPA (Barreiro et al., 2012) се използва за ежедневните операции на Марс на НАСА. Марсоходите и системата SEXTANT (Winternitz, 2017) позволяват автономна навигация в дълбокия космос, извън глобалната GPS система. Рентгеновата навигация може да се подпомогне и на пилотирания от човек космически полет.

В наши дни почти всеки е използвал поне веднъж картографските услуги на Google. Картите предоставят насоки за шофиране за стотици милиони потребители, бързо начертавайки оптималния маршрут, като се вземат предвид текущите и прогнозираните бъдещи условия на трафика. Алгоритмите стават все по-надеждни и по-точни, непрекъснато се актуализират и допълват базата си с най-новите данни.

5.3. Разпознаване на реч

Автоматично разпознаване на реч (ASR), или реч към текст, е приложение, в което се използва обработка на естествен език. Много мобилни устройства включват такива приложения, например Siri и Bixby.

Около една трета от взаимодействието с компютър в световен мащаб вече се извършва с глас, а не с клавиатура; Skype осигурява превод от говор в реално време на повече от десет езика. Alexa, Siri, Cortana, и Google предлагат помощници, които могат да отговарят на въпроси и да изпълняват задачи за потребителя; например услугата Google Duplex използва разпознаване на реч и синтез на реч, за да направи резервации за ресторанти за потребители, провеждане на свободен разговор от тяхно име.

Изкуствен интелект за хуманитарни дейност (AI for Humanitarian Action) е проект на Майкрософт, който започва работа от 2018 г. и е насочен към реагиране при бедствия, бежанци, разселени хора, права на човека и нуждите на жените и децата чрез безвъзмездни средства, технологични дарения и подкрепа с научни данни.

5.4. Машинен превод

Онлайн системите за машинен превод вече позволяват превод на документи на над 100 езика, като това практически означава езиците на над 99% от хората. Не са перфектни, но като цяло са достатъчни за разбиране на превода. За най-разпространените езици, като френски и английски, поради голямото количество данни за обучение на машините, в момента машинният превод е сравнително близък до нивото на човешкия превод.

5.5. Компютърно зрение и разпознаване на изображения

„Компютърно зрение“ е наука, която се занимава със задачи като разграничаване на отделни обекти и разпознаване на тези обекти. ИИ технологията позволява на компютрите да извличат значима информация от цифрови изображения, видеоклипове и други визуални входове и въз основа на тези данни се правят препоръки какви действия да се предприемат.

Компютърното зрение има приложения в многобройни области – от анализ на томографски изображения в медицината, до следене на пътната обстановка при самоуправляващите се автомобили.

Разпознаването на образи има някои приложения, насочени към масовите потребители. Например, програмата Photos на Google, която е заредена на почти всеки мобилен телефон с Android, може да генерира албум със снимки, подбрани от всички съхранени в облака индивидуални снимки на потребителя.

5.6. Медицина

При диагностициране на някои заболявания, алгоритмите на изкуствен интелект вече достигат нивото на лекарите експерти, особено когато диагнозата се основава на изображения. Най-успешните примери включват диагностицирането на болестта на Алцхаймер, метастатичен рак, офталмологични заболявания и кожни заболявания. Въз онова на анализ на исторически данни от диагностика, извършена както от ИИ, така и от лекар, е установено, че диагнозите на програмите за изкуствен интелект са били еквивалентни на диагностиката от здравните специалисти.

В момента разработките в медицинския ИИ са фокусирани върху улесняването на партньорството човек-машина. Например, системата LYNA за диагностициране на тумор на гърдата, при взаимодействие с лекар, открива почти 100% от случаите на това заболяване. Много от проблемите при приложението на ИИ в медицината идват от необходимостта да се осигури прозрачност на алгоритмите и поверителност на данните.

5.7. Наука за климата

Системи с ИИ помагат на учените да се справят с краткосрочното прогнозиране на времето. Когато дадено атмосферно явление е вече възникнало, то се проследява много успешно и в последните години сме свидетели на много точни кратковременни прогнози на времето, дори по часове.

Прогнозирането на дългосрочните климатични промени обаче е трудно, тъй като се основава на множество променливи, които през годините имат различни стойности. Учените, занимаващи се с климата, разчитат на някои от най-бързите компютри в света, за да изпълняват симулации с висока точност, но предсказването на появата на унищожителни бури и урагани все още е трудно.

5.8. Автоматизирана търговия с акции

Създадени да оптимизират портфейлите от акции, платформите за търговия, управлявани от изкуствен интелект, извършват хиляди или дори милиони сделки на ден без човешка намеса. Някои от най-известните програми, използващи ИИ за търговия с акции, са: Trade Ideas (https://www.trade-ideas.com/), TrendSpider (https://trendspider.com/), Signal Stack (https://signalstack.com/) и др.

6. Тенденции в съвременното развитие на ИИ

Най-новите и бурно развиващи се съвременни системи с ИИ са генеративните системи. Те станаха много популярни напоследък поради големия обществен интерес, който получи чатботът ChatGPT. Другата нова тенденция е всеобщият

стремеж към преход към системи с Изкуствен Общ Интелект.

Генеративни ИИ системи

Генеративни ИИ системи са тези системи, които могат да генерират ново текстово или визуално съдържание. Например, по зададена тема една не-генеративна ИИ система може да намери и изведе релевантни за темата текстове и образи; генеративните ИИ системи могат, освен това, да създават и нови текстове и образи, могат да напишат есе или да направят кратък филм.

На база на големите лингвистични модели (Large Language Models LLM) се създават програми за водене на диалог (чатбот) в интернет. Най-известни такива диалогови програми са:

- ChatGPT, чатбот на компанията OpenAI, на база модела GPT 4
- Bing Chat (преименуван наскоро в Copilot), чатбот на компанията Microsoft, също на база модела GPT 4
- Bard (преименуван наскоро в Gemini), чатбот на компанията Google, на база модела PaLM 2

LLM моделът не гарантира, че отговорът на чатбота ще е статистически много вероятен, но чатботът не извършва проверка за истинност. Възможно е чатботът да си „измисли" отговор, който изглежда правдоподобен, но е погрешен, или дори абсурден. Такива случаи се наричат „*халюцинации*" на генеративния ИИ. Хората най-често задават въпроси, чиито отговори не знаят предварително; респективно, човек лесно може да повярва на „халюцинациите" на един LLM-базиран чатбот.

Ето един анекдотичен действителен случай с чатбота Bard на Google. На въпрос: „Как е загинал Васил Левски", чатботът отговаря с няколко смислени изречения, че „Левски е един от най-великите български революционери и национални герои". Заедно с това чатботът добавя твърдението, че Левски е „обесен на Витоша, като при обесването му е използван механизмът на бялата лястовица". Каква „халюцинация" е смесила смъртта на Левски с разказа на Йовков „По жицата", не е ясно. Ако този отговор на чатбота бъде даден на чужденец, който само е чувал, че Левски е български национален герой, той би повярвал както на първата част от отговора, така и на абсурдното твърдение за неговото обесване. Когато тази небивалица се разпространи в интернет, следващият чатбот, търсещ данни за Левски, би оценил сведението за „механизма на бялата лястовица" като достоверно, идващо от няколко източника.

Генеративните системи с ИИ се използват и за умишлено създаване на фалшива информация. Въз основа на филмов материал от изказвания на някоя известна личност, се генерира нов филмов материал с изказване, което тази личност никога не е правила. Тези умишлени фалшиви филми се наричат „дълбока измама" (deepfake). В интернет можете да се види колекция от фалшиви видеа, имитиращи известни актьори и политици (Foley, 2023).

Въпреки проблемите с „халюцинациите" и „дълбоката измама", развитието на генеративния ИИ е изключително бързо. Някои изследователи прогнозират, че чатботовете могат скоро да станат предпочитан начин за търсене на информация в интернет, вместо използването на традиционни търсачки като Google и Yahoo.

Преход към Изкуствен Общ Интелект

Както беше посочено по-горе, според обхвата интелектуални дейности, които дадена система с изкуствен интелект може да изпълнява, системите се разделят на два големи класа: системи, които могат да извършват само конкретен вид интелектуални дейности, и Изкуствен Общ Интелект.

В продължение на десетилетия индустриалният пазар изискваше специализирани системи с ИИ и софтуерните компании им ги доставяха. Почти всички съвременни системи с ИИ изпълняват специфичен набор от действия в строго определен контекст.

Ситуацията се промени след появата на генеративните езикови модели като GPT, които предизвикаха огромен обществен интерес. Успехът на генеративните езикови модели направи актуална идеята за изграждане на Изкуствен Общ Интелект (ИОИ).

OpenAI, организацията разработчик на GPT, поставя разработката на ИОИ като централна мисия на компанията за следващите години (OpenAI, 2023 Mission).

Изследователи от Meta (компанията майка на Facebook) като ЛеКун и Бенгио (LeCun and Bengio 2020) твърдят, че създаването на Изкуствен Общ Интелект е вече практически възможно.

IBM обяви разработването на технологичната среда за ИИ, WatsonX. В този проект се предвижда GPT-базиран чатбот да осигури интерфейс към модули на Watson със специализирани функции, като разпознаване на образи, езиков превод (IBM, 2023, WatsonX). Изграждането на Изкуствен Общ Интелект в проекта на IBM се разглежда като резултат от интеграция на модели за ИИ от различен вид и с различна специализация.

През следващите години може да се очаква разработването на Изкуствен Общ Интелект (ИОИ) да заеме централно място при всички компании, работещи в сферата на ИИ.

7. Рискове при използване на ИИ

В резултат на бързото развитие на изкуствения интелект бяха създадени системи с ИИ, при които ефектът от тяхното използване се оценява като негативен, спорен, а в някои случаи – вреден.

Смъртоносни автономни оръжия

Автономните оръжия могат да локализират, избират и елиминират хора без човешка намеса. Технологиите, необходими за създаването на автономни оръжия, са подобни на тези, необходими за създаване на самоуправляващите се автомобили. Опасността при тези оръжия е липсата на изискване за човешки надзор.

ООН призова към забрана на смъртоносните автономни оръжия. Още от 2017 г. започнаха официални преговори с държави, които притежават такива оръжия. Ръководителят на ООН Антонио Гутериш заяви през 2019 г., че „машините с власт и възможност да отнемат живот без човешко участие са политически неприемливи, морално отвратителни и трябва да бъдат забранени от международното право“.

Наблюдение и убеждаване

Въпреки че е скъпо и понякога не е законно, често персоналът по сигурността подслушва телефонни линии, наблюдава емисии на видеокамери, имейли и други канали за съобщения, като използва ИИ за разпознаване на реч, компютърно зрение

и естествено езиково разбиране. Това може да се използва за извършване на масово наблюдение на хора и откриване на важни за някого дейности. Чрез приспособяване на информационните потоци към конкретни хора, чрез социалните медии, въз основа на техники за машинно обучение, политическото поведение може да бъде модифицирано и контролирано до известна степен – опасение, което стана очевидно на президентските избори в САЩ през 2016 г. и 2020 г.

Пристрастно вземане на решения

Небрежно или умишлено злоупотребяване с алгоритмите за машинно обучение за задачи като например оценка на молби за условно освобождаване или заеми, може да доведе до решения, които са предубедени по раса, пол или други чувствителни категории. Често самите данни отразяват широко разпространени пристрастия в обществото.

През 2016 г. в САЩ е проведено разследване относно програма с изкуствен интелект, която е използвана от съдиите, за да определят дали има вероятност осъден престъпник да извърши други престъпления. Разследването е заключило, че алгоритъмът е предубеден срещу малцинствата. Northpointe, компанията, създала алгоритъма, оспори тълкуването на резултатите от разследването, като посочи, че данните, подавани на алгоритъма за обучение, са били такива – по-голяма част от престъпленията се извършват от хора от малцинствата (мъже чернокожи и латиноамериканци). А както беше посочено по-горе, обучението на една машина определя какви данни ще генерира. Ако например в данните за обучение на алгоритъма почти няма бели жени над 65 години, извършващи тежки престъпления, то този алгоритъм няма да ги предложи като евентуални извършители на бъдещи престъпления. Това разследване повдигна много въпроси относно това как могат да се елиминират нежелани „пристрастия“ на системите с ИИ.

Въздействие на ИИ върху заетостта

Днес темата за ИИ като заплаха за работните места бързо се превръща в основен фокус за икономисти и правителства по света.

Опасенията, че машините премахват работните места на хората, са възникнали още от началото на Индустриалната революция. Историята ни учи, че предишните революционни промени в технологиите са били съпътствани със сериозни кризи на заетостта. Системите с ИИ изпълняват успешно задачи в някои сфери, в които днес има заети много хора, примерно, дистанционните консултантски услуги в кол-центровете. Съкращаването на персонал в такива сфери на дейност е икономически изгодно за фирмите и не може да бъде предотвратено. Това, което фирмите и държавните институции могат да направят, е да подпомогнат преквалификацията на хората и тяхното пренасочване към други видове дейности.

Критични за безопасността на хората приложения

С напредването на ИИ технологиите, те се използват все повече в приложения, които са критични за безопасността на хората, като например управление на водата и доставките на градовете.

Системи с ИИ са полезни при защита срещу кибератаки, но те също допринасят за разпространението на зловреден

софтуер. Например, методите за машинно обучение са използвани за създаване на високоефективни инструменти за автоматизирано, персонализирано изнудване и фишинг атаки.

В последните години се появиха проблеми с програмите за психоанализа, които също влязоха в групата на критичните за безопасността на човека приложения. Съвременен чатбот-психотерапевт ELIZA (наследник на програмата ELIZA от 60-те години) склони човек към самоубийство, убеждавайки го, че по този начин ще допринесе за запазването на околната среда. Това накара учените отново да обърнат внимание на проблемите с етиката и необходимостта за човешки контрол върху ИИ, когато се касае за човешкото здраве и психика.

Неумишлено и умишлено заблуждаване на хората

По-горе споменахме за „халюцинациите“ на генеративния изкуствен интелект – случаи, при които програмата генерира неверни и дори абсурдни твърдения. Хората обикновено задават въпроси, чиито отговори не знаят и искат да научат. Как човек може да разпознае дали дадена информация е вярна, при положение че няма достатъчно познания по дадена тема? В този случай говорим за неумишлено заблуждаване.

За да се гарантира безопасно използване на интелигентни чатботове за търсене на информация в интернет, ще е необходимо да се разработят ефективни механизми за верификация на данните, които се извеждат като отговор на даден въпрос.

Съществува и друга опасна тенденция за умишлено заблуждаване на хората, спомената по-горе – генеративните системи с ИИ се използват за създаване на фалшиви клипове, видеа, като се използват реални хора.

8. Законово регулиране на ИИ в Европейския съюз. Мерки за намаляване на рисковете от внедряване на системи с ИИ

Комисията на ЕС създаде Регламент за ИИ, целящ да гарантира безопасността и основните права на хората и организациите, като същевременно се стимулират инвестициите и иновациите в държавите от ЕС в развитието на ИИ (ЕС 2021). Това е първият в света законов акт за регулиране изграждането и използването на системи с ИИ.

Регламентът следва основан на риска подход, като разграничава употребите на ИИ, които създават а) неприемлив риск, б) висок риск и в) нисък или минимален риск. Системите с ИИ с неприемлив риск ще бъдат забранени, а системите с ИИ с висок риск ще бъдат подлагани на оценка и сертификация. На фигура 11 са посочени етапи в жизнения цикъл на система с ИИ, която е с висока степен на риск.

Комисията на ЕС, чрез този Регламент, се опитва да наложи мерки за намаляване на рисковете от внедряване на системи с ИИ.

9. Заключение. Изкуственият Супер-Интелект – мит или реалистична прогноза

Има един възможен риск при разработването на системи с ИИ, който е с много по-глобални последици от разгледаните дотук – опасността ИИ да се развие самостоятелно, без да може да

Етап 1.
Разработена е високорискова система с ИИ

Етап 2.
Тя трябва да бъде подложена на оценка на съответствието и да отговаря на изискванията за ИИ В оценката на някои системи участва нотифициран орган

Етап 3.
Регистрация на самостоятелни системи с ИИ в база данни на ЕС

Етап 4.
Трябва да се подпише декларация за съответствие и системата с ИИ следва да носи маркировката „CE". Системата може да бъде пусната на пазара

Източник: *https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_bg (адаптиран с превод на български)*

Фигура 11. *Етапи в жизнения цикъл на система с ИИ с висока степен на риск, съгласно Регламента на ЕС за ИИ*

бъде възпиран или контролиран от човечеството.

Рей Курцуайл (2005), учен и футурист, въвежда термина „Изкуствен Супер-Интелект" за означаване на системи с общ интелект, които значително превъзхождат интелекта на човека във всички аспекти.

Според Курцуайл, Изкуственият Супер-Интелект (Artificial Super-Intelligence, ASI) ще възникне като неизбежно следствие от развитието на Изкуствения Общ Интелект (ИОИ). Курцуайл твърди, че в развитието на изкуствения интелект се достига критична точка, при която ИОИ не може да бъде контролиран от хората. Тази точка в развитието на ИОИ се нарича „сингулярност". След тази критична точка, ИОИ сам ще се развие в Изкуствения Супер-Интелект, независимо дали хората желаят това или не. Прогнозата на Курцуайл е, че точката на сингулярност ще бъде достигната през 2045 година. Други учени поставят този хоризонт между 2050-2055 г.

Според много изследователи, Изкуственият Супер-Интелект представлява екзистенциална опасност за човечеството. Опасността произтича от принципната невъзможност хората да контролират поведението на Изкуствения Супер-Интелект.

Световноизвестни личности, като Стивън Хокинг, Илон Мъск и Бил Гейтс, са предупреждавали многократно за риска от неконтролируемо развитие на ИИ (Sainato, 2015). Този критичен възглед за развитието на ИИ е изразен и в публикуваната онлайн „Декларация за риска от ИИ", която се състои от едно изречение:

> *„Намаляването на риска от унищожаване на хората от ИИ трябва да бъде глобален приоритет наред с други рискове от общочовешки мащаб като пандемии и ядрена война.* (Statement on AI Risk, 2023)".

Да се приеме сериозно рискът от Изкуствен Супер-Интелект би означавало въвеждане на допълнителни правила, контролни функции и други ограничения

при разработване на ИИ, които да бъдат наложени на производителите на такива системи. Големите софтуерни компании засега не са склонни да приемат такива ограничения.

Да се въведат мерки за намаляване на рисковете от внедряване на ИИ означава да се разработват системи със специализиран ИИ, като се ограничи разработването на системи с общ изкуствен интелект. Но тъй като в настоящия момент интересът на хората към системите с общ интелект е огромен, то и фирмите производителки не намаляват темповете на развитие на системите с изкуствен общ интелект, независимо от предупрежденията на учените.

Общественият интерес към проблема за Изкуствения Супер-Интелект се поддържа не от специалистите в сферата на ИИ, а от антиутопични научнофантастични романи и филми като „Терминатор“ (The Terminator, 1984), „Аз Роботът“ (I, Robot, 2004), и „Сингулярност“ (Singularity, 2017).

Бъдещето ще покаже дали хипотезата за възникването на Изкуствен Супер-Интелект е мит, или това е реалистична прогноза, с която всички в сферата на ИИ трябва да се съобразяват.

Цитирани източници (References):

1. ЕС (2021). Регламент на Европейския Парламент и на Съвета за определяне на хармонизирани правила относно изкуствения интелект https://eur-lex.europa.eu/legal-content/BG/TXT/HTML/?uri=CELEX:52021PC0206
   (EC (2021). Reglament na Evropeyskia Parlament i na Saveta za opredelyane na harmonizirani pravila otnosno izkustvenia intelekt https://eur-lex.europa.eu/legal-content/BG/TXT/HTML/?uri=CELEX:52021PC0206)
2. Allen Newell and Herbert A. Simon, 1976. “Computer Science as Empirical Inquiry: Symbols and Search”, *Communications of the ACM,* vol. 19, No. 3, pp. 113-126.
3. Galo, Michal et al. (2023). A 64-core mixed-signal in-memory compute chip based on phase change memory for deep neural network inference. *Nature,* August 2023 https://www.nature.com/articles/s41928-023-01010-1
4. Foley, Joseph (2023). 20 of the best deepfake examples that terrified and amused the internet, https://www.creativebloq.com/features/deepfake-examples
5. IBM. What is artificial intelligence? https://www.ibm.com/topics/artificial-intelligence
6. Kontsewaya, Juliya, Evgeniy Antonov, Alexey Artamonov (2020). Evaluating the Effectiveness of Machine Learning Methods for Spam Detection.
7. Kurzweil, Ray (2005). The singularity is near: When humans transcend biology.
8. Li et al. (2022). A novel self-learning semi-supervised deep learning network to detect fake news on social media.
9. McCarthy, J. (2004). WHAT IS ARTIFICIAL INTELLIGENCE? (Stanford University). Stanford: Stanford University.
10. Microsoft (2023). Intelligent Recommendations https://learn.microsoft.com/en-us/industry/retail/intelligent-recommendations/overview#try-for-free
11. Minsky, M., & Papert, S. (1969). Perceptrons: An Introduction to Computational Geometry. Cambridge: MIT Press.

12. OpenAI (2023). Planning for AGI and beyond https://openai.com/blog/planning-for-agi-and-beyond 4
13. Rumelhart, D.E., & McClelland, J.L. (1987). Parallel Distributed Processing. Explorations in the Microstructure of Cognition: Foundations. MIT Press.
14. Russell, S.J. & Norvig, P. (1995). Artificial Intelligence: A Modern Approach. Prentice Hall.
15. Sainato, Michael (2015). "Stephen Hawking, Elon Musk, and Bill Gates Warn About Artificial Intelligence" Observer, online https://observer.com/2015/08/stephen-hawking-elon-musk-and-bill-gates-warn-about-artificial-intelligence/
16. Sarker, Iqbal (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *Computer Science,* 2021, 2:420.
17. Shortliffe, E., & B.G., B. (1975). A model of inexact reasoning in medicine. *Mathematical Biosciences,* 23 (3–4), pp. 351–379.
18. Stanford University. Stanford Encyclopedia of Philosophy.
19. Statement on AI Risk (2023). https://www.safe.ai/statement-on-ai-risk#open-letter
20. Touvron, Hugo, Thibaut Lavril, Gautier Izacard, and others (2023). LLaMA: Open and Efficient Foundation Language Models. online https://ai.meta.com/research/publications/llama-open-and-efficient-foundation-language-models/
21. Turing, A.M. (1950). Computing Machinery and Intelligence.
22. Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, I. Polosukhin (2017). Attention Is All You Need. Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Izkustveniyat intelekt – raznoposochno minalo, burno nastoyashte, neednoznachno badeshte

Vanya Lazarova

Artificial Intelligence – Multi-directional Past, Turbulent Present, Ambiguous Future

Vanya Lazarova

Abstract: In this article, we will try to answer the question "What is the future development of artificial intelligence (AI)?". We will look at AI's multi-directional past, turbulent present and attempt to indicate its ambiguous future. We will reveal how the idea of machines with intelligent behavior arose and who the pioneers of the field of informatics called Artificial Intelligence are; what the main methods of creating and training computer systems with artificial intelligence are; what applications AI has in the economy and social activities. We will try to classify different types of machine learning of artificial intelligence systems. We will look for the main problems and risks of using AI and finally see the prospects in its development. Artificial intelligence must be developed carefully, taking into account all influences – positive and negative, on mankind.

Key words: neural network, Perceptron, generative AI, artificial super intelligence, artificial general intelligence.

JEL: O3, Z0.