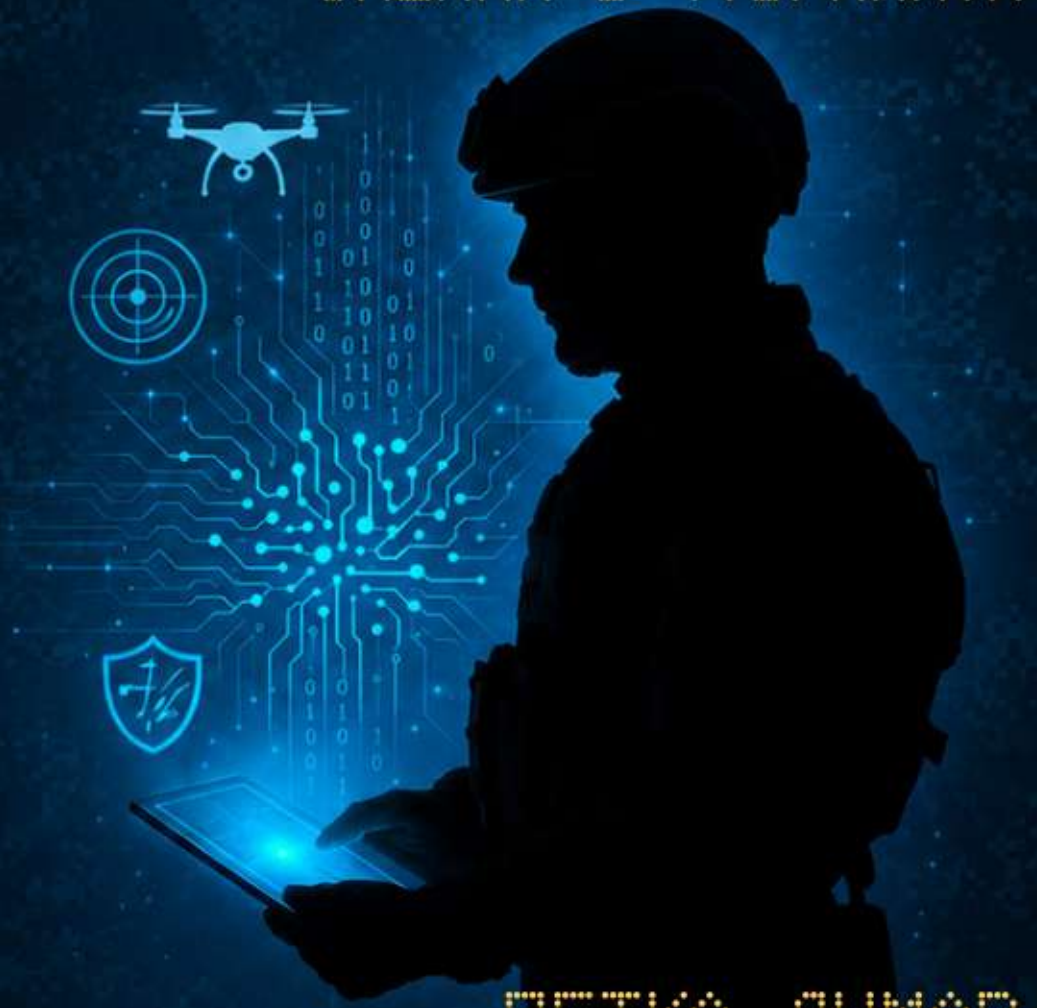
ИЗКУСТВЕН ИНТЕЛЕКТ

И ДИГИТАЛНИ УМЕНИЯ В ОТБРАНАТА



ПЕТКО ДИМОВ

Военна академия „Георги Стойков Раковски“

Петко Димов

ИЗКУСТВЕН ИНТЕЛЕКТ
И ДИГИТАЛНИ УМЕНИЯ В ОТБРАНАТА

София ● 2026

Монографията „Изкуствен интелект и дигитални умения в отбраната“ от полк. доц. д.н. Петко Димов предоставя систематизиран анализ на стратегическото значение на дигиталните компетенции и технологиите с изкуствен интелект (ИИ) в сектора на националната сигурност и отбраната.

Разглежда се историческата еволюция на изкуствения интелект, развитието на дигиталната инфраструктура, невронните мрежи и машинното обучение, като поставя специален акцент върху необходимостта от непрекъснато повишаване на AI грамотността сред военнослужещите. Изследвани са ключовите практически приложения на изкуствения интелект във военното дело, въпросите за киберсигурността и автономните системи, както и конкретни препоръки за изграждане на нормативна рамка и развитие на българския потенциал в областта на изкуствения интелект за целите на отбраната.

ИЗКУСТВЕН ИНТЕЛЕКТ
И ДИГИТАЛНИ УМЕНИЯ В ОТБРАНАТА
Монография
Българска, първо издание, 2026 г.

© Петко Стефанов Димов автор, 2026 г.

© Военна академия „Г. С. Раковски“, издател, 2026 г.

ISBN: 978-619-7711-85-1

СЪДЪРЖАНИЕ

Увод 5
Дигиталната компетентност като стратегически ресурс в ерата на изкуствения интелект 6
1.1. Европейската рамка за дигитални компетентности (DigComp) 7
1.2. От дигитална компетентност към AI интелигентност 13

Еволюция на изкуствения интелект и дигиталната инфраструктура 20
2.1. Философско-математически основи на изкуствения интелект 20
2.2. Основни научни подходи за изучаване на AI 24
2.3. Ранен ентусиазъм и първите AI зими 28
2.4. Експертни интелектни системи 30
2.5.Изграждане на дигиталната инфраструктура (WEB 1.0, 2.0, 3.0) 33
2.6. Ерата на дълбокото обучение и големите езикови модели 34
2.7. AI и съвременната дигитална среда 37

Наука за невронните мрежи и машинното обучение 41
3.1 От символен към статистически изкуствен интелект 41
3.2. Класически методи на машинно обучение 59
3.3. Математически модел на изкуствен неврон и архитектура на перцептрон 72
3.4. Дълбоко обучение и дълбоки невронни мрежи 77

Оптимизационни алгоритми в изкуствения интелект 87
4.1. Алгоритми за групиране без учител (K-means, Йерархична клъстеризация) 91
4.2. Еволюционни алгоритми 96

Езикови модели на трансформаторите в съвременната архитектура 112

Промпт инженеринг и контекст инженеринг 128

Практики за работа с AI инструменти за създаване и управление на дигитално съдържание 141
7.1.Интерфейс и класификация на видовете AI инструменти 141
7.2. Примери от практиката за създаване на съдържание 146

Визуално програмиране (Vibe coding) 152

Оптимизация на сайтове за генеративни системи 159

Автономни AI агенти и интелигентна автоматизация на процеси.....165
10.1. Автономни интелигенти агенти 165
10.2 Протоколи за оперативна съвместимост
и платформи за автоматизация.....169
10.3. Архитектурна интеграция на LLM, RAG и MCP
с различни изпълнителни инструменти 175

Киберсигурност и уязвимости на AI системите.....183
11.1. Кибератаки с помощ от специализирани AI инструменти.....183
11.2. Уязвимости на архитектурата и промпт инжектиране 184
11.3. Оперативни рискове: Shadow AI и човешкият фактор 186

Приложения на AI в отбраната и националната сигурност.....189
12.1.Стратегическо значение на изкуствения интелект в отбраната.....189
12.2. AI в цикъла на вземане на решения и верига на унищожение.....191
12.3. Методи, приложения и основни решения 195
12.4. Емпирично изследване на AI грамотността.....200

**Международен опит и добри практики за употреба
на изкуствен интелект в отбраната**.....211

Препоръки за развитие на AI в отбраната на България.....221
14.1 Нормативна рамка за употреба на изкуствен интелект
в сигурността и отбраната.....221
14.2 Бъдещето на военните длъжности и компетенции
в контекста на изкуствения интелект:.....227
14.3 Българският потенциал в областта на AI.....233

Заключение 243
Приложения 245
Литература 250

УВОД

Предсказването на бъдещето не е магия,
а изкуствен интелект.
Дейв Уотърс

През последния век малко области на науката са получи толкова внимание от обществеността, колкото изучаването на изкуствения интелект. Мислители от Алън Тюринг до Стивън Спилбърг са споделяли своите виждания за това как компютрите някой ден могат да достигнат или надминат човешките възможности. Днес изкуственият интелект е навсякъде, с негова помощ се автоматизират всякакви процеси, усъвършенстват се анализите и управлението на данни, оптимизират се системите за управление, взимане на решения, логистика, разузнаване, комуникация и автоматизация.

Тези технологични пробиви водят до нови възможности, но и до някои предизвикателства в областта на националната сигурност и отбраната. Дигиталните технологии се превръщат във водещ фактор за трансформацията на отбранителните способности и стратегическото планиране. Тяхното усвояване е от изключително значение за запазване на националния суверенитет и повишаване на оперативната ефективност на въоръжените сили.

Високото ниво на дигитална компетентност сред военнослужещите е предпоставка за ефективно прилагане на изкуствения интелект и иновативните инструменти в тяхната дейност. Развитието на аналитични умения, способността за работа с платформи за автоматизация на личния състав и бързото адаптиране към дигиталната среда на военнослужещите са ключови за изграждането на стратегически предимства и конкурентоспособност в глобалната среда на сигурност.

Настоящият труд систематизирано представя основните принципи, добрите практики и актуалните приложения на изкуствения интелект, подходящи за използване в сферата на отбраната. Особено внимание е отделено на въпросите на информационната сигурност, защитата на данни и управлението на дигиталните рискове. Целта е да се формира стабилна основа от знания и практически умения, които да подготвят бъдещите офицери и професионалисти за отговорна и иновативна работа с най-новите технологии в услуга на националната сигурност.

ДИГИТАЛНАТА КОМПЕТЕНТНОСТ КАТО СТРАТЕГИЧЕСКИ РЕСУРС В ЕРАТА НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

Интелигентността не е способността
да трупаш знания, а да знаеш къде да ги откриеш.
Алберт Айнщайн

В епохата на ускорена дигитализация и динамични заплахи за националната и международната сигурност дигиталните компетенции се открояват като стратегически активи в системата за отбрана. За съвременния военнослужещ и експерт по сигурност умението да борави с цифрови технологии не е просто техническо предимство, а критична необходимост за гарантиране на националния суверенитет и оперативната ефективност[1].

Разбирането за дигиталните умения еволюира значително през последните две десетилетия. Още през 2006 г. Европейският съюз признава дигиталната компетентност за една от осемте ключови компетенции за учене през целия живот, наред с езиковата и многоезикова компетентност, STEM компетентностите, личната и гражданска компетентност, културната осведоменост и предприемачеството[2]. Ала-Мутка отбелязва, че развиването на дигиталната компетентност е непрекъснат процес, който преминава през три нива – от инструментални умения, формиращи дигитална грамотност, до най-високото стъпало на критично и отговорно прилагане на технологиите, което води до дигиталната компетентност (Ala-Mutka 2011).

Дигиталната грамотност представлява съвкупност от дигитални умения, осигуряващи ефективно използване на информационните и комуникационните технологии. Дигиталните умения са съставни елементи на дигиталната грамотност, а в комбинация с аналитично и творческо мислене изграждат дигиталната компетентност.

---

1. EUROPEAN COMMISSION. 2023. Digital skills. Shaping Europe's digital future. Available from: https://digital-strategy.ec.europa.eu/en/policies/digital-skills.
2. EUROPEAN PARLIAMENT, 2006. Recommendation on Key Competences for Lifelong Learning. *Official Journal of the European Union*, L 394/10.



Фигура 1. Дигитална компетентност, грамотност и умения

1.1. Европейска рамка за дигитални компетентности (DigComp)

Европейската рамка DigComp дефинира дигиталната компетентност като „набор от знания, умения и нагласи, необходими за уверено, критично и отговорно използване на информационните и комуникационните технологии“ (Vuorikari, Kluzer & Punie 2022). Тази рамка е основният инструмент за измерване на дигиталните умения от 2013 г. насам. През 2022 г. беше приета

DigComp 2.2, която добави над 250 нови примера за знания и умения, включително по темите за дезинформацията и киберсигурността. Поради бурното развитие на изкуствения интелект в края на 2024 г. излезе DigComp 3.0 в синхрон с новото европейско законодателство – Регламента за изкуствения интелект (AI Act)[3], Акта за киберсигурност[4] и Акта за цифровите услуги[5].

Стратегическите политики на ЕС – Digital Compass[6], Digital Decade[7] и инициативата Union of Skills[8] – определят пет основни области на дигитална компетентност, пряко приложими в сектора на отбраната:

1. информираност и грамотност по отношение на данни – търсене и филтриране на информация, управление на данни;

2. комуникация и сътрудничество – споделяне чрез облачни услуги, етикет в мрежата, дистанционна работа;

3. създаване на дигитално съдържание – включително използване на AI аватари и интерактивни графики;

4. безопасност и отговорна употреба – киберсигурност, защита на лични данни, цифрово благополучие;

5. критично мислене и решаване на проблеми – анализ на информация, програмиране, изчислително мислене.

В рамката за дигитална компетентност, показана на фигура 2, се измерват 21 основни компетентности в 8 нива на владеене: два вида базово ниво, средно ниво, ниво за напреднали и силно специализирано ниво (Vuorikari, Kluzer & Punie 2022). Според обновената рамка DigComp 3.0 ефективната интеграция на изкуствения интелект изисква нов тип компетенции, които надхвърлят фундаменталната дигитална грамотност (Cosgrove, Cachia 2025). Тези компетенции се оценяват по четири ключови компонента:

[3.] EUROPEAN PARLIAMENT, 2024. Regulation (EU) 2024/1689 Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*.

[4.] EUROPEAN PARLIAMENT AND THE COUNCIL, 2024. Regulation (EU) 2024/2847 On horizontal cybersecurity requirements for products with digital elements (Cyber Resilience Act). *Official Journal of the European Union*.

[5.] EUROPEAN PARLIAMENT AND THE COUNCIL, 2022. Regulation (EU) 2022/2065 On a Single Market for Digital Services (Digital Services Act). *Official Journal of the European Union*.

[6.] EUROPEAN COMMISSION, 2021. 2030 Digital Compass: the European way for the Digital Decade. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*.

[7.] EUROPEAN COMMISSION, 2022. The Digital Decade Policy Programme 2030. *Decision of the European Parliament and of the Council establishing the policy programme 2030*.

[8] EUROPEAN COMMISSION, 2025. Union of Skills. Commission initiative/communication on skills, upskilling and reskilling.

– познаване на принципите и разбиране как функционира AI;

– практическо прилагане и умение за ефективно използване на AI инструменти;

– оценяване и критичен анализ на генерираните от AI решения;

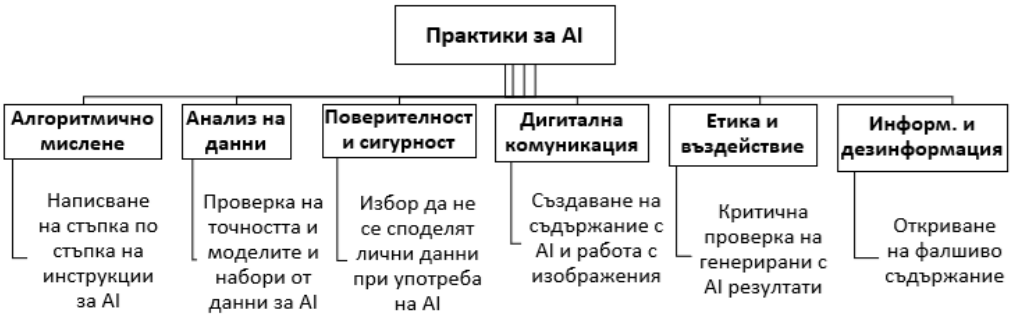
– етика и отговорна употреба, разбиране на социалното въздействие и избягване на пристрастия в алгоритмите.



Фигура 2. Области на дигитална компетентност според DigComp 3.0

За поддържане на високо ниво на дигитална компетентност сред военнослужещите е необходимо постоянно обучение и развитие на личния състав. Това включва създаване на нови учебни програми и практически тренировки по алгоритмично мислене, умения за създаване на правилни инструкции за AI, анализ на данни, работа с лични данни, създаване на

съдържание с AI инструменти, разпознаване на дезинформация и етично използване на изкуствения интелект.[9]



Фигура 3. Практики за натрупване на умения при работа с AI

В рамките на настоящото изследване е създаден инструмент за бърза самооценка на дигиталните умения – тест, представен в Приложение 1, който се базира на адаптиран и опростен вариант на индикаторите, формулирани в европейската рамка DigComp. Този инструмент осигурява структурирана и обективна основа за идентифициране на индивидуалните силни страни и зони за усъвършенстване, като същевременно създава предпоставки за системна работа по надграждане на дигиталната и AI грамотност в сигурността и отбраната.

С развитието на технологиите фундаменталната грамотност постепенно се надгражда, като преминава през етапите на компютърно мислене, дигитално гражданство, медийна и информационна грамотност, за да се превърне в съвременна AI грамотност. През етапа на компютърното мислене хората разбират, че вече не им е достатъчно просто умение за четене, а се нуждаят от разбиране на алгоритми и автоматизиране на процеси. След това социалното общуване от физическия свят се превръща в дигитално гражданство. Способността за разбиране на текст се разширява до критичен анализ на съдържанието и медийна грамотност, която позволява различаване на дезинформацията. В сферата на отбраната значението на посочените умения нараства многократно, защото внедряването на изкуствен интелект позволява автоматизация на административни процеси, оптимизация на

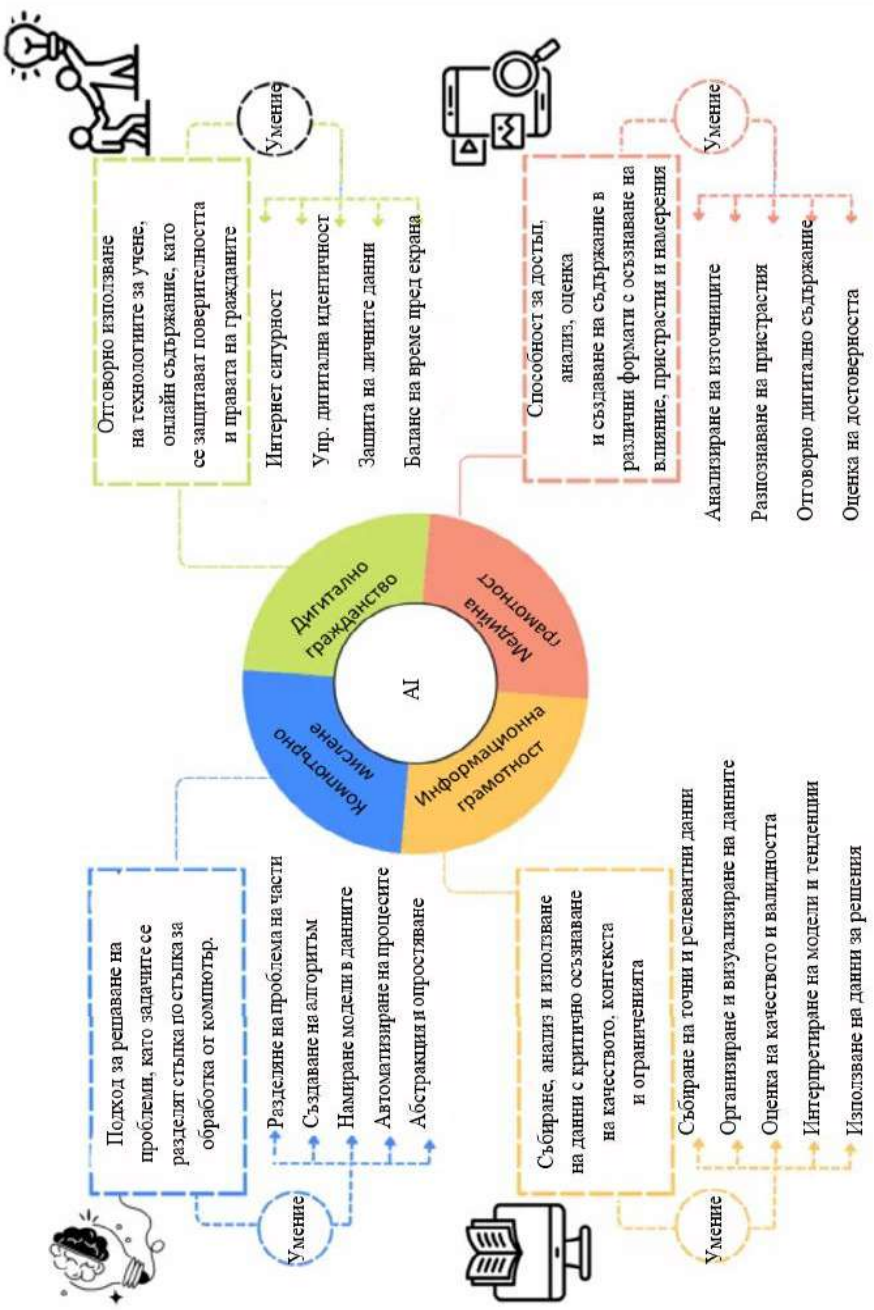
[9] NATO, 2023. *Digital Transformation in Defence*. NATO Standardization Office: Brussels

логистиката, усъвършенстване на анализите в реално време, подобряване на разузнаването, въоръжението и защитата на критична инфраструктура.

Европейската рамка DigComp не е единствената система за оценка и развитие на дигитални компетенции. Във военния контекст съществуват специализирани рамки, които допълват или адаптират DigComp за специфичните изисквания на отбраната и националната сигурност. DAPA-HR (Digital Acceleration Program for Advanced Human Resources) например поставя специален акцент върху AI-human teaming – способността на военнослужещите да работят ефективно с AI системи като партньори, а не просто като потребители на инструменти. Това изисква развитие на метакогнитивни умения – разбиране кога да се довериш на AI препоръка и кога да задействаш човешка преценка, а ICDL (International Computer Driving Licence) предоставя стандартизирана основа, върху която могат да се надграждат специфичните военни компетенции.

NATO Digital Transformation Framework предоставя собствена рамка за цифрова трансформация, адаптираща DigComp за нуждите на Алианса. В нея компетенциите са структурирани около способността за действие едновременно в космоса, киберпространството, въздуха, сушата, морето и информационното пространство. Мултидомейн операциите изискват умение да се интегрират данни от различни среди. Има необходимост както от стандартизирани компетенции, гарантиращи, че военнослужещи от различни страни могат да работят заедно в мултинационални сили, така и от хибридни компетенции – умения за работа с цифрови и с аналогови резервни системи.

Връзката между дигиталните компетентности и съвременните разбирания за интелект става все по-значима, особено на фона на разширяващата се роля на изкуствения интелект. За да се овладеят AI компетенциите, е необходимо базово разбиране на разликата между човешки и изкуствен интелект – не като технически детайл, а като основа за критична оценка на AI възможностите.

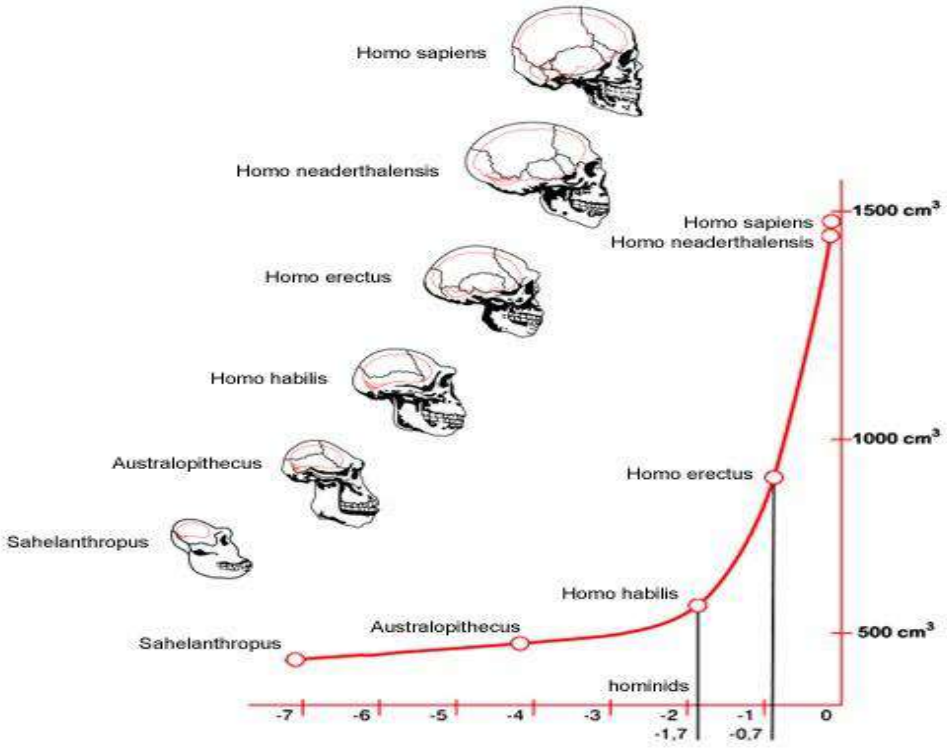


Фигура 4. Компоненти на AI грамотност

1.2. От дигитална компетентност към AI интелигентност

Интелигентността на животните се определя като комбинация от умения и способности, които им позволяват да преживяват и да се адаптират към специфичната си среда (Shettleworth 2010). Най-често се дават примери за животинска интелигентност с октоподите, делфините, враните и шимпанзетата, като общото между тях е наличието на сложна нервна система и мозък.

Човешкият мозък съдържа около 86 милиарда неврона и е структуриран в отделни области на функциониране (Herculano-Houzel 2009). Вероятно стремителното нарастване на обема на човешкия мозък преди два милиона години е свързано с преминаването от индивидуално към социално поведение и развитието на езика (Dunbar 1998).



Фигура 5. Еволюция на човешкия мозък

Една от първите дефиниции за интелигентността е на Карл Берейтър, който я определя като *това, което използвате, когато не знаете какво да правите* (Bereiter & Scardamalia 1993). Роберто Колом дефинира интелигентността като *общата умствена способност за разсъждение, решаване на проблеми и самообучение* (Colom, Stein et al. 2013). Интелигентността се разбира като способност за учене от опита, решаване на проблеми и адаптиране към средата. Робърт Стърнбърг разграничава три вида интелигентност: аналитична, творческа и практическа (Sternberg 1985). Напоследък се обръща сериозно внимание и на емоционалната интелигентност като основен фактор за социалната адаптация (Goleman 1995).

Ключово за компетентността е, че човешкият интелект интегрира емоционална и етична димензия, позволяващи стратегическо мислене и автентично творчество. Фундаменталната разлика между човешкия и изкуствения интелект се състои в наличието на съзнание, което обуславя реакции, базирани на вътрешна мотивация и субективна ценностна система. Това определя границите на отговорност при вземане на решения с AI подкрепа.

Изкуственият интелект е научна област от компютърните науки, насочена към проектирането на системи, способни да изпълняват задачи, които традиционно изискват човешка когнитивна намеса (Simon 1969). По тази причина учените дефинират AI като *способност на цифровата машина или компютъра да изпълнява задачи, като имитира интелигентни същества* (Russell & Norvig 2021).

Важно е да се разбере, че макар AI системите да дублират функции на човешкия интелект, те използват коренно различни методологични подходи. Днешните модели са предимно обучени алгоритми, които работят с масиви от данни и висока изчислителна мощ, за да симулират разбиране на сложни ситуации с минимална човешка намеса. Основните характеристики, които ги отличават от биологичния интелект, включват липса на емоционален интелект, флуиден коефициент на интелигентност (AI не притежава фиксиран IQ) и прецизност при големи данни (LeCun, Bengio & Hinton 2015).

Изкуственият интелект преминава през три еволюционни фази на развитие (Goertzel & Pennachin 2007):

● Изкуствен интелект с *тесен (ограничен) интелект* (ANI – Artificial Narrow Intelligence) – текущото състояние на технологията. Това е ограничен изкуствен интелект, който може да изпълнява само конкретни задачи. Типичен пример е системата Deep Blue на IBM, победила световния шампион по шахмат Гари Каспаров през 1997 г. (Campbell, Hoane & Hsu 2002). След това дойде ред на известната и доста по-трудна китайска игра Го. През 2016 г.

Google DeepMind успя да победи най-добрите професионалисти на тази игра със своята програма AlphaGo (Silver et al. 2016). Обаче програмата, която победи Каспаров, не може да победи шампиона на Го и обратното – тези ограничени типове изкуствен интелект блестят във всяка конкретна задача, но не могат да постигнат нищо извън областта, в която са тренирани. За военния специалист е критично да разбира, че работи с ANI системи – мощни в специфични задачи, но изискващи човешки надзор и валидация.

● *Обобщен универсален* изкуствен интелект (AGI – Artificial General Intelligence) – следващият етап от еволюцията, който може да изпълнява широк кръг от задачи. Той притежава елементарна форма на самосъзнание и подобрени умствени способности, които го правят сравним с човешкия интелект (Goertzel & Pennachin 2007). AGI представлява теоретична система с интелектуални способности еквивалентни на човешките, умееща да трансферира знания между различни области.

● *Изкуствен суперинтелект* (ASI – Artificial Superintelligence) – последната фаза, в която машината надминава колективния човешки интелект във всички сфери (Bostrom 2014). Това е хипотетична точка на „технологична сингулярност“, при която AI системите стават способни да подобряват себе си експоненциално, изпреварвайки човешките възможности за контрол.



Фигура 6. Технологична сингулярност

Една деликатна тема за бъдещето на изкуствения интелект е хипотезата за технологичната сингулярност. В основата на тази идея е допускането, че съществува огромна разлика в темповете на развитие на човешкия и напредъка на машинния интелект (Kurzweil 2005). В книгата „Сингулярността е близо“ той се аргументира с експоненциалния растеж на изчислителната

мощ. Законът на Мур предсказва ежегодно удвояване на капацитета. До достигането на физическите си граници капацитетът на интегралните схеми вероятно ще продължи да нараства, доближавайки се до изчисли-телната мощ на човешкия мозък и ще се удвоява всяка година. (Kurzweil 2025). Хронологията на Кърцуайл за еволюцията предвижда, че през 2029 г. ще достигнем нивото на AGI, а през 2045 г. – сингулярността, при която AI ще е по-добър от хората във всички задачи.

Идеята за технологичната сингулярност предполага, че поради бързото развитие на технологиите в определен момент изкуственият интелект ще се доближи до човешкия и двата ще се интегрират. Тази интеграция би могла да доведе до експоненциално нарастване на интелектуалните възможности на човека. Поради тази причина Илон Мъск инвестира значителни средства в компанията Neuralink, която се опитва да помогне за осъществяването на идеята за сингулярност с помощта на чипове в човешкия мозък (Musk 2020). При успех тези технологии биха могли да облекчат условията на живот на пациенти, страдащи от болестта на Паркинсон или различни видове парализа.

Макар оптимистичните прогнози да сочат бързо постигане на AGI, научната общност идентифицира няколко критични проблеми (Marcus & Davis 2019):

– енергийната ефективност на настоящите големи езикови модели (LLM), които консумират между 1 и 10 kWh на милион токена. В контраст, човешкият мозък функционира с константна мощност от около 20 W. Постигането на биологична ефективност изисква хилядократно подобрение в архитектурите или преминаване към нови парадигми (Patterson et al. 2021);

– насищането с данни – моделите изчерпват наличните качествени текстови данни в интернет. Истинският AGI изисква преход към обучение в реални физически взаимодействия и мултимодално разсъждение (Villalobos et al. 2022);

– изчисление в тестово време – пробиви като OpenAI o1 показват, че логическото мислене се подобрява, когато моделът разполага с повече време за „размисъл“. Това обаче създава латентност, която е неприложима в реално време за автономни системи (Wei et al. 2022).

Изкуственият интелект в момента широко се употребява от търсещите машини в интернет, спам филтрите, социалните мрежи и навигационните приложения, но в бъдеще неговата употреба ще нараства значително. В основата на това твърдение е превръщането на изкуствения интелект в съществен икономически фактор. Затова в близко бъдеще се очаква да има

пробив в автономните превозни средства (autonomous vehicles) и метавселената (metaverse).

Изследвания от 2024 г. сочат значителна диференциация в приемането на AI технологиите. Докато в САЩ внедряването на генеративен AI е достигнало 45,9%, в Европа и конкретно в българските въоръжени сили, процесът е забавен от социално-икономически и технически бариери (Singla et al. 2024). Трансформацията не означава пълна замяна на човека, а промяна в ролята на специалиста към дейности с висока добавена стойност – стратегическо планиране и етичен контрол. В контекста на отбраната, дигиталната трансформация превръща информационната среда в ново, динамично бойно поле.[10]

Отличен пример за малка държава, достигнала стратегическо предимство чрез всеобхватна цифрова трансформация и масови инвестиции в дигитална компетентност, е Естония. Това е първата държава, предложила „електронно гражданство“ (e-Residency), като успешно прилага уменията, заложени в рамката DigComp. В областта на информираността и управлението на данни естонската отбрана използва X-road – децентрализирана платформа за обмен на данни между държавни агенции. Военният персонал в Естония може да използва кръстосан AI анализ на данни, като комбинира информация от МВР, здравеопазване и отбраната. След мащабната кибератака през 2007 г. – първата по рода си кибервойна срещу държава – Естония развви уникална култура на „киберхигиена“. Всеки военнослужещ преминава редовно обучение за разпознаване на усъвършенствани постоянни заплахи (APT – Advanced Persistent Threat). Критичното мислене е друг ключов компонент на естонския подход. Иновативната концепция data embassies – сървъри, разположени в чужди държави и съхраняващи критични национални данни – изисква от командирите да разбират в дълбочина правните и техническите аспекти на облачната сигурност.

Друг пример е войната в Украйна – първият мащабен конфликт, където AI инструменти са достъпни за масова употреба от двете страни. Украинските сили използват AI за бързо генериране на пропагандни материали, но също така за автоматизирано разпознаване на руски военни единици от сателитни снимки. Това изисква компетенции по prompt engineering за създаване на съдържание с точно формулиране на заявки към AI системи. С помощта на уменията за критично мислене и оценяване украинските аналитици разработиха методи за валидация на AI генерирани разузнавателни

[10] NATO, 2023. Digital Transformation in Defence. NATO Standardization Office: Brussels

данни, комбинирайки автоматичен анализ с човешка експертиза. Приложението Signal и различни AI базирани инструменти за превод позволиха координация между украински сили и международни доброволци без общ език. И двете страни използват deepfake технологии за дезинформация. Украинските военни получиха обучение за разпознаване на манипулирани видеа – приложение на медийната грамотност в бойни условия. Урок за България е, че AI компетенциите стават критични за оперативното превъзходство в реален конфликт, но изискват баланс между автоматизация и човешки контрол.

Съединените американски щати създадоха Project Maven – първата мащабна военна програма за интеграция на AI в бойни операции, фокусирана върху анализ на видео от дронове. Системата обработва милиони часове видеозаписи, които човешки аналитици не могат да прегледат. Военният персонал трябва да формулира правилни заявки към системата и да интерпретира нейните оценки на сигурност. Проектът предизвика вътрешни протести в Google, което доведе до приемането на етични принципи за изкуствен интелект в Министерството на войната на САЩ[11]. Това наложи нов компетентностен слой – способността на офицерите да оценяват етичните аспекти на препоръки, дадени от AI. Компетентността се състои в разбиране кога AI е надежден (разпознаване на превозни средства) и кога не (разпознаване на намерения в тълпа). Това води до урок за нас, че внедряването на AI в бойни системи изисква предварително изграждане на етична рамка и компетенции за отговорно вземане на решения, а не само технически умения.

Дигиталната компетентност в ерата на изкуствения интелект обединява традиционни дигитални умения с нови компетенции, свързани с разбирането и прилагането на AI технологии. Европейската рамка DigComp 3.0 предоставя структуриран подход за оценка и развитие на тези компетенции, който е пряко приложим в сектора на сигурността и отбраната. Разграничаването между човешки и изкуствен интелект е от съществено значение за разбирането на възможностите и ограниченията на AI системите. Докато човешкият интелект се характеризира със съзнание, емоционална интелигентност и етика, изкуственият интелект предлага прецизност при обработка на големи данни и способност за автоматизация на специфични задачи. Внедряването на

---

11. US DEPARTMENT OF WAR, 2020. Department of Defense Adopts Ethical Principles for Artificial Intelligence. [Onlirne]. Available from: https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

AI в отбраната изисква системен подход към обучението на персонала, развитие на критично мислене и етична рамка за отговорна употреба на технологиите. Бъдещото развитие на AI ще продължи да трансформира отбранителните способности, като същевременно поставя нови предизвикателства пред сигурността и стратегическото планиране.

ЕВОЛЮЦИЯ НА ИЗКУСТВЕНИЯ ИНТЕЛЕКТ И ДИГИТАЛНАТА ИНФРАСТРУКТУРА

Всяка изчислителна машина,
която решава сложна математическа задача,
трябва да бъде програмирана за тази цел.

Джон фон Нойман

Фундаменталните въпроси, свързани с възможността формалните правила да пораждат валидно знание, с произхода на мисълта от материята и с трансформацията на знанието в действие, са поставяни още от античността. Математици като Евклид (ок. 300 пр.н.е.) полагат основите на аксиоматичния метод, докато през XIX век Джордж Бул разработва формалната логика като система, позволяваща извеждането на заключения независимо от конкретното съдържание (Boole 1854). Икономическата наука допринася с рамки за рационален избор чрез теориите за вероятността и полезността (Neumann & Morgenstern, 1944), докато невронауката и психологията започват да разглеждат мозъка и човешкото поведение като информационни системи (Ashby 1956). Лингвистикатаизследва критичната връзка между езика и мисленето, оформяйки теоретичните предпоставки за машинен интелект още преди появата на съвременните компютри (Chomsky 1957).

2.1. Философско-математически основи на изкуствения интелект

При изучаването на всяка област на човешкото знание съществуват три фундаментални въпроса: „какво?“, „защо?“ и „как“. Те определят три важни вида теории:

– класификационните теории се занимават с въпроса „какво?“, като целта е да се установи какви феномени съществуват в изследваната област. Резултатът е класификационен списък;

– епистемологичните теории се занимават с въпроса „защо?“. Тук стремежът е да се намери обяснение на установените феномени чрез анализ

в контекста на целите и причините за тяхното появяване, както и на взаимовръзките с други феномени;

– механизиращите теории се занимават с въпроса „как?“. Този въпрос цели разкриването на механизми (структури и процеси), които довеждат до проявата на наблюдавания феномен. Тези механизми обикновено се използват за възпроизвеждане на феномена в изкуствена среда.

Изкуственият интелект се различава от другите когнитивни науки не толкова по обекта на изследване (интелекта и интелигентното поведение), колкото по използваната методология. В повечето когнитивни науки се използва индуктивният метод: започва се с наблюдение на факти за поведението на естествена система, следва преход към формулиране на теория чрез обобщаване, а проверката се състои в съпоставяне на генерираните хипотези с реално наблюдавани факти. При несъвпадение настъпва промяна на теорията.

Синтезният (инженерният) метод, характерен за AI, започва по същия начин, но вместо да се генерират хипотези, теорията се използва като основа за създаване на изкуствени системи, които пораждат екземпляри на съответно поведение. Проверката на теорията се състои в оценка на поведението ѝ по определен критерий. При отрицателна оценка причината може да е както в теорията, така и в технологията на изграждане.

Идеята за автоматизирани механизми, имитиращи човешки действия, има дълбоки исторически корени. През 1206 г. арабският инженер Ал-Джазари конструира сложни механични автомати, включително музикална лодка с четирима автоматични музиканти (Al-Jazari 1974). През XVIII – XIX век става популярен шахматният автомат „Турчинът“, макар впоследствие да бива разкрит като измама със скрит човешки оператор (Standage 2002).

Истинският научен пробив настъпва през 40-те години на XX век с концепцията, че мисленето може да бъде моделирано като изчисление. През 1943 г. Уорън Маккълок и Уолтър Питс публикуват пионерска статия, в която предлагат първия математически модел на изкуствен неврон. Те демонстрират, че мрежа от прости логически елементи може да изчислява всяка функция, която е изчислима от машина на Тюринг (McCulloch & Pitts 1943).

През 1950 г. Алън Тюринг публикува есето Computing Machinery and Intelligence, където не просто пита дали машините могат да мислят, а предлага операционален критерий за разпознаване на интелигентност – прочутата „имитационна игра“, по-късно известна като Тест на Тюринг (Turing 1950). Въпреки че Тюринг прогнозира, че до 2000 г. машините ще успеят да измамят 30% от оценителите, дори и днес няма консенсус относно машина, преминала теста при строги условия (French 2000).

Критичните технически предпоставки за развитието на AI се формират през периода на Втората световна война и непосредствено след нея. Клод Шанън формулира математическата теория на комуникацията (Shannon 1949), Норберт Винер основава кибернетиката като наука за управлението и комуникацията в животни и машини (Wiener 1948), а Джон фон Нойман разработва архитектурата на съхраняващите програми компютри (Neumann 1945). Тези интелектуални постижения създават необходимата теоретична и техническа база за възникването на изкуствения интелект като самостоятелна научна дисциплина.

През 1956 г. се провежда лятната конференция в Дартмутския колеж, организирана от Джон Маккарти, Марвин Мински, Натаниел Рочестър и Клод Шанън, където официално е въведен терминът Artificial Intelligence (McCarthy et al. 2006). Участниците формулират оптимистичната хипотеза, че „всяка характеристика на ученето или всяка друга характеристика на интелекта могат в принцип да бъдат описани толкова прецизно, че да бъдат симулирани от машина“. В рамките на конференцията са очертани седем ключови изследователски направления: автоматични компютри, програмиране на естествен език, невронни мрежи, теория на сложността на изчисленията, самоусъвършенстване, абстракции и случайност и творчество.

Като наименование на научна област, терминът „изкуствен интелект“ бива критикуван поради смесването на наименованието на обекта на изследване с наименованието на самата област. През 1980 г. председателят на Европейската асоциация по AI Волфганг Бибел предлага термина „интелектика“, който все по-често се употребява като официално име на областта в редица европейски университети.

Изкуственият интелект е познавателна и инженерна научна област, която се занимава с разбиране и синтезиране на пораждането от интелекта интелигентно поведение. Под „интелигентно поведение“ най-общо се разбира взаимодействие между система и среда, характерно за което е възможността да се знае, разсъждава и действа целенасочено. Обикновено то включва перцепция (възприемане, обработка и интерпретиране на информация събрана чрез сензори от заобикалящата среда), логически разсъждения, обучаване, вземане на решения и действане.

„Интелект“ е множество от специфични способности на една система, насочени към пораждане на интелигентно поведение. Всяка система, която е способна да води интелигентно поведение, се нарича „интелектна система“.

Като познавателна наука ИИ се занимава (наред с психология, невробиология, философия, лингвистика и др.) с формиране на терминология,

модели и теории с цел разбиране на принципите, които правят интелигентното поведение възможно. Това, което различава ИИ от чисто познавателните науки, е насочеността му към синтезирането и изграждането на изкуствени (синтетични) интелектни системи, които от една страна служат като средство за тестване на създаваните теории, а от друга – са практически полезни на хората. Именно това определя и инженерния аспект на ИИ – разработването и използването на подходи, формализми и методи за синтез на реални системи с интелигентно поведение.

Основна научна цел на ИИ е разбиране на принципите, които правят интелигентното поведение възможно. Основна инженерна цел на ИИ е изграждане на интелектни системи.

В съвременната наука доминира разбирането, че няма принципни ограничения за моделиране на естествения интелект и за възпроизвеждане в изкуствена среда на различни специфични интелектуални способности на човека. Това разбиране е свързано с теоремата на Гьодел, която (съдържателно) гласи, че вътрешната непротиворечивост на една логическа система не може да се докаже със средствата на самата система. Съвременната логика не доказва невъзможността, нито пълнотата на възможността за възпроизвеждане на интелигентно човешко поведение в изкуствени среди.

Начинът на възпроизвеждане не е задължително да бъде еквивалентен на начина на осъществяване на феномена в биологичните системи. Първият полет на човека е осъществен след като той най-после се е отказал да имитира полета на птиците. Това е подход към отговор на въпроса „Може ли една машина реално да мисли?“ чрез разглеждане на аналогичния въпрос „Може ли една машина реално да лети?“.

Провеждани са редица научни изследвания върху ефективността на подходите на ИИ при решаването на разнообразни задачи от практиката. Установено е, че за широк клас задачи, от една страна, класическите подходи и компютърни програми не са подходящи, а от друга страна, хората не са в състояние да вземат правилни решения с необходимата точност и бързина. Това обяснява мотивите за бурното развитие на ИИ.

Разработените технологии на ИИ са намерили изключително широки приложения – от системи, основани на знания (експертни системи), с приложения в химията, биологията, медицината, транспорта, образованието, правото, икономиката, държавната администрация, електрониката, военното дело, машиностроенето, енергетиката и др., до интелигентни интерфейси, системи за автоматичен превод, системи за автоматизация на проектирането, интелигентни системи за обучение, интелектни роботи и автономни агенти.

Съществува клас интелектни системи, които могат да бъдат по-интелигентни от човешкия индивид – класът на организациите. Една фирма като цяло може да проектира, произвежда, продава хиляди компютри, за което са необходими сумарни интелектуални умения, надхвърлящи възможностите на един човешки индивид.

Изкуственият интелект се стреми да проверява валидността на хипотезите за естеството на интелигентното поведение, като синтезира интелектни системи, които не са задължително структурни аналози на човешкия мозък. Предмет на изучаване и синтез е интелигентното поведение както на биологични, така и на организационни обекти.

Целта на ИИ не е компютърната симулация (както често се смята), а разбирането на интелигентното поведение чрез синтез на реални системи, които го осъществяват. От една страна, компютърът е физическа знакова система и осъществява интелигентно поведение на системи с перцепция и действия в еднородна, чисто компютърна (знакова), среда. От друга страна, компютърното симулиране на движението на манипулатора на един робот, например, поражда знакови описания на екземпляри на поведение, които не са реално възпроизведени, тъй като извън компютъра сили и моменти не съществуват. Аналогичен е и случаят при симулиране на визуална, вербална и др. видове перцепция – липсва връзката с реалния свят. За нейното осъществяване са необходими нетривиални интерфейси (сензори и изпълнителни механизми), които изискват среда за реализация, различна от компютърната.

От тази гледна точка, роботиката играе важна методологична роля не само за ИИ, но и за науката въобще. Роботиката е научна област, същността на която е интеграцията на перцепция, разсъждения и действия. Роботите са полезни интелектни системи за действия в нееднородни физически среди. Чрез конструирането на роботи се адресират не само въпросите за ролята на вътрешните механизми (знаковите преобразувания) в интелигентното поведение на една система, но също така и въпросите за ролята на онези механизми, които определят връзката с реалния свят.

2.2. Основни научни подходи за изучаване на AI

Въпреки огромния напредък в науката за изкуствения интелект, феномените „интелект“ и „интелигентно поведение“ имат редица аспекти и проблеми. Много от тях все още са само частично разбрани и съответно остават нерешени, но има разнообразие от подходи към тяхното изучаване и

възпроизвеждане. Още от Дартмутската конференция от 1956 г. до 1969 г., развитието на изкуствения интелект се характеризира с два отделни и до известна степен противостоящи подхода – знаков (символен) и поведенчески (числов).

Знаковият подход има своите корени в логиката, лингвистиката и психологията. В основата му стои хипотезата на Алън Нюел, известна като *Хипотеза за физическата знакова (символна) система*. Тя твърди, че интелигентното поведение изисква знакови разсъждения върху знаково представени знания. На тази основа се приема, че вътрешната структура на една интелектна система се основава на знаково представяне на информацията, която тя използва. Съществено тук е, че представянето се извършва на ниво смислова информация (знания), а не на ниво данни. Като следствие от това вътрешните за системата процеси се състоят в обработване на семантично натоварени знакови структури. Методите за такъв тип знаково представяне и обработка на информация са няколко вида: продукционни системи, фрейми, логическо програмиране. Това представяне се оказва забележително стабилно и предпочитано през 80-те години на XX в.

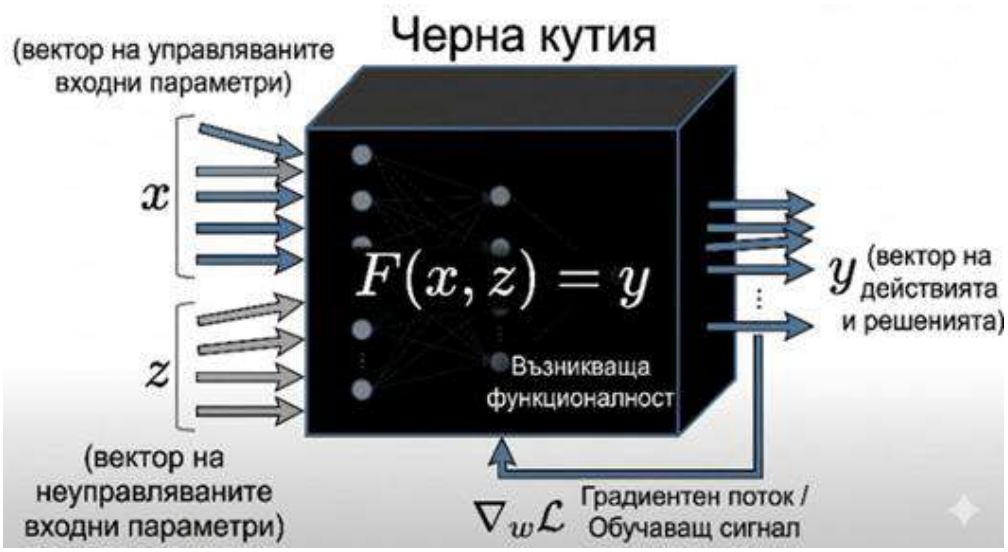
Поведенческият подход в изкуствения интелект има своите корени в биологията, неврофизиологията, етологията и еволюционната теория. При него акцентът е снет от вътрешните знакови модели и процесите над тях и е поставен върху интерактивните процеси между външната среда и изкуствената система. При този подход тя се разглежда като сложна динамична система, намираща се в непрекъснато взаимодействие с процесите от реалния свят. Съществено тук е, че директно се поставя въпросът как една физическа система може да извършва комплексно необходимите (за определена задача) перцепция, взимане на решения и действия, за да оперира автономно (независимо от човека) в реално време и в динамично променяща се среда.

Основната хипотеза, която се издига в отговор на този въпрос, е, че интелигентно поведение може да бъде породено от голяма колекция относително независими, елементарни (не предполагащи разумност) поведения без каквото и да е централизирано управление над тях. Всяко елементарно поведение се реализира от строго специализиран (само за това поведение) модул, който разполага с необходимите за целта сензори и изпълнителни механизми. Всеки такъв модул възприема външната среда от „своя гледна точка“ (т.е. извлича от сензорите само информацията, необходима за постигането на поведението, за което той е отговорен) и е в състояние да реагира адекватно и незабавно чрез изпълнителните си органи.

Поведенческият подход е една разновидност на динамичния еволюционен подход, който се опитва пряко да емулира дейността на централната нервна система на човека и висшите животни. Нервната система е организирана от огромно количество елементарни клетки (неврони), които сами по себе си не са интелигентни. Интелектът се поражда от поведението на цялата сложна динамична система, съставена от елементарните обработващи единици – невроните. Мрежата от взаимносвързани неврони до голяма степен прилича на електронна схема, която преработва непрекъснати числови сигнали. Съгласно този подход интелектните системи формират адаптивно непрекъснати функции, които апроксимират големи количества данни, без при това да специфицират явно (чрез знакови модели) входно-изходните зависимости на системата. В този смисъл, една система има интелигентно поведение, ако при някакви входни стимули тя адаптивно ги апроксимира и генерира подходящи за решаване на дадени задачи въздействия. Много често за онагледяване се използва принципът на черната кутия, в която се намира изследваната система, а отвън ѝ въздействат контролирани и управлявани параметри, промяната на които води до промяна на действията на системата, без да знаем какво се случва в нея.

Обща основа и на двата подхода са кибернетиката и ранните опити от средата на XX век да се анализира устройството на сложни естествени системи. И в двата подхода интелектната система преработва информация – в знаков вид при знаковия подход и във вид на сигнали при числовия. Конкретните механизми обаче са различни, знаковият подход използва предимно формалната логика, докато числовият се основава главно на теорията на динамичните системи и системотехниката. И двата подхода имат своите положителни и отрицателни страни, затова се налага емпирична проверка.

Развитието на двете направления върви успоредно – Франк Розенблат разработва в Корнел теорията на определен тип прости невронни мрежи, наречени от него *перцептрони*, а Нюел и Саймън създават нова и още по-впечатляваща знакова система, наречена „*Общ решител на задачи*“ (General Problem Solver). През 1960 г. Уидроу и Хоф от Станфорд откриват начин да обучават перцептроните на Розенблат според принципи, известни от психологията и неврофизиологията.



Фигура 7. Принцип на черната кутия за изследване на системи

Това паралелно развитие на двата подхода спира през 1969 г., когато Марвин Мински и Сеймур Пейпърт от Масачусетския Технологичен Институт доказват, че елементарните невронни мрежи на Розенблат не притежават достатъчни изчислителни качества и по този начин не могат да конкурират знаковите системи. Разбира се, това доказателство не е задължително валидно за по-сложни невронни мрежи, но този факт се пропуска от множеството учени по това време и води до изоставяне на числовия подход за повече от 15 години. За щастие, други дисциплини (породени също от кибернетиката) продължават да го използват. В теорията на управлението и системотехниката се разработват адаптивни системи за управление, които притежават някои от качествата, характерни за интелигентните системи – например способността да прилагат подходящ управляващ сигнал в променящи се условия. В областта на изследване на операциите пък се разработват алгоритми за планиране в реално време посредством динамично програмиране, които са директно приложими в системите с изкуствен интелект. Изследването на поведението на големи системи от прости и взаимносвързани обработващи елементи продължава в статистическата физика. Приложимостта на такива системи като модели на човешката памет се изследва в когнитивната психология.

В рамките на изкуствения интелект периодът между 1970 и 1985 г. е изцяло доминиран от знаковите системи. През това време знаковият подход отбелязва няколко важни постижения, най-значимото от които са експертните системи – програми, в които е заложено в явен вид знанието на експерти в дадена предметна област, например медицина, право, счетоводство. Успоредно с това става ясно, че извличането на експертно знание е сложен и скъп процес, което ограничава разпространението на експертните системи. Необходимостта от обучение и самообучение става все по-явна в светлината на тези проблеми.

Насърчен от затрудненията на знаковите системи, числовият подход се възражда отново при ИИ през втората половина на 80-те години на XX в. и бързо набира популярност под името *конекционизъм*. Основното влияние за този процес идва от страна на когнитивната психология и отчасти от страна на статистическата физика.

Знаковият и числовият подход не си противоречат, а се допълват. Първият предоставя по-удобни средства за възпроизвеждане на „съзнателното“ (разсъдъчното, глобалното) поведение, а вторият – на „подсъзнателното“ (рефлексното, реактивното) поведение. В общия случай, една пълноценна интелектна система трябва да разполага със средства за съчетаване и на двата класа поведение.

В самия ИИ се появиха нови подобласти, каквито са например: вероятностните мрежи, които на практика извършват знакови разсъждения по числов път; системите с размита логика, използватщи знакови правила за обработка на числова информация; невронните мрежи, основани на знания, позволяващи въвеждане и извеждане на знания (от обучение) във вид на знакови правила. Научните области теория на управлението и изследване на операциите също се сближават с ИИ. Съвсем определено вече се очертават рамките на една по-пълна теория на интелигентното поведение, която използва и двата подхода, като се стреми към взаимното им проникване и интегриране.

2.3. Ранен ентусиазъм и първите AI зими

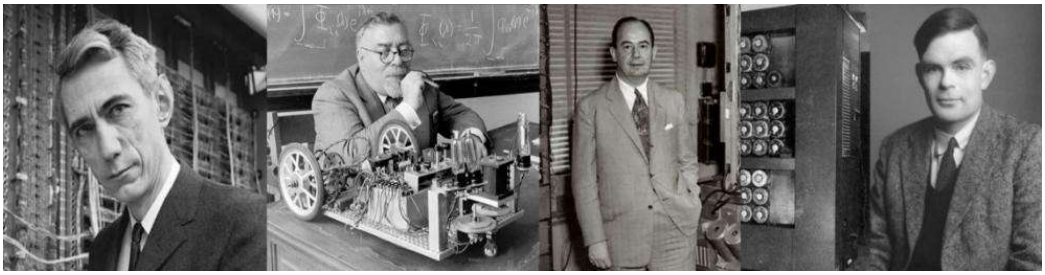
Първото десетилетие след Дартмут е период на значителен ентусиазъм и сравнително щедро финансиране, включително от военни агенции като DARPA (Defense Advanced Research Projects Agency), които виждат потенциал за стратегическо предимство в контекста на Студената война

(Crevier 1993). През 1957 г. Франк Розенблат разработва Перцептрона – първият изкуствен неврон, и компютърна система, способна да се обучава без пряко програмиране, базирана на хардуерна реализация на невронна мрежа (Rosenblatt 1958). Идеята изпреварва времето си, защото все още липсва необходимата изчислителна мощност и подходящи алгоритми за обучение с достъпни данни. През 1960 г. Бърнард Уидроу и Тед Хоф предлагат модела ADALINE (Adaptive Linear Neuron), който демонстрира възможността за аналогово машинно обучение чрез адаптивни линейни елементи (Widrow & Hoff 1960).

Паралелно се развиват символните подходи. Алън Нюел и Хърбърт Саймън разработват Logic Theorist (1956) и по-късно General Problem Solver (1959), поставяйки началото на евристичното търсене и символното разсъждение (Newell & Simon 1963). През 1965 г. Едуард Фейгенбаум и Джошуа Ледърбърг разработват първата експертна система DENDRAL, способна да анализира химични структури (Lindsay et al. 1980).

Въпреки тези постижения, оптимистичните прогнози на пионерите се оказват преждевременни. През 1958 г. Саймън и Нюел прогнозират, че до десет години компютър ще стане световен шампион по шахмат и ще открие важна математическа теорема (Simon & Newell 1958). През 1965 г. Саймън заявява, че до 20 години машините ще могат да изпълняват всяка работа, която човек може да свърши (Simon 1965), а през 1967 г. Мински прогнозира решаването на проблема за изкуствения интелект в рамките на едно поколение (Minsky 1967).

Реалните технологични ограничения обаче са съществени: компютрите разполагат с минимална памет (първите експерименти с обработка на естествен език (NLP) използват речници от едва 20 думи) и ниска скорост на обработка. През 1969 г. Марвин Мински и Сеймур Паперт публикуват книгата Perceptrons, в която математически доказват ограниченията на еднослойните перцептрони – конкретно невъзможността им да решат задачата за изключващото (XOR) (Minsky & Papert, 1969). Този резултат, в комбинация с неизпълнените обещания, води до рязко намаляване на финансирането и настъпването на първата „зима на изкуствения интелект“ (Crevier 1993), когато разочарованите правителствени агенции спират финансирането на научното направление за повече от десетилетие (Rickards, 2025).



Фигура 8. Първите учени на изкуствения интелект (от ляво надясно) – Клод Шенън, Норбърт Винер, Джон фон Нойман и Алън Тюринг

2.4. Експертни интелектни системи

През 70-те и 80-те години на XX в. доминират експертните системи – подход, фокусиран върху кодирането на специализирано знание под формата на правила „ако – тогава“ (if-then). През 1976 г. е разработена системата MYCIN за медицинска диагностика (Shortliffe 1976), а през 1982 г. – първата търговска експертна система R1/XCON за конфигуриране на компютърни системи Digital Equipment Corporation (McDermott, 1982). Въпреки търговския успех на някои приложения, експертните системи се оказват крехки – знанието е трудно за извличане от експерти, правилата са статични и изискват ръчна актуализация, а системите не могат да се справят с неочаквани ситуации извън предварително дефинираната област.

През 1986 г. настъпва решаващ пробив в областта на невронните мрежи с публикуването на метода за обратно разпространение на грешката (backpropagation) от Дейвид Румелхарт, Джефри Хинтън и Роналд Уилямс (Rumelhart, Hinton & Williams 1986). Този алгоритъм позволява обучението на многослойни невронни мрежи, като преодолява ограниченията, идентифицирани от Мински и Паперт, и поставя началото на съвременното дълбоко обучение. През 1997 г. системата Deep Blue на IBM побеждава световния шампион по шахмат Гари Каспаров (Campbell, Hoane & Hsu 2002), а през 2016 г. AlphaGo на DeepMind побеждава шампиона по Го Ли Седол (Silver et al. 2016).

Системите с изкуствен интелект, основаващи се на правила, използват инструкции от типа „ако – тогава“, за да достигат до заключения на базата на първоначален набор от факти. В най-простата си форма едно твърдение "ако – тогава" казва на системата: „Ако това условие е вярно при дадените факти, то направи така. Ако е невярно, не прави нищо“. Добавянето на

опцията „иначе“ позволява формулирането на по-сложни твърдения: „Ако това твърдение е вярно, направи така, а ако не е (иначе) – направи нещо друго“. Основаващите се на правила системи са предвидими, надеждни и „прозрачни“, т.е. лесно е да се разбере кои правила прилага ИИ. Този тип ИИ обаче не може да се адаптира към нови ситуации извън предварително програмираните правила.

Компютърните програми, имитиращи знанията и моделите на разсъждение на човешките специалисти, се наричат експертни системи. Информацията, която те съдържат, се осигурява от хора специалисти и се програмира в системата от „програмист на знания“. Всяка система има три части:

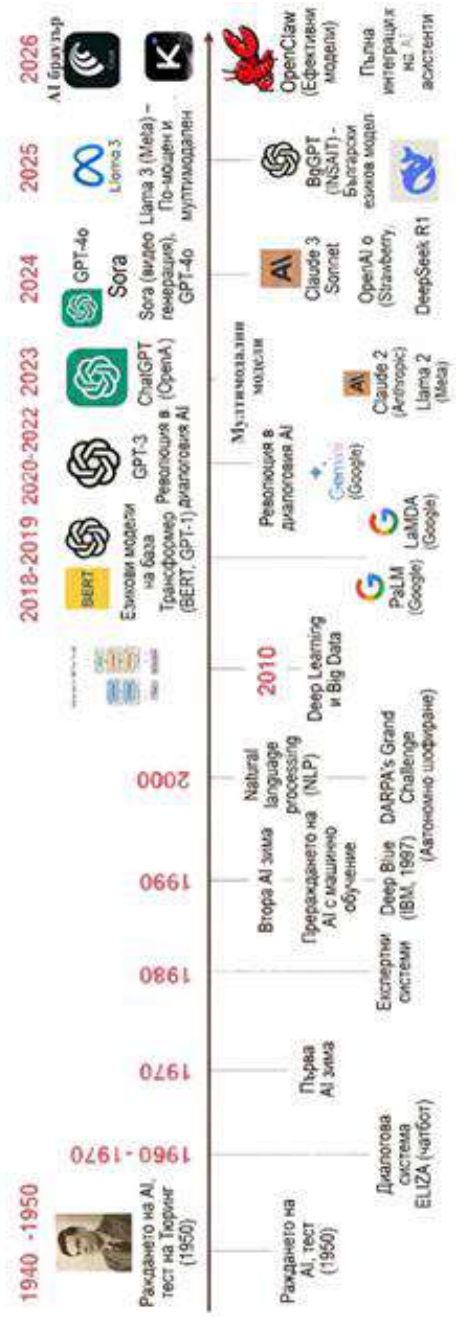
1. База от знания (Knowledge Base) – съдържа фактите и правилата (обикновено във формата „ако-тогава“), използвани от човешките експерти в съответната предметна област.

2. Механизъм за достигане до заключения (Inference Engine) – прилага правилата към фактите в базата от знания, за да достигне до заключения и отговори на зададените от потребителите въпроси. Този компонент реализира логическото разсъждение в системата в разбираем формат.

3. Потребителски интерфейс (User Interface) – приема запитванията на потребителите и им показва решенията, намерени от системата.

Експертните системи могат да отговарят на сложни въпроси и да осигуряват на потребителите по-широк достъп до експертни съвети. Те се използват в много области, включително медицината, където сравняват симптоми с вероятни причини и назначават подходящо лечение.

Исторически, експертните системи бележат първата вълна от търговски приложения на ИИ през 70-те и 80-те г. на XX в. с представители като MYCIN (медицинска диагностика) и DENDRAL (анализ на химични структури). Въпреки че са изместени от статистическите методи и машинното обучение в много области, принципите на експертните системи остават релевантни за системи, изискващи прозрачни и обясними решения – качества от критична важност за военни и правителствени приложения.



Фигура 9. Линия на времето за развитие на изкуствения интелект

2.5. Изграждане на дигиталната инфраструктура (WEB 1.0, 2.0, 3.0)

Паралелно с развитието на AI методологиите се формира и критичната комуникационна инфраструктура. През 60-те г. на XX в. в RAND Corporation Пол Баран разработва концепцията за разпределени мрежи с блоково предаване (packet switching), проектирани да устоят на ядрена атака (Baran 1964). През 1969 г. ARPANET свързва първите четири възела (UCLA, SRI, UCSB, University of Utah), демонстрирайки възможността за надеждна комуникация между разпределени компютри (Abbate 1999).

През 1974 г. Винт Сърф и Боб Кан публикуват протокола TCP/IP (Transmission Control Protocol/Internet Protocol), който позволява интернет работа между различни мрежи. През 1984 г. е въведена системата за имена на домейни (DNS – Domain Name System), превръщаща числовите IP адреси в четими имена (Mockapetris 1983). През 1989 г. Тим Бърнърс-Лий в CERN предлага системата за хипертекст World Wide Web, базирана на HTML, HTTP и URL (Berners-Lee 1989), превръщайки интернет от инфраструктура за специалисти в универсална среда за съдържание.

Инфраструктурата за дигитална интелигентност е готова. Появява се Web 1.0, който дава възможност за дигитализация на информацията. Той е статичен, еднопосочен и доминиран от институции и експерти, но браузъри като Mosaic и Netscape демократизират достъпа до информация. Интернет се превръща в глобален архив – решаващо условие за бъдещето обучение на AI.

През 2004 г. Тим O'Райли формулира концепцията Web 2.0 – платформа, при която съдържанието се генерира от потребителите (user-generated content), включваща блогове, социални мрежи (Facebook, 2004; Twitter, 2006; YouTube, 2005) и уики (O'Reilly 2005). Този етап създава социалния граф, реалновремеви комуникации и огромни корпуси от текст, изображения и видео – „горивото“ за съвременното машинно обучение.

Комбинацията от евтини изчисления (законът на Мур), мащабни мрежи (законът на Меткалф), облачни платформи и данни от Web 2.0 води до възраждането на невронните мрежи. Съоснователят на Intel Гордън Мур през 1965 г. твърди, че броят на транзисторите в един интегрален чип се удвоява приблизително на всеки две години, което води до експоненциално увеличаване на изчислителната мощ и същевременно до намаляване на цената на изчисленията. Това прави сложните изчисления и обработката на огромни обеми данни достъпни за все по-широк кръг от изследователи и

компании. Робърт Меткалф твърди, че стойността на една мрежа нараства пропорционално на квадрата на броя на свързаните към нея потребители. Това е основата за мащабното развитие на интернет и социалните мрежи, които осигуряват огромни масиви от данни за обучение на изкуствени невронни мрежи. В съчетание с развитието на облачните платформи, които позволяват лесно мащабиране и достъп до огромни изчислителни ресурси по заявка, и експлозията на данни, генерирани от потребителите в Web 2.0, тези фактори създават идеалната среда за възраждането и практическото приложение на невронните мрежи и дълбокото обучение. В резултат, изкуственият интелект прави скок от теоретична концепция към широко използвана технология, която трансформира редица индустрии и сфери на живота. Облачните изчисления (Cloud Computing), популяризирани от Amazon Web Services (2006), Google Cloud (2008) и Microsoft Azure (2010), предоставят мащабируем достъп до изчислителни ресурси (CPU, GPU, памет, съхранение) по заявка, елиминирайки необходимостта от капиталови инвестиции в хардуер (Armbrust et al. 2010).

2.6. Ерата на дълбокото обучение и големите езикови модели

През 2006 г. Джефри Хинтън и сътрудниците му демонстрират възможността за ефективно предварително обучение на дълбоки невронни мрежи (Hinton, Osindero & Teh, 2006), поставяйки началото на ерата на дълбокото обучение (Deep Learning). Идеята на дълбокото обучение е, че могат да бъдат обучени сложни компютърни модели от разнообразни структури на невронни мрежи, базирани на различни слоеве от изкуствени неврони. Най-известната архитектура е предложена през 2012 г. – AlexNet, разработена от Алекс Крижевски, Иля Суцкевър и Джефри Хинтън. Тя постига революционен успех в състезанието ImageNet, като намалява грешката при класификация на изображения с 10,8 процентни пункта (Krizhevsky, Sutskever & Hinton 2012). Този пробив привлича вниманието на технологичните гиганти и стимулира масивни инвестиции в AI изследвания.

В резултат болшинството от водещите учени, свързани с изкуствения интелект, вече е съсредоточено в технологичните гиганти. Фундаменталните изследвания на изкуствения интелект се трансформират в индустриални изследвания, т.е. фокусът е върху решаването на бизнес проблеми. Положителното в тази фаза е, че огромната концентрация на таланти и изчислителен

капацитет създава много нови методи, алгоритми и софтуерна инфраструктура на изкуствения интелект за кратък период от време.

Особено важна тенденция е споделянето от страна на повечето технологични гиганти на технически постижения с академичните и бизнес средите. В резултат едрият бизнес постепенно започна да използва споделените възможности и инфраструктурата на изкуствения интелект, за да решава своите проблеми по ефективен начин. Разходите за разработване на решения, базирани на изкуствен интелект, значително намаляват. В рамките на 10 – 15 години изкуственият интелект се внедрява с бясно темпо в бизнес гиганти като Walmart, Ford, Tesla, AT&T, General Electric и др.

Изкуственият интелект може да се обучава с помощта на невронните мрежи, дълбокото машинно обучение, машините с опорни вектори и алгоритъм на К-най-близки съседи (KNN). Друго функционално направление е умението на изкуствения интелект да предсказва и прогнозира причинно-следствени връзки чрез Байесово обучение, дърво на решенията и др. Все по-нарастващо е значението на основната функция на изкуствения интелект да разсъждава и да предлага автоматично решения на сложни проблеми. Тази функция се базира на методите на размита логика, експертни системи, графи на знанието и когнитивни модели. Съществува и обширна област за решаване на проблеми, където изкуственият интелект може да помогне да се оптимизират процеси с алгоритми като генетични алгоритми, оптимизация с колонии от мравки (Ant Colony Optimization – ACO) и оптимизация чрез симулация на ято птици.

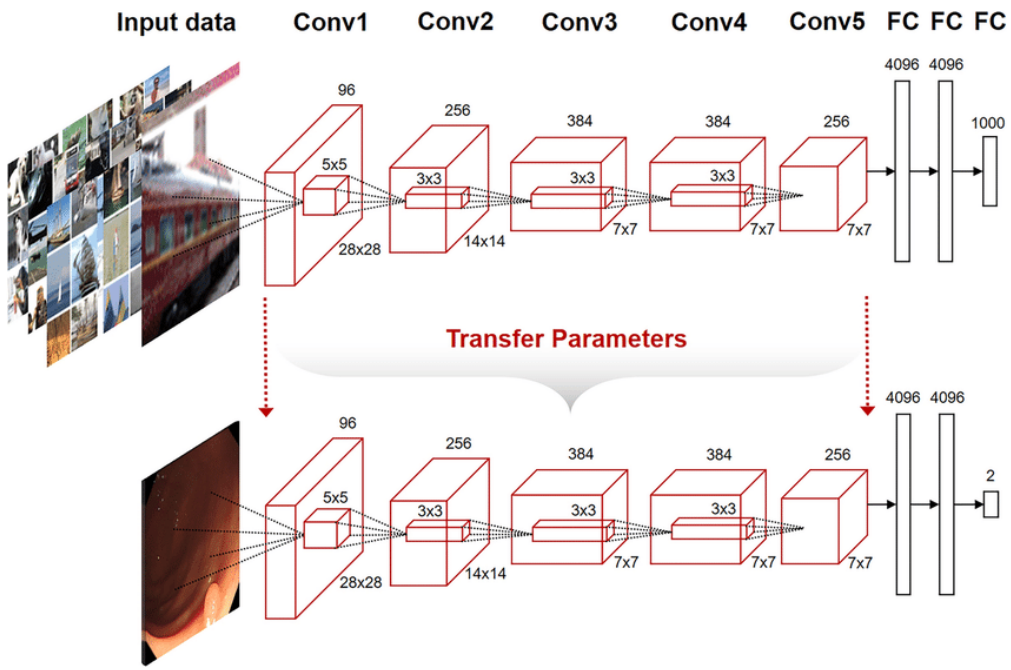
През 2017 г. изследователи от Google Brain публикуват статията „Attention Is All You Need“, въвеждайки трансформаторната архитектура, която позволява паралелна обработка на последователности и решава проблемите с дългосрочните зависимости в рекурентните мрежи (Vaswani et al. 2017). Това открива ерата на Големите езикови модели (Large Language Models – LLM).

През 2018 г. OpenAI публикува GPT-1 (Generative Pre-trained Transformer), демонстрирайки ефективността на предварителното обучение с последващо фино настройване (Radford et al. 2018). След 2020 г. настъпва генеративен взрив с GPT-3 (175B параметъра), който дори започва да води вестникарска колона и изкуственият интелект навлиза в масовото съзнание. Следва появата на множество мащабни модели: LLaMA, Claude, Mistral, GPT-4 и пускането на мултимодалните системи и AI агенти през 2023 г.

Появява се съвременната дигитална среда Web 3.0, която включва автономни участници (AI агенти), Интернет на нещата (IoT) за телеметрия от

физическия свят и блокчейн за сигурност и идентичност. Облачните изчисления се превръщат в двигател, предлагайки мащабируемост (GPU ресурси) и достъпност. ИИ вече не само намира, но и синтезира знание. Според прогнози на Gartner, до 2026 г. около 90% от онлайн съдържанието ще бъде генерирано от AI (Gartner 2023).

През 2023 – 2024 г. се наблюдава експлозия от модели: GPT-4 (OpenAI), LLaMA (Meta), Claude (Anthropic), Gemini (Google), Mistral (Mistral AI), както и появата на мултимодални системи и AI агенти. През 2024 г. китайският модел DeepSeek-R1 демонстрира конкурентоспособни възможности при значително по-ниски изчислителни разходи (DeepSeek-AI 2025). През 2024 г. навлизат нови агенти като DeepSeek и интегриране на ИИ директно в браузърите (Edge с Copilot, Chrome с Gemini).



Фигура 10. Архитектура на AlexNet, базирана на няколко нива на невронни мрежи, означени с отделни блокове (conv1 – conv5) и три функционални нива (fc)

2.7. AI и съвременната дигитална среда

Изкуственият интелект не съществува във вакуум. Той е върхът на пирамида, изградена върху десетилетия развитие на компютърните мрежи, хардуера и протоколите за обмен на данни. За да се разбере как AI „мисли“ и оперира, първо трябва да се познава средата, в която той „живее“ – инфраструктурата на интернет. Всяко AI приложение разчита на „светата троица" на уеб инфраструктурата:

● Домейн (Domain) – адресът на системата в мрежата (напр. mod.bg). Той е човешкият интерфейс към сложната цифрова адресация.

● Хостинг (Hosting) – физическото място (сървър), където „живее мозъкът“ на AI. Хостингът предоставя дисково пространство и изчислителна мощ, където се съхраняват алгоритмите, моделите и базите данни.

● DNS (Domain Name System) – системата, преобразуваща домейните в IP адреси. Компрометирането на DNS може мигновено да прекъсне достъпа до критични AI инструменти.

Развитието на големи езикови модели (LLM) и невронни мрежи изисква огромен изчислителен ресурс, който е трудно постижим с локални сървъри. Облачните инфраструктури предоставят достъп до хиляди процесори (GPU) за обучение на модели и дават възможност AI услугите да бъдат достъпни от всяка точка на света.

Съвременната дигитална среда преминава през трансформация на браузъра – от пасивна програма за четене на HTML уеб страници те се превръщат в активен AI асистент. Интегрират се системи: Microsoft внедрява Copilot в Edge, а Google – Gemini в Chrome. Нови играчи като Brave и Vivaldi залагат на сигурност и вградени локални AI асистенти. Проекти като ChatGPT Atlas, Perplexity Comet, Open Browser, Dia и Sigma не просто показват страници, а изпълняват сложни задачи: резервации, извличане на данни в таблици и дълбоки проучвания (Deep Research) чрез автономно управление на уеб интерфейси.

Този бум беше продиктуван от предхождащото го развитие на търсещите машини, които първоначално използваха прости механизми за съвпадение на ключови думи, за да се развият до съвременни системи, способни на дълбоко разбиране на съдържанието. Ключов момент в прехода на интернет инфраструктурата към интелигентни системи е въвеждането на алгоритъма PageRank, който бележи първия значим напредък в оценката на уеб съдържание.

Борбата с уеб спама продължава с въвеждането на Trust Rank – техника, която оценява качеството на страниците на база връзки от ръчно избрани доверени източници. Впоследствие, с развитието на алгоритмите на Google като Флорида, Панда, Пингвин и Колибри, се появяват все по-усъвършенствани средства за филтриране, разбиране на езика и разчитане на потребителските намерения – процеси, които все по-силно интегрират изкуствения интелект в сърцето на търсачките. Така се стига до RankBrain през 2015 г. – първата самообучаваща се AI система на Google, която бележи прехода към дълбоко, контекстуално и мултимодално разбиране на информацията (BERT, MUM). В резултат търсачките еволюират до интелигентни агенти, които не само откриват, но и синтезират знание. Така изкуственият интелект се превръща във водещ интерфейс за достъп до информация и основен двигател на дигиталната трансформация, което намира отражение във все по-широкото навлизане на AI в различни сфери, включително националната сигурност, отбраната и автоматизираното съдържание онлайн.

Развитието на изкуствения интелект оказва значително влияние върху сектора на националната сигурност и отбраната, като трансформира начина, по който се проектират и използват военни системи и технологии. Още през 40-те г. на XX в. радарните системи и противовъздушната отбрана започват да използват транспондери и автоматизирани методи за идентификация на самолети като приятели или врагове. С течение на времето тези технологии се развиват и системите за предупреждение започват да разпознават обекти чрез сравняване на радарни сигнали и топлинни сигнатури с бази данни за заплахи. През 70-те години се проектират ракети с автономно насочване и системи за управление на огъня с ограничена автономия (Arkin 2013).

Съвременните AI системи в отбраната обхващат: компютърно зрение за анализ на разузнавателни изображения и видео; автономни безпилотни системи (дронове, наземни и подводни апарати); системи за поддръжка на вземането на решения (Decision Support Systems); киберзащита и разузнаване; както и логистика и предиктивна поддръжка на военна техника (Horowitz et al. 2018).

Критичните въпроси на автономните оръжейни системи (LAWS – Lethal Autonomous Weapons Systems) и машинната етика във военния

контекст стават предмет на международни дебати в рамките на Конвенцията за определени конвенционални оръжия (CCW) на ООН[12].

Големите пробиви в AI през края на 90-те и началото на XXI век, като победата на IBM Deep Blue над световния шампион по шах и триумфът на AlphaGo срещу топ играча на Го, демонстрират потенциала на интелигентните системи да решават изключително сложни задачи с огромен брой възможни варианти. Съвременният изкуствен интелект, базиран на машинно обучение и невронни мрежи, води до значителен напредък в области като компютърно зрение, разпознаване на реч и обработка на естествен език, което позволява на военните системи да анализират видео данни, да идентифицират обекти и да взаимодействат с оператори чрез естествен език.

Еволюцията от дигиталната среда към AI инфраструктура представлява комплексен процес, обединяващ философско-математически концепции, технологични иновации и социални трансформации. Изкуственият интелект не възниква изолирано, а е резултат от съвместната еволюция на формалната логика, теорията на изчисленията, комуникационните мрежи и социалните технологии. Дигиталната среда преминава през последователни фази – от статичен Web 1.0 през социалния Web 2.0 до интелигентния Web 3.0, като на всеки етап се натрупват критични маси от данни, изчислителна мощ и алгоритмични иновации. Съвременната AI инфраструктура се характеризира с:

– масова достъпност на облачни изчислителни ресурси;

– наличие на огромни корпуси от обучителни данни;

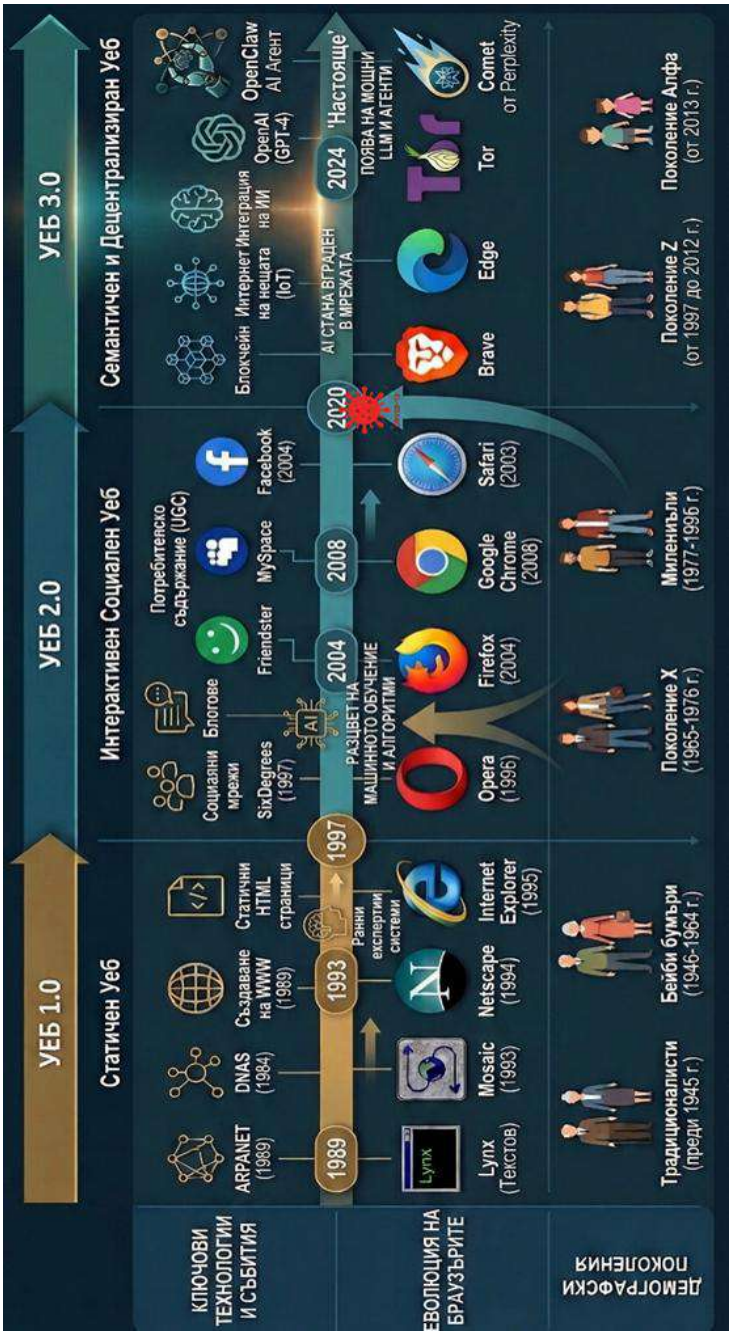
– развити алгоритми за дълбоко обучение;

– интеграция в крайни потребителски устройства и приложения;

– нарастващо присъствие в критични сектори, включително отбраната и националната сигурност.

Разбирането на тези инфраструктурни компоненти и тяхната историческа еволюция е задължителна предпоставка за ефективното проектиране, управление и регулиране на AI системи в съвременната дигитална среда.

[12.] UNITED NATIONS. Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems [онлайн]. Geneva: United Nations, 2019. Doc. CCW/GGE.1/2019/3. [Прегледан на 3 юни 2026]. Достъпен на: https://documents.unoda.org/wp-content/uploads/2020/09/CCW_GGE.1_2019_3_E.pdf.



Фигура 11. Развитие на AI технологиите спрямо различните поколения

НАУКА ЗА НЕВРОННИТЕ МРЕЖИ И МАШИННОТО ОБУЧЕНИЕ

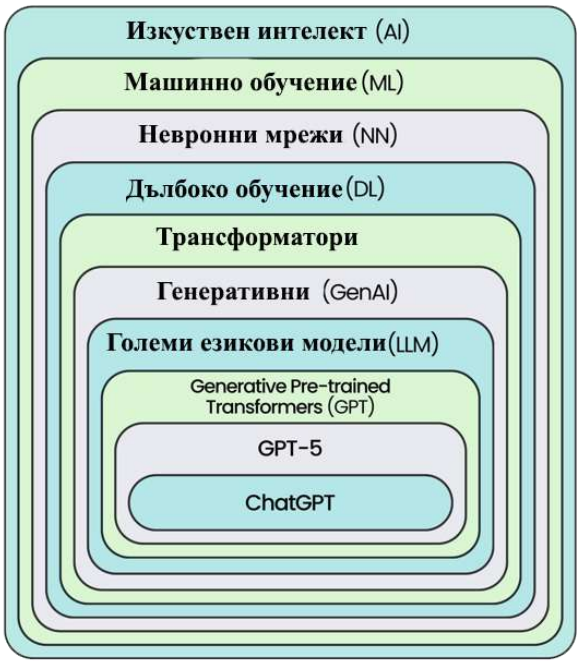
По-голямата част от дейността ни
с машинното обучение
се случва под повърхността.

Джеф Безос

3.1. От символен към статистически изкуствен интелект

Въпреки изключителния обществен и медиен шум, генериран от неотдавнашното обявяване на сложни генеративни AI модели, е необходимо да се подчертае, че тази технология не е внезапно откритие, а продукт на десетилетия последователно научно развитие (Dimitrov & Nikovski 1999). За да се разбере как работи съвременният изкуствен интелект, е необходимо първо да се анализира йерархията между изкуствения интелект, машинното обучение и дълбокото обучение, като се изяснят математическите принципи, позволяващи на машините да се „учат“ от данни.

В научната литература изкуственият интелект се дефинира като широкообхватна научна област и инженерна дисциплина, изследваща методологиите за създаване на машини и софтуерни агенти, способни да демонстрират интелигентно поведение (Dimitrov & Nikovski 1999). AI включва теории, алгоритми, методологии и стратегически цели, обхващащи области като машинно обучение, компютърно зрение, обработка на естествен език (NLP), роботика и автоматизирано планиране. В рамките на тази дисциплина съществува строго дефинирана йерархия – машинното обучение се разглежда като поддисциплина на изкуствения интелект, дълбокото обучение е специализиран дял от машинното обучение, използващ многослойни невронни мрежи, а генеративният AI представлява специфичен клас модели в рамките на дълбокото обучение.



Фигура 12. Вложена мрежа от технологии от ИИ до трансформаторите

В контекста на отбраната е от критично значение да се направи разграничение между научната област „Изкуствен интелект“ и нейната конкретна реализация – интелектната система. Докато AI е дисциплината, интелектната система е конкретното приложение или машина, която може да възприема своята среда, да разсъждава върху събраната информация и да действа автономно или полуавтономно, за да постигне определена военна или стратегическа цел. Основната разлика между интелектната система и традиционния софтуер се състои в това, че тя не следва просто статични, предварително кодирани инструкции. Вместо това тя демонстрира динамично поведение, включващо учене, адаптация към променящите се условия на бойното поле и способност за разрешаване на неструктурирани проблеми.

Архитектурата на една интелигентна система се определя от синергията между нейните компоненти, превръщащи суровите данни в оперативно превъзходство. Основните елементи са:

1. Сензорна периферия (P – Perception) – осъществява връзката с физическия свят. Във военен контекст това са радари, камери и средства за

радиоелектронно разузнаване. Те дефинират вектора на наблюдение и осигуряват потока от данни в реално време.

2. База знания/база данни (K) – съхранява фактическата информация и научените модели. Тя служи като цифрова когнитивна памет на системата.

3. Механизъм за извод и разсъждения, като логически мотор (L) – изпълнява процеса на разсъждение, използвайки базата знания и текущите данни от сензорите. Използва логически правила или вероятностни модели (например байесови мрежи) за оценка на ситуацията.

4. Модул за обучение и адаптация (E) – сърцето на машинното обучение, което позволява на системата да подобрява своите алгоритми на базата на натрупания опит и обратната връзка от околната среда.

5. Изпълнителни механизми – актьори (A) или устройства, чрез които системата взаимодейства със средата – от управление на полета на дрон до задействане на системи за киберзащита.

6. Потребителски интерфейс (U) – осигурява комуникацията между човешкия оператор и AI системата, позволявайки „човек в контура“ (human-in-the-loop) при вземането на критични решения. Много често това е комуникационен модул, приемащ информация чрез разпознаване на реч и разбиране на текст или изображения.

За да представим взаимодействието между тези елементи като единна функция на системната архитектура S, можем да използваме следната формула:

$$S = E \,.\, L(P, K) \otimes (A + U),$$

където $L(P, K)$ е логическата функция на разсъждение, която съпоставя входящите данни от перцепцията с базата знания;

E – коефициент на обучение и адаптивност, който мащабира ефективността на логическите изводи;

$\otimes$ – оператор за координация между автоматизираното действие и човешкия контрол;

$(A + U)$ – синергията между автономното изпълнение и намесата на оператора.

Оптимизацията на тези шест компонента води до повишаване на ситуационната осведоменост. В борбата с дезинформацията, например, сензорите са алгоритми за обработка на естествен език (NLP), които идентифицират семантични аномалии в информационното пространство, а базата знания съдържа патерни на известни хибридни заплахи. В контекста на националната

сигурност и военното дело, тази структура осигурява затварянето на **НОРД цикъла** (Наблюдение, Ориентация, Разсъждение, Действие) и веригата на унищожаване (Kill Chain) с висока степен на автоматизация.

3.1.1. Ролята на формалната логика за представяне на знанието

Представянето на знания е концептуалният мост, който свързва суровите данни с интелигентното разсъждение. Още от древността човечеството се опитва да формализира процесите на логическо мислене, но в контекста на изкуствения интелект се оформят два основни методологични подхода – символни методи и методи, базирани на статистически данни (Goodfellow, Bengio & Courville, 2016).

Системите с изкуствен интелект се класифицират в два основни вида, които определят начина на представяне на знания и вземане на решения:

1. Класически или символен AI е доминираща концепция в научните изследвания от 50-те до 80-те години на XX век. Системите тук са базирани на формална логика и правила. Интелектът се симулира чрез манипулация на символи, които представят обекти и понятия от реалния свят. Тези системи са детерминистични и лесно интерпретируеми, но страдат от трудности при обработката на несигурност и шум в данните (Dimitrov & Nikovski 1999).

2. Модерен или статистически AI са група системи, основани на архитектури, които се обучават директно от данни, като централно място заемат изкуствените невронни мрежи. Те набраха скорост от 90-те години на XX в. насам и станаха доминиращи през последното десетилетие благодарение на възхода на „дълбокото обучение“. Статистическият AI е изключително мощен за откриване на сложни, скрити модели и аномалии в огромни обеми от неструктурирани данни, което го прави незаменим за съвременните отбранителни приложения, включително разузнаване, наблюдение и целеуказване.

Символните методи представят знанието експлицитно – ясно записано, като символи, логически твърдения и правила. Машината манипулира тези символи според строги правила, за да симулира интелигентност. Този подход е в основата на експертните системи, логическото програмиране и автоматичното планиране (Dimitrov & Nikovski 1999). Типични структури в тази категория са:

– логическото представяне на информация може да бъде ясно, логично и недвусмислено от типа „Том е една котка“;

– факти и правила — основата на продукционните системи („Ако А, тогава Б“);

– семантични мрежи – знанието се представя визуално като граф, където възлите са обекти или концепции, а ребрата – релациите между тях. Семантичното представяне е за взаимовръзките между фактите: Том – котка – бозайник;

– фреймове – информацията се представя в елементарни таблици, съдържащи „слотове“ за специфична информация. Например: име – Том, вид – котка, група – бозайници;

– онтологии — формални и споделени спецификации на дадена област, дефиниращи йерархията от класове, свойства и ограничения.

Методите, базирани на данни и статистика, съхраняват знанието имплицитно. В този случай то не се кодира ръчно от експерти, а AI системата го извлича чрез анализ на огромни обеми от данни, съхранявайки го като числени параметри – тегла в невронни мрежи или вектори на вграждания (Goodfellow, Bengio & Courville, 2016).

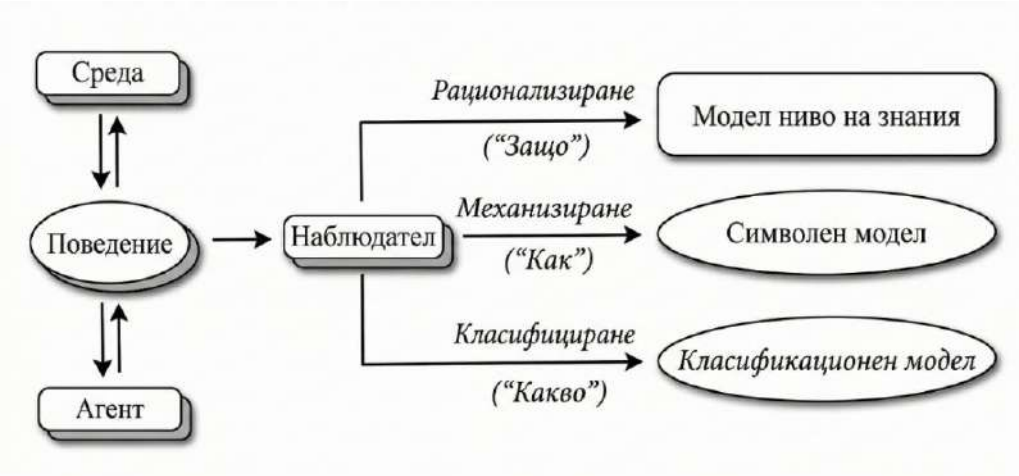
Съвременните информационни системи в Индустрия 4.0 се управляват виртуално от **цифрови близнаци** на реални физически обекти чрез постоянен поток от данни от сензори, което често съхранява и историческия профил на обекта. Цифровият близнак дава възможност на оператора да извлече полезни данни за функционирането му и да изпробва различни сценарии. Основните предимства включват намаляване на разходите за поддръжка и престой на устройствата, оптимизиране на времето на процеса и съкращаване на времето за навлизане на нови продукти на пазара. Същевременно те имат относително висока цена на разработка, тъй като изискват фундаментални AI модели на базата на законите на физиката, химията, икономиката и т.н.

Разработването на сложни цифрови близнаци е изключително трудно и скъпо, а внедряването им в реално време изисква значителна изчислителна мощ. Алтернативното решение са **емпиричните емулатори** – модели, основани и задвижвани от данни, които частично заменят фундаменталните принципи. Докато цифровият двойник използва сложни диференциални уравнения и изисква висококвалифицирани специалисти, емпиричните емулатори се базират на статистика и методи на изкуствения интелект (невронни мрежи, генетично програмиране) за оптимизация, класификация и предсказване.

Целта и в двата случая е максимално точно и евтино да се моделира физическото устройство. Така се създава интелектен агент – система, притежаваща способност за:

– извличане, представяне, натрупване, коригиране, интерпретиране и обобщаване на знания чрез активно взаимодействие със средата;

– използване на знания за водене на целенасочено поведение.



Фигура 13. Модели на интелигентно поведение

Чрез наблюдаване на екземпляри от поведението на агента наблюдателят създава класификационен „какво“ модел. Символното ниво осигурява средства за „механизиране“ на наблюдаваното поведение. Нивото на знания добавя поглед на целеустременост в поведението, изграждайки модел в термините на знания за средата, целите и потенциалните средства (действия) на агента.

За интелигентните системи е характерен преходът от схемата: данни + алгоритъм към схемата: знания + разсъждения. Данните са относително динамичната част от информацията – конкретни сведения за обектите в даден момент. Знанията са по-общата, относително постоянна и по-неизменяема част, валидни за цели класове обекти.

Знанието най-общо означава това, което човек е научил и разбрал за света около себе си. То представлява подредена система от твърдения, основани на реални факти. Ако погледнем знанието от лична гледна точка, то може да изглежда като мнение или убеждение, което човек има, защото

вярва, че е истина, като се базира на това, което е наблюдавал. В изкуствения интелект знанието се разглежда по-тясно – тоест, интересува ни как знанието може да помогне за решаването на конкретни задачи в дадена област. Когато говорим за знание в ИИ, имаме предвид събраната и подредена информация с факти, която помага на системата да взема решения или да решава проблеми.



Фигура 14. Размер на знанието

Предметната област на изследване е всяка област от живота, която изучаваме и ни дава информация. Например, всяко понятие – като „котка“, „дърво“ или „училище“ – определя група от обекти, споделящи общи черти. Всички неща, попадащи в тази група, образуват нейния обем. Така например, категорията „животно“ включва всички видове животни, а „котка“ – само котките. Съдържанието на едно понятие показва какви важни връзки и свойства имат обектите в тази група. Когато определим тези връзки с думи или символи, можем да проверим дали са верни или не. Такава проверена информация се счита за знание в изкуствения интелект.

Информацията е интересна, когато има смисъл и може да се използва за решаване на проблеми. Например, ако знаем формулата за лицето на квадрат, това ни помага да изчислим площта на всякакъв квадрат – това е знание. Ако знаем само дължината на една страна на конкретен квадрат, това са данни. Знанието е по-общо и по-трайно, докато данните са по-конкретни и временни. Стойностите на признаците, които получаваме от наблюдения или измервания (например дължина, температура, цвят), се наричат данни. Те описват обектите в даден момент. Знанието, от друга страна, показва по-общи закономерности, които важат за повече обекти или случаи.

Предположение е всяко изказване, което може да бъде записано в езика за представяне и подлежащо на обосновка. Хипотезата е обосновано предположение (за което не е доказано че е истинно, но в достатъчна степен се смята, че може да е такова). Знанието е обосновано истинно предположение доказано с факти или правила. Фактите носят истинска информация, а правилата – при определени условия. Базата знания е съвкупност от истинни предположения, записани в езика на представяне.



Фигура 15. Ролята на контекста в представянето на знания

Знанието може да се представя и използва по различни начини. Например, има декларативни знания – това са факти и правила, които са записани и могат да се проверят независимо от начина, по който ще се използват

– „Всеки квадрат има четири страни“. Ако знаем също, че „фигура А е квадрат“, можем да заключим, че „фигура А има четири страни“.

Процедурните знания, от друга страна, са инструкции как да направим нещо – например, стъпките за намиране на площта на квадрат. Те са като рецепта, която трябва да следваме, за да постигнем определен резултат. Тези знания са нужни, за да изпълним задача правилно. Основни модели за представяне на знания в изкуствения интелект са: логически модели, които боравят с понятията „истина“ и „лъжа“, но се използват и мрежови, фреймови и продукционни модели.

Механизмите за разсъждение могат да бъдат:

– дедуктивен извод – от общи правила и конкретни факти се извеждат конкретни заключения. Пример: ако е известно, че „всяко кубче има отвор“ и че „обектът О е кубче“, то може да се заключи, че „обектът О има отвор“;

– индуктивен извод – от конкретни примери се извеждат общи правила;

– абдуктивен извод – формиране на най-доброто обяснение за наблюдавани факти.

3.1.2. Съждително, предикатно смятане и размита логика

Формалната логика е двигателят на всяка интелигентна система. В науката за AI се разграничават три основни нива на логическа сложност:

1. Съждително смятане – най-простият формален език, работещ с прости твърдения, които могат да бъдат само истинни (1) или неистинни (0). Използва логически оператори като И (AND), ИЛИ (OR) и НЕ (NOT). Основното му ограничение е невъзможността да работи с променливи или да изразява вътрешната структура на обектите.

2. Предикатно смятане – разширява съждителното смятане чрез въвеждането на обекти, предикати, променливи и квантори. Това позволява моделиране на сложни знания за света, като йерархии и отношения между обекти (напр. „Всички дронове от клас X имат сензор Y“). Тази логика е в основата на езика Prolog и експертните системи за планиране (Dimitrov & Nikovski 1999).

3. Размита логика – концепция, предложена от Лотфи Заде през 1965 г., която адресира несигурността и неточността на реалния свят (Zadeh 1965). Вместо бинарните стойности 0 и 1, размитата логика използва степени на истинност. Това е критично за отбранителните системи, работещи с понятия без строги граници като „висока температура“, „бърза скорост" или „значителен риск“.

Съждителното смятане е формална система, която превръща човешките разсъждения в математически модели. В неговия център стои елементарното съждение – твърдение, изречение на естествен език, което притежава точно една от две възможни стойности: истина (1) или лъжа (0). То не може да бъде и двете едновременно, нито пък въпрос или заповед.

– Пример за съждение А: „Слънцето е звезда.“ (Това е истина).

– Пример за съждение В: „2 + 2 = 5.“ (Това е лъжа).

– НЕ е съждение: „Колко е часът?“ или „Затвори вратата!“.

От формална гледна точка елементарно съждение е двойката <знак, стойност>. Сложните мисли се изграждат чрез свързване на прости съждения (променливи като A, B, P, Q, R) с помощта на логически връзки. От елементарните съждения могат да се образуват съставни чрез преобразуване посредством частицата „не“, както и чрез съчетаване с помощта на съюзите (наричани съждителни връзки) „и“, „или“, „ако... то...“, „тогава и само тогава“, „ако и само ако“. Истинностната стойност на всяко съставно съждение се определя в зависимост от истинностните стойности на участващите в него елементарни съждения. Ето как те определят истинността на цялото изречение:

Таблица 1. Операции и символи на елитарни съждения

Операция	*Символ*	*Смисъл*	*Кога е Истина?*
Отрицание	¬	„не“	Когато оригиналът е лъжа
Конюнкция	∧	„и“	Само ако всички части са истина
Дизюнкция	∨	„или“	Ако поне една част е истина
Импликация	→	„ако... то...“	Винаги, освен при комбинацията (истина → лъжа)
Еквивалентност	↔	„само ако“	Когато и двете страни са еднакви (И – И или Л – Л)

Ако имаме две прости съждения: А: „Навън вали“ и В: „Аз нося чадър.“, можем да ги свързваме, точно както в езика използваме съюзи. Тези връзки се наричат логически оператори. Най-лесният оператор е отрицанието не (¬). Той просто обръща стойността на съждението. Например, ако А е „Истина“, то ¬ А е „Лъжа“.

Еквивалентни формули се обозначават с ∧ за конюнкция (логическо „и“), ∨ за дизюнкция (логическо „или“), ¬ за отрицание, → за импликация, ↔ за еквивалентност. Основни логически закони са:

– идемпотентност: x ∨ x = x; x ∧ x = x;

– комутативност: x ∨ y = y ∨ x; x ∧ y = y ∧ x;

– асоциативност: (x ∨ y) ∨ z = x ∨ (y ∨ z); (x ∧ y) ∧ z = x ∧ (y ∧ z);

– дистрибутивност: x ∧ (y ∨ z) = (x ∧ y) ∨ (x ∧ z); x ∨ (y ∧ z) = (x ∨ y) ∧ (x ∨ z);

– Закони на Де Морган: ¬(x ∧ y) = ¬x ∨ ¬y; ¬(x ∨ y) = ¬x ∧ ¬y;

– двойно отрицание: ¬¬x = x;

– Закон за изключеното трето: x ∨ ¬x = истина (тавтология);

– Закон за противоречието: x ∧ ¬x = лъжа (контрадикция).

За да разберем едно сложно съждение, обикновено се използва истинна таблица, която показва всички възможни комбинации. Например, ако вземем две прости съждения: P: „Слънцето грее.“ Q: „Отивам на плаж.“, а ние разглеждаме формулата: P ∧ ¬Q, която се чете: „Слънцето грее И не отивам на плаж“, ето как изглежда таблицата за истинност в тази ситуация:

Таблица 2. Таблица на истинността

P (Слънце)	*Q (Плаж)*	*¬Q (Не на плаж)*	*P∧¬Q (Резултат)*
И (Истина)	И (Истина)	Л (Лъжа)	Л (Лъжа)
И (Истина)	Л (Лъжа)	И (Истина)	И (Истина)
Л (Лъжа)	И (Истина)	Л (Лъжа)	Л (Лъжа)
Л (Лъжа)	Л (Лъжа)	И (Истина)	Л (Лъжа)

Обърнете внимание на втория ред, където резултатът е **Истина**. Това е единственият сценарий, при който цялото твърдение е вярно: когато слънцето наистина грее, но човека по някаква причина **не** е на плажа.

Въпрос за размисъл. Ако променим формулата на **P → Q** („Ако слънцето грее, то отивам на плаж“), при коя комбинация от таблицата твърдението ще бъде **Лъжа**? (Погледнете частта за импликация, ако имате нужда от жокер).

Предикатното смятане е фундаментален раздел от математическата логика, въведен от германския математик Готлоб Фреге. То играе ключова

роля в развитието на изкуствения интелект, тъй като предоставя мощен формален апарат за описание и анализ на знания в различни предметни области (Dimitrov & Nikovski 1999). Предикатното смятане е следващото ниво в логиката след съждителното смятане. Ако съждителното смятане разглежда целите изречения като „черни кутии“ (истина или лъжа), предикатното смятане „влиза“ вътре в изречението, за да разбере връзките между обектите и техните свойства.

Ако в съждителното смятане казваме: P: „Иван е човек“ и Q: „Мария е човек“, то в предикатното смятане използваме предикати – това са функции, които описват свойство или връзка. Нека H(x) означава „x е човек“. Тогава: H(Иван) е истина, H(Мария) е истина. Тук H е предикатът (свойството), а Иван е обектът.

Предикатите могат да се разглеждат като съждения, които съдържат променливи. Всяка променлива може да приема стойности от определена предметна област, а истинността на съждението зависи от конкретните стойности, които тези променливи получават. Така предикатите позволяват да се формулират общи твърдения за обекти и техните свойства или отношения.

Също така се използват квантори – за „всички“ и за „някои“. Именно това е „суперсилата“ на предикатното смятане. То ни позволява да говорим за групи от обекти, използвайки два специални символа – универсален квантор (∀), който означава „за всеки“ или „всички“ и екзистенциален квантор (∃), който означава „съществува поне един“.

– ∀ x H(x) означава „Всички обекти са хора“;

– ∃ x H(x) означава „Съществува поне един обект, който е човек“.

Синтаксисът на предикатното смятане се основава на ясно определен набор от символи и правила за построяване на валидни изрази. Азбуката на предикатното смятане включва:

– константи – имена на конкретни обекти (обикновено малки латински букви или цифри);

– променливи – означават произволни обекти (главни латински букви);

– функции – имена на функционални зависимости с определена арност (брой аргументи);

– предикати – имена на релационни свойства с определена арност;

– логически символи – за съждителни връзки (∧, ∨, ¬, →, ↔), скоби;

– квантори – ∀ (квантор за общност, „за всяко“) и ∃ (квантор за съществуване, „съществува поне едно“).

Например, имаме група от плодове: ябълка, банан и круша. Нека предикатът F(x) означава „x е жълт“. Ако знаем, че: ябълката е червена; бананът е жълт; крушата е зелена, кое от двете твърдения е вярно за тази група плодове $\forall$ x F(x) (Всички плодове са жълти) или $\exists$ x F(x) (Съществува поне един жълт плод). Кой е правилният отговор? Тъй като само бананът е жълт втората опция $\exists$ x F(x) е истина, а първото твърдение е лъжа, защото не всички плодове в нашата група са жълти.

Докато съждителното смятане описва само свойства (като „Иван е човек“), предикатното може да описва отношения между двама или повече души. Например: L(x, y) означава „x обича y“. С това можем да изразим много различни неща, в зависимост от това как подредим кванторите:

– $\forall x \forall y L(x,y)$ – всеки обича всекиго;

– $\exists x \exists y L(x,y)$ – някой обича някого.

– $\forall x \exists y L(x,y)$ – всеки обича някого (всеки си има някой, когото обича).

Забележете как редът и видът на кванторите променят смисъла на изречението.

За упражнение, нека дефинираме следното: S(x) означава „x е студент“. R(x) означава „x чете книги“. Как би изглеждала на обикновен език формулата ∃x (S(x)∧R(x).

Отговорът е „Съществува поне един студент, които чете книги.“ Кванторът за съществуване ($\exists$ x), който ни казва, че има „поне един“ такъв обект. Конюнкцията ($\wedge$), която свързва двете свойства – да бъдеш студент и да четеш книги.

Предикатното смятане ни позволява да бъдем много прецизни и дори да използваме конкретни термове – основните градивни елементи на изразите. Това са имената, с които назоваваме обектите. Всеки термин представлява обект от предметната област и може да бъде константа (конкретни имена – Иван, Мария), променлива (x, y, z, които могат да бъдат всеки обект) или резултат от прилагане на функция върху други термове. Например, ако имаме функцията баща(x), то баща (Иван) е терм, който сочи към конкретен човек (бащата на Иван).

– Човек (Иван) – предикат с един терм (свойство);

– Брат (Иван, Петър) – предикат с два терма (връзка).

Кванторите позволяват да се изразяват обобщени твърдения. Формулите от вида (∀v α) и (∃v α) се наричат квантифицирани, като обхватът

(областта на действие) на квантора е формулата, към която е приложен. Променливата се нарича „свързана“ във формулата, ако се намира в обхвата на квантор, в противен случай е „свободна“. Формула без свободни променливи се нарича „затворена“.

– свободна – в израза „Обича(x, сладолед)“ променливата x е свободна. Не знаем кой е x, той е просто „празно място“;

– свързана: В израза $\forall$ x Обича(x, сладолед), променливата x вече е „свързана“ от квантора $\forall$ (всеки). Сега вече знаем за кого говорим – за всички.

Светът, в който работи предикатното смятане, се нарича „предметна област“. Тя включва всички обекти, техните свойства и връзките помежду им, които са важни за решаването на проблема. Ако решаваме логическа задача за шах, предметната област обхваща шахматната дъска, нейните 64 квадрата и всички фигури. В този контекст фигурите и квадратите са нашите обекти. Всичко извън тази област (например времето навън или цветът на масата) е несъществено и се изключва от логическия анализ.

Обект е всяко отделно нещо, което можем да различим. Това може да е реален предмет (като фигура на кон) или абстрактна идея (като числото 5). За да бъде нещо обект, то трябва да е ясно дефинирано и различно от останалите.

Предметът е обект, върху който сме фокусирали вниманието си в конкретния момент. Когато анализираме свойствата на един обект (например „Конят е бял“), той се превръща в наш предмет на изследване. Предметът е обект, който е подложен на изследване, анализ или наблюдение, като се разглежда с оглед на определени свойства и взаимоотношения. Предметът има свойства и връзки, но при абстракция на реалността, при която се игнорират несъществени характеристики. Така „предметна област“ фиксира гледната точка, от която се извършва анализът, и определя кои свойства и връзки между обектите са значими за конкретната задача.

Както отбелязва Нилс Нилсон: „Всеки език за представяне на знания в интелектните системи трябва да бъде най-малкото толкова изразителен, колкото предикатното смятане от първи ред“ (Nilsson 1980). Това подчертава значението на предикатното смятане като минимален стандарт за формализиране на знанието.

Размитата логика е класическа логикаq в която всичко е черно или бяло (0 или 1). Например: „Един човек е или висок, или не е“. Но в реалния живот, ако някой е висок 179 см, той висок ли е? А този, който е 181 см?

Размитата логика казва: „Този човек е 80% висок“. Тя работи със степени на истина (числа между 0 и 1).

Ако трябва да изградим интелигентна система за управление на дрон, която трябва да реши колко силно да надуе двигателите според вятъра, размитатата логика постановява изграждането в три стъпки:

1. **Фазификация** (Размиване) – дронът измерва скорост на вятъра точно 45 км/ч. Системата превръща това число в понятия – „Вятърът е 30% СИЛЕН и 70% УМЕРЕН“.

2. **Оценка на правилата – с**истемата гледа своята база от знания: „АКО вятърът е СИЛЕН, ТОГАВА увеличи мощността ЗНАЧИТЕЛНО“. Тук се пресмята как точно се комбинират тези „размити“ условия.

3. **Дефазификация** (Изчистване) – накрая дронът не може да изпълни команда „малко мощност“. Той се нуждае от число. Дефазификацията превръща размития резултат в конкретна стойност, например „Подай 12,5 волта към мотора“.

3.1.3. Семантични мрежи и онтологии: структуриране и представяне на знания

Докато логическите изчисления (съждително, предикатно смятане) ни дават правилата за извеждане на заключения, семантичните мрежи предоставят графичния модел за това как различните обекти, концепции и техните релации се съхраняват в паметта на една интелигентна система. Семантичната мрежа е модел за представяне на знания, който използва структурата на насочен граф:

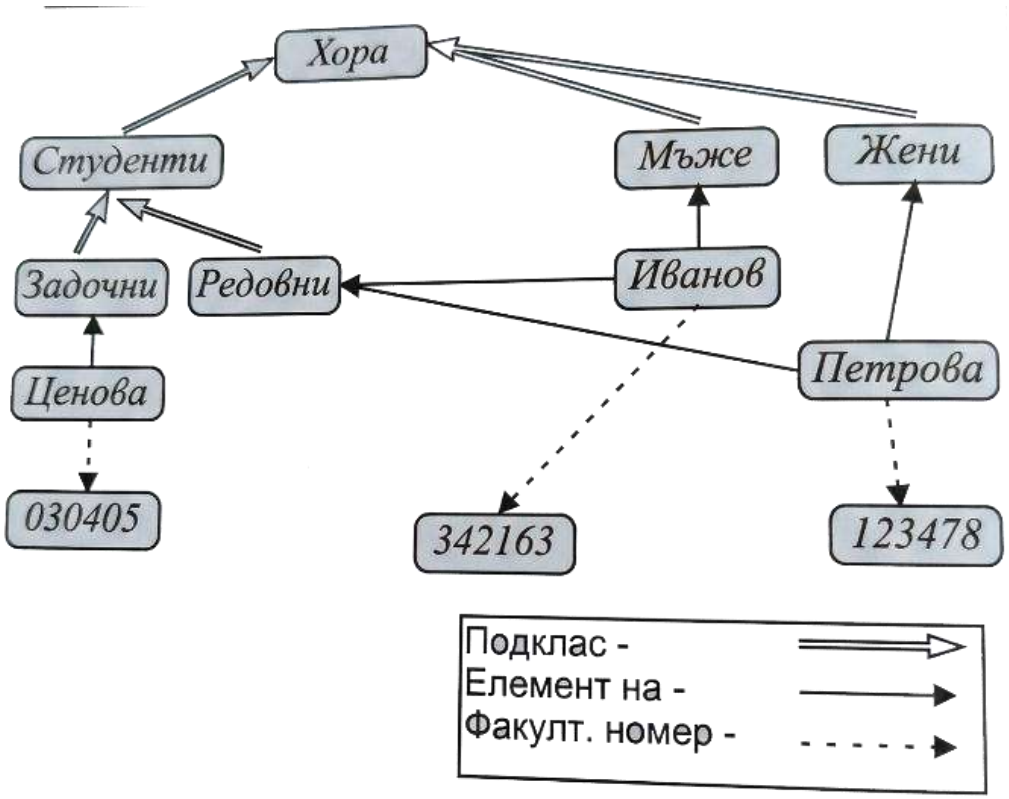
– Възлите (Nodes) представляват конкретни военни обекти (екземпляри) или абстрактни концепции (класове). В една военна база данни възли могат да бъдат „Т-72“, „1-ва бронетанкова бригада“ или „дизелово гориво“.

– Дъгите/ръбовете представляват релациите между тях (напр. „Е вид“, „Притежава“, „Намира се в“, „Е съюзник на“).

Едно от най-мощните свойства на семантичните мрежи е автоматичното извеждане на нови знания чрез транзитивни релации. Ако знаем, че „Т-72“ е вид „Основен боен танк“ и „Основният боен танк“ има „броня“, системата автоматично извежда заключението, че „Т-72“ има броня, дори това да не е изрично записано за конкретния модел.

В съвременните военни доктрини (като мрежовоцентрични войни), семантичните мрежи са критични за ситуационната осведоменост и създаване на единна картина на бойното поле, където сензори, платформи и

човешки ресурси са свързани чрез семантични релации (кой обект е заплаха, кой е неутрален и каква е връзката между тях). Семантичните мрежи могат да се използват при управление на веригата за доставки и автоматизирано проследяване на ресурси, например – „Склад А“ съдържа „Резервни части“, които са необходими за „Вертолет Б“. Също така се употребяват в платформите за подпомагане вземането на решения (C2 – Command and Control), които използват онтологии (разширени семантични мрежи), за да разпознаят сложни тактически модели.



Фигура 16. Семантична мрежа

В дигиталните умения за отбрана често се преминава от прости мрежи към военни онтологии. Въпреки че често се използват взаимнозаменяемо в по-общ контекст, в инженерството на знанията съществува ясна

разграничителна линия между двете концепции. Докато семантичната мрежа се фокусира върху структурата на данните и асоциативните връзки, то онтологията е формално описание на знанието в конкретна област (напр. „Градски бой“ или „Киберсигурност“). Обикновено тя включва:

– Таксономия – йерархия на обектите (напр. противовъздушна отбрана (ПВО) -> С-400).

– Аксиоми – правила, които винаги са верни (напр. „Никоя единица не може да бъде едновременно в движение и в ремонт“).

Таблица 3. Семантична мрежа и онтология

Характеристика	*Семантична мрежа*	*Онтология*
Основна цел	Представяне на релации между обекти	Дефиниране на структура и логически правила
Ниво на формализъм	По-ниско (асоциативно)	Високо (математическа логика)
Компоненти	Възли и дъги	Класове, свойства, аксиоми, инстанции
Приложение	Търсене на пътища, асоциации	Разсъждаване (Reasoning), оперативна съвместимост

Да разгледаме следния пример за радарното засичане:

1. Сензорно ниво – радарът подава сурови данни: обект *O1*, скорост *v=2.5 “Mach”*.

2. Онтологично ниво – системата проверява дефинираните категории. В онтологията за „Въздушни заплахи“ съществува правило:

“If” v> “Mach”2.0⟹“Category: Ballistic/High-Speed Threat”.

3. Семантично ниво – в мрежата се активира възелът „Заплаха“. Чрез релацията „Намира се в“ системата веднага идентифицира кои приятелски единици са в обсега на траекторията на *O1*.

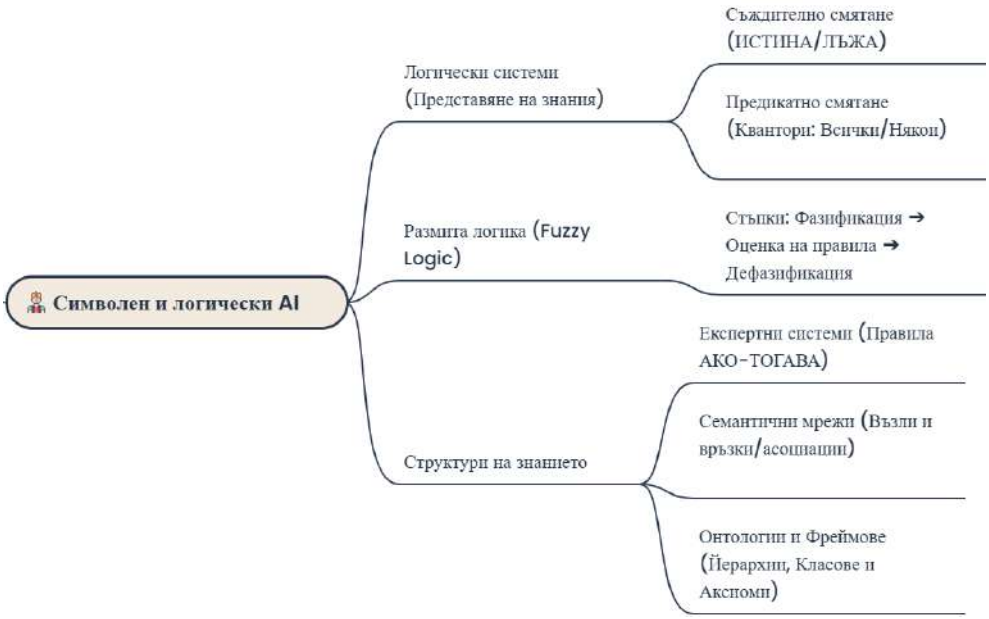
Този подход е в основата на борбата с дезинформацията и хибридните заплахи, тъй като позволява на алгоритмите да откриват логически противоречия в постъпващата информация чрез съпоставяне с установени онтологични аксиоми. Предимствата на тези мрежи са, че човекът оператор може лесно да разбере защо машината е взела дадено решение и позволява бързо търсене в сложни структури от данни. Недостатъците са, че при разрастване на мрежата става трудно да се избягват логически противоречия и трудно се

справят с несигурност или вероятностни събития (тук се налага намесата на размитата логика или статистиката).

Въпреки своята прецизност, семантичните мрежи изискват ръчно дефиниране на всяка връзка от експерти. В динамичната среда на модерната отбрана, където данните са огромни и неструктурирани, този „ръчен“ подход става недостатъчен. Това налага прехода към статистически модели и машинно обучение, където машината сама открива зависимостите в данните, вместо да разчита на предварително начертани карти на знанието.

3.1.4. Преходът към статистически модели и анализ на големи масиви от данни

Преходът от символен към статистически AI е обусловен от няколко фундаментални ограничения на класическите подходи. Символните системи изискват ръчно кодиране на знанията от експерти, което е трудоемък и скъп процес, наречен „бутилково гърло на знанието“. Те са чувствителни към шум в данните и не могат ефективно да се справят с неструктурирана информация като изображения, звук и естествен език.



Фигура 17. Класификация на алгоритмите за символен и логически AI

Статистиката и машинното обучение предлагат алтернатива – вместо да се програмират правила, системите се обучават от примери. С нарастването на изчислителната мощ и наличността на големи масиви от данни (Big Data) статистическите методи демонстрират превъзходна производителност в задачи като разпознаване на образи, обработка на естествен език и предсказателен анализ. Това доведе до доминирането на подходите, базирани на данни, в съвременния AI (Goodfellow, Bengio & Courville, 2016).

3.2. Класически методи на машинно обучение

Машинното обучение (Machine Learning) представлява междудисциплинарна област на изследване, която се занимава с разработването на алгоритми, позволяващи на компютърните системи да подобряват производителността си при изпълнение на специфични задачи чрез натрупване на опит от данни, без да бъдат изрично програмирани за всяка конкретна ситуация (Samuel 1959; Mitchell 1997). Според формалната дефиниция на Том Мичъл, компютърна програма се счита, че учи от опит E по отношение на клас задачи T и мярка за производителност P, ако нейната производителност при задачите в T, измерена с P, се подобрява с опита E (Mitchell 1997).

Основната парадигма на машинното обучение изисква наличието на три фундаментални компонента (Mitchell 1997; Goodfellow, Bengio & Courville, 2016):

– модел – хипотеза или структура, която се обучава;

– данни – обучителни примери, представляващи опита E;

– алгоритъм за обучение – оптимизационен метод за напасване на параметрите на модела към данните.

В зависимост от начина на подаване на обратна връзка по време на тренировъчния процес, машинното обучение се систематизира в три основни вида:

● **Контролирано обучение** (Supervised Learning) – подход, при който алгоритъмът се обучава върху множество от етикетирани примери, състоящи се от входни данни (признаци) и съответните желани изходи (целеви стойности). Целта е изграждане на модел, способен да прогнозира верния изход за нови, непознати данни чрез минимизиране на функцията на грешката между прогнозираните и реалните стойности (Bishop 2006; Alpaydin 2010). Този подход намира основно приложение в задачите за класификация (присвояване на дискретни категории) и регресия (предсказване

на непрекъснати числови стойности). В контекста на отбраната и сигурността, контролираното обучение се използва за разпознаване на вражеска техника в сателитни снимки чрез свръхточно класифициране на визуални модели (Dimitrov & Nikovski, 1999).

● **Неконтролирано обучение** (Unsupervised Learning) – функционира при липса на етикетирани данни, като алгоритмите се стремят да открият скрити структури, клъстери (групи от сходни обекти) или разпределения в данните без предварителна информация за целевите категории (Alpaydin 2010). Методите на неконтролираното обучение са фундаментални за откриване на аномално поведение в мрежовия трафик (киберсигурност), идентифициране на нови видове заплахи, за които липсват предварително дефинирани правила, както и за сегментация на клиенти и намаляване на размерността на данните (Bishop 2006).

● **Обучение с подсилване** (Reinforcement Learning) – изчислителен подход, при който агент се учи да взема решения при спазване на последователност чрез взаимодействие с динамична среда, получавайки числени сигнали за награда или наказание като обратна връзка (Sutton & Barto 2018). За разлика от контролираното обучение, където се предоставят готови примери с правилни отговори, агентът сам открива оптималната стратегия (политика) чрез метода на пробата и грешката, целейки максимизиране на кумулативната награда в дългосрочен план. Фундаменталните елементи включват агент (вземащ решения), среда (реагираща на действия), състояние (текуща ситуация), действие (изборът на агента), награда (обратна връзка от средата), политика (стратегията на поведение) и функция на стойността (оценка на дългосрочната полезност на състоянията) (Sutton & Barto 2018).

Обучението с подсилване може да се класифицира в три основни подкатегории според наличието на модел на средата:

– Model-Free RL – агентът не изгражда модел на средата, а директно учи оптимална политика чрез Q-learning, Deep Q-Networks или Policy Gradient методи. В рамките на управлението на политиките се използват два алгоритъма – PPO и GRPO, които сравняват резултатите и му дават награди за развитие на най-добрия от тях;

– Model-Based RL — агентът изгражда модел на преходите между състоянията и използва планиране, както при Dyna-Q и MuZero;

– Inverse Reinforcement Learning – агентът извлича функцията на наградата от поведението на експерти (Sutton & Barto 2018)

През периода 2024 – 2025 г. областта на обучението с подсилване претърпя революционни скокове в способностите, захранвайки модели като

OpenAI o1, o3 и DeepSeek R1. Тези модели използват разширени вериги на разсъждение и търсене по време на извод, при което генеративният модел извършва вътрешно разсъждение за оптимизиране на стратегията преди финалния отговор. Моделът o3 постига бенчмарк 87,7% експертно ниво в науката според GPQA Diamond, 71,7% умения за софтуерно инженерство в SWE-bench Verified и Elo рейтинг 2727 на Codeforces, докато DeepSeek R1 демонстрира способност за самостоятелно развитие на поведения като рефлексия и самокорекция чрез чисто подсилващо обучение с оптимизация с относителни правила за групата (DeepSeek 2025).

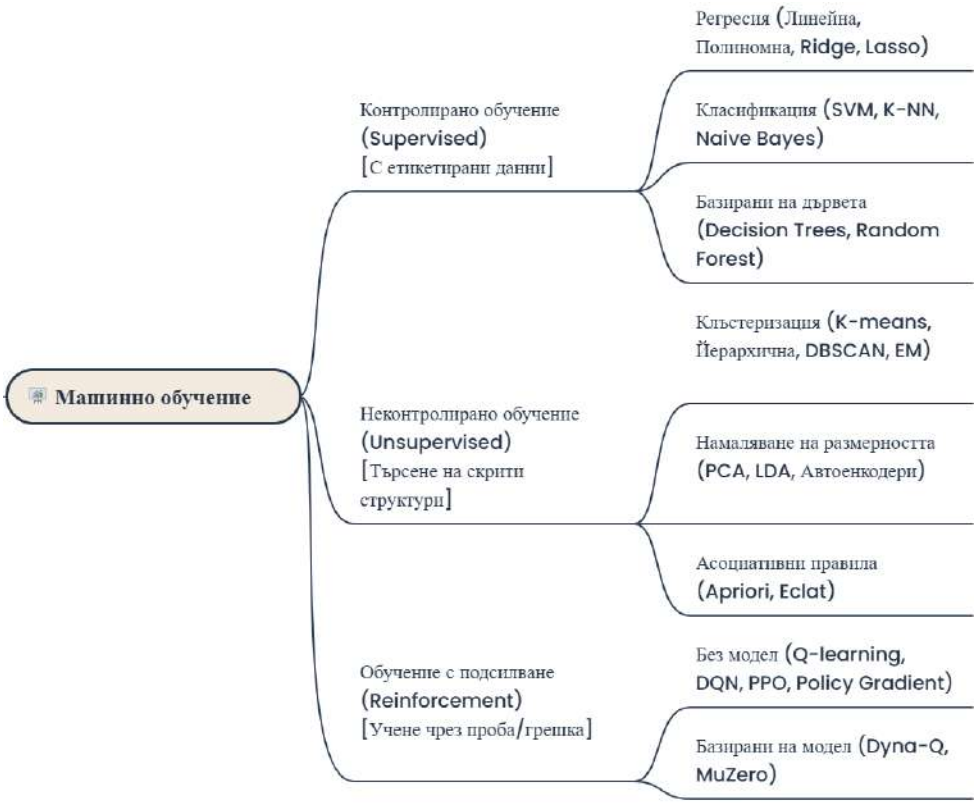
Процесът на обучение може да се разглежда като вид „откривателска“ дейност, много подобен на основната парадигма на експерименталната наука – натрупване на експериментални данни и извеждане на закони въз основа на тези данни. Единствената разлика е, че експерименталната наука се провежда от хора, докато машинното обучение е процес, изпълняван от интелектен агент.

Във всеки процес на машинно обучение могат да се различат четири основни елемента, илюстрирани чрез примера с Галилео Галилей и закона за свободно падане:

1. входна (независима) променлива – времето t;
2. изходна (зависима) променлива – изминатото разстояние s;
3. функционална зависимост – изведеният закон $s = \frac{1}{2}gt^2$;
4. метод за избиране на правилната функционална зависимост – при Галилей – шанс, първоначални опити и интуиция, а при автономните системи – систематични алгоритми за откриване на решение.

В процеса на индуктивното разсъждение се различават два последователни етапа – извличане на знание от примери и използване на това знание върху нови входни данни. Способността за обобщение въз основа на примери е най-характерната особеност на системите с възможности за обучение.

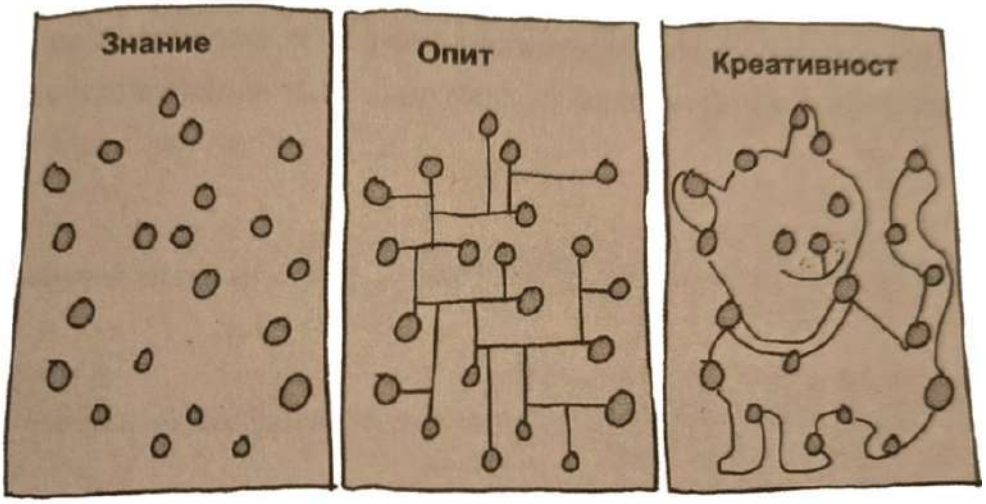
От теоретична гледна точка е важно да се знае, че индуктивното разсъждение е недостоверно и няма същия статус като дедуктивното разсъждение. Изведеното знание е само една хипотеза, която добре обяснява експерименталните данни. Колкото и добро да е това обяснение, отникъде не следва, че хипотезата е вярна и важи винаги.



Фигура 18. Класификация на алгоритмите за машинно обучение

Уклон (bias) на една обучаваща се система е нейното предпочитание към едни или други хипотези в процеса на търсене на функционална зависимост. Уклонът представлява подреждането на предпочитанията към различните хипотези – винаги се избира възможно най-предпочитаната хипотеза, която обяснява експерименталните данни. Повечето уклони са подчинени на общ принцип, известен като **бръсначът на Окам** – предпочитание се дава на най-простото удовлетворително обяснение на дадено явление.

Ако се формулира целта на процеса на машинно обучение като максимизиране на някакъв критерий *P*, от математическа гледна точка този процес представлява оптимизация на функция върху множеството на всички възможни решения. Тази връзка между алгоритмите за машинно обучение и тези за оптимизация е много важна, защото ни позволява да използваме мощните методи за нелинейно оптимизиране.



Фигура 19. Знание, опит креативност (адаптирано по MacLeod, 2014)[13]

Критерии за спиране на обучението са:

– прагова стойност – спиране при достигане на определен праг за критерия;

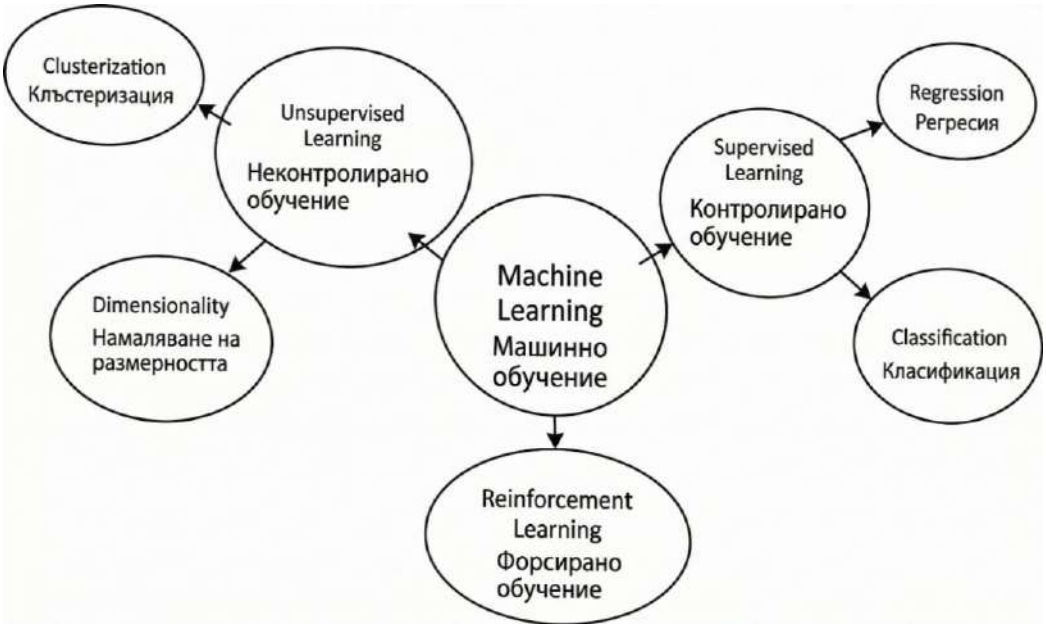
– стагнация – спиране, когато всяка следваща стъпка не подобрява съществено критерия;

– валидация – специфичен метод за машинно обучение, който осигурява способността за обобщение.

За да се избегне преобучение, се съставя отделен набор примери, наричан тестващо множество. Тези примери не се използват за определяне каква ще е следващата стъпка в търсенето на по-добро решение. Обикновено, в началото с подобряването на критерия върху обучаващото множество, се подобрява и този върху тестващото – системата се учи успешно. В определен момент обаче критерият върху тестовото множество започва да се влошава и това е сигнал за спиране на обучението.

Обучението на AI не е магия, а сложен процес на математическа оптимизация. Той включва настройка на милиони тегла (w) и отклонения (b), за да се минимизира разликата между прогнозата и реалността.

[13.] MACLEOD, H. Information vs. Knowledge. [Online]. Gapingvoid, 14 януари 2014. [Прегледан на 3 юни 2026]. Достъпен на: https://www.gapingvoid.com/blog/2014/01/14/information-knowledge/.



Фигура 20. Класификация на видовете машинно обучение

3.2.1. Дърво на решенията, класификация и регресия

Методът „Дърво на решенията“ има за цел да преобразува наличните данни в логическа последователност от правила и действия. Алгоритъмът за създаване на такава структура е основан на идеята за постепенно разделяне на данните на непрекъснато нарастващи клонове на дървото на решенията.

Основна мярка в случая е ентропията, характеризираща степента на подреденост на дадена структура. Източникът на обучение при генериране на дървовидни структури е минимизиране на ентропията чрез максимизиране на критерий, наречен информационна печалба (Information Gain):

$$Information\ Gain = Entropy\ (преди) - Entropy\ (след)$$



Фигура 21. Дърво на решенията

Той определя количествено ентропията преди и след разделянето на дадено дърво на нови клонове. Алгоритъмът за обучение няма да раздели дървото и да добави клон, ако не увеличим информационната печалба. Коренът на дървото на решенията (root) съдържа основния въпрос и свързаните с него променливи. Дървото показва как решението зависи от различни стойности на променливите, които разделят данните на клонове. Процесът на разделяне дефинира съответни възли на решения.

Предимства на тези модели са, че те са лесни за разбиране от хората, може лесно да се дефинират бизнес правила и избор на променливи, но са ненадеждни при различен набор от данни. Често се прилагат при инвестиционни решения, банкови заеми, анализ на отлив на клиенти и др.

3.2.2. Опорни векторни машини

Съществува социална аналогия, която може да помогне за разбирането на понятието „опорен вектор“. Например, ако имаме две политически партии, които не променят подкрепата си лесно, а резултатът от изборите зависи от около 20% от нерешителните избиратели. Тези гласоподаватели са аналог на „социалните опорни вектори“.

В машинното самообучение опорните вектори се използват основно за класификация, тоест за разделяне на данни в различни категории (например спам срещу нормален имейл, болен срещу здрав пациент). Теорията на статистическото обучение показва, че не се нуждаем от всички данни, за да

решим някои проблеми, а просто е необходимо да съпоставим някои точки. Ето как се съпоставят нещата:

● Твърдият електорат (Партия А и Партия Б) – това са точките от данни, които са дълбоко в своята категория. Те са ясно разпознаваеми и са далеч от „фронтовата линия“. За алгоритъма те не са толкова интересни, защото позицията им не влияе на това къде точно ще се прокара границата.

● Колебаещите се избиратели (Опорните вектори) – това са точките от данни, които се намират най-близо до другата категория. Те са „най-трудните“ за класифициране. Точно както в изборите, само тези точки определят крайния резултат (къде ще мине линията на разделение). Те буквално „опират“ или „поддържат“ границата. Ако се премахнат всички сигурни гласоподаватели от модела, границата няма да помръдне. Но ако бъде премахнат или преместен един колебаещ се (опорен вектор), границата ще се промени.

За да стане разбираема пълната картина, ето трите основни технически термина, обяснени просто:

● Хиперравнина (Hyperplane) – това е самата граница, която разделя данните. Ако данните са на 2D лист хартия, хиперравнината е просто една права линия. Ако са в 3D пространство, тя е двуизмерен лист (равнина). Целта на алгоритъма е да намери най-добрата хиперравнина, която разделя двете класи.

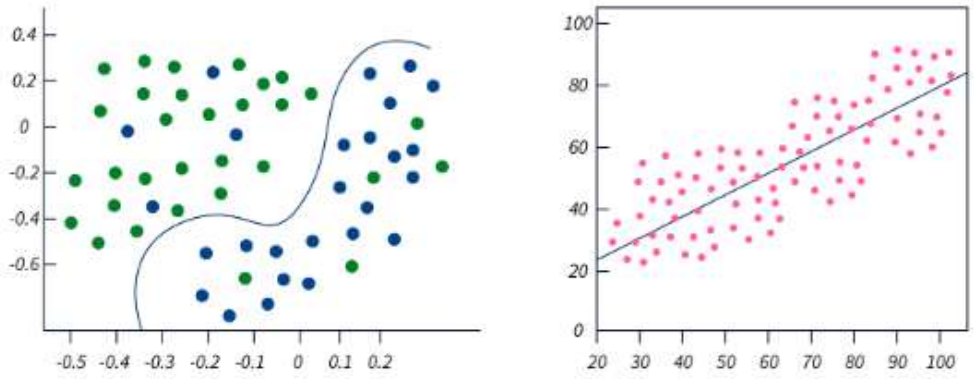
● Марж (Margin) – това е „ничията земя“ или буферната зона около хиперравнината. Целта не е просто да начертае линия между двете класи, а да начертае линията така, че тя да е максимално отдалечена от най-близките точки (опорните вектори) и на двете групи. По-голям марж означава по-сигурна и стабилна класификация при нови, непознати данни.

● Опорни вектори (Support Vectors) – както вече уточнихме, това са данните, които лежат на самия ръб на този марж. Те са скелетът на модела.

Ако си представим, че избирателите не са ясно разделени на „ляво“ и „дясно“, а са напълно смесени (например, живеят в едни и същи квартали), една права линия няма да може да ги раздели. Тук методът използва нещо, наречено „Ядрен трик“ (Kernel Trick). Алгоритъмът добавя ново измерение (например, освен по квартал, започва да ги разглежда и по доходи или възраст), „повдига“ данните в 3D пространство и изведнъж става възможно да се прокара права линия (равнина) между тях. Тази характеристика дава две уникални възможности:

– да се сравни информационното съдържание на всяка нова точка от данни със съществуващите опорни вектори и да се определи дали тя включва неизвестна информация (новост);

– да се кондензира големият брой записи на данни в малък брой опорни вектори с високо информационно съдържание – често 10 – 15% от оригиналните записи.



Фигура 22. В ляво – класификация, а в дясно – регресия

На практика това се прави със сложна математика, но днес има готови библиотеки в езици за програмиране като Python, най-често библиотеката scikit-learn. Ето как изглежда процесът стъпка по стъпка в реалния свят след като импортираме библиотеката:

– имаме таблица с данни (например характеристики на тумори и информация дали са доброкачествени или злокачествени);

– зареждане в модела тези данни на алгоритъма;

– обучаваме алгоритъм, като той анализира данните и намира оптималната граница и опорните вектори;

– даваме на алгоритъма нови, невиждани досега данни и той предсказва и ни казва в коя категория попадат.

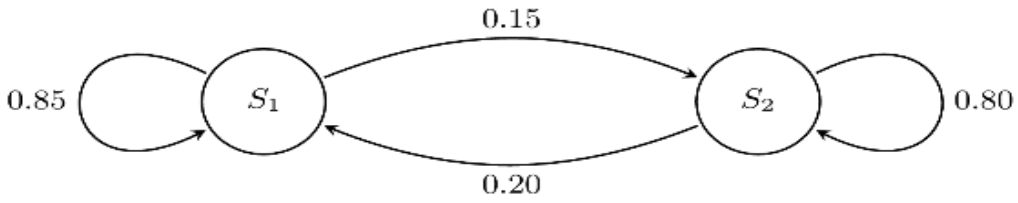
Недостатъкът е, че има множество параметри за настройка, включително сложни математически трансформации, наречени „ядра“ (kernels). Въпреки това често се прилага в медицински приложения (откриване на левкемия, рак, анализ на мамограми), биоинформатиката (класификация на тъкани, структура на протеини), производство (косвени сензори, усъвършенстван контрол на процесите), категоризиране на текст.

3.2.3. Вериги на Марков

Веригите на Марков представляват математически модели, описващи поредица от вероятни събития, в които вероятността за всяко събитие зависи от състоянието, постигнато по време на предходното. Моделът прогнозира резултати, базирани на вероятности и наличните данни. След като бъде обучен, той трябва само да знае условията в непосредственото минало за да извлече свързана информация и да прогнозира бъдещето.

Тези вериги се основават на формализирането на динамични системи, чието бъдещо състояние зависи единствено от настоящото, но не и от поредицата събития, които са довели до него. Това е известно като Свойство на Марков. Формално за последователност от състояния S_1, S_2, ..., S_n вероятността за преход към състояние S_{t+1} се определя условно само от текущото състояние S_t:

$$P(S_{t+1} \mid S_t, S_{t-1}, ..., S_1) = P(S_{t+1} \mid S_t)$$



Фигура 23. Верига на Марков

Ако си представим, че имаме AI система за противовъздушна отбрана (ПВО), която решава дали да стреля по цел. Диаграмата описва логиката на софтуера, който управлява радара и ракетите:

– S1 е режим „Наблюдение“. AI следи небето и всичко е нормално, което съвпада 0,85 (85%) шанс това да е така. AI преценява, че обектите на екрана са птици или облаци и остава в режим наблюдение в състояние S1. Той остава в режим наблюдение и не хаби скъпи ракети;

– При шанс 0,15 (15%) AI засича скорост и траектория, които приличат на вражеска ракета. Той взема решение да премине в Режим „Прицелване“ или състояние;

– S2. Тук AI е насочил оръжията и е готов за стрелба. При 0,80 (80%) шанс системата потвърждава, че това наистина е заплаха. Тя задържа фокуса върху целта, докато чака идеалния момент за изстрел;

– 0,20 (20%) шанс – AI прави допълнителен анализ и разбира, че това е „фалшива тревога“ (например пътнически самолет с грешен сигнал). Той веднага сваля готовността и се връща в Режим „Наблюдение“ (S1).

Военните инженери използват тези числа, за да настроят агресивността на изкуствения интелект:

– чувствителност (0,15) – ако увеличим това число на 0,50, AI ще бъде „параноичен“ – ще се прицелва във всичко, което лети;

– скептицизъм (0,20) – това е защитата срещу грешки. Ако това число е 0,01, веднъж прицелил се, AI почти никога няма да се откаже, което може да доведе до трагични инциденти (приятелски огън).

Веригите на Марков имат широко приложение в изкуствения интелект – от прогнозиране на климатични модели и финансови пазари до системите за предлагане на текст. Съвременните смартфони използват тези модели за предвиждане на думи при въвеждане на текст – функцията, която познава думата по първите ѝ букви, се базира на вероятностни модели, обучени върху големи корпуси от езикови данни. Веригите на Марков са фундаментални за разбирането на скритите модели, които намират приложение в разпознаването на реч, обработката на естествен език и биоинформатиката.

3.2.4. Алгоритъм за обратно разпространение на грешката

Тъй като директното изчисление е невъзможно, при сложни мрежи се използва **стохастично градиентно отклонение** – итеративен процес, при който алгоритъмът прави малки стъпки в посока, която намалява грешката. Модерният стандарт за оптимизация е алгоритъмът Adam (Adaptive Moment Estimation), предложен от Кингма и Ба (Kingma & Ba, 2014), който адаптивно регулира скоростта на обучение за всеки параметър.

Обратното разпространение на грешката е ключовият механизъм, чрез който мрежата се учи от грешките си. Ако например обучаваме един робот-стрелец да уцелва центъра на мишена, в началото той е пълен новобранец и стреля напосоки. Обратното разпространение е начинът, по който той се учи от всеки пропуск и то минава през следните стъпки:

1. Изстрелът (изчисляване на грешката). Роботът стреля, данните за насоката минават през неговата мрежа, но уцелва далеч от центъра.

2. Колко много сбъркахме? (Връща грешката през слоевете). Роботът изважда една линийка и измерва разстоянието от дупката до центъра на

мишената. Това разстояние е грешката (или „загубата“). Колкото по-голямо е то, толкова по-лошо се е справил.

3. Кой е виновен? (Backpropagation). Роботът започва да мисли отзад напред: „Ръката ми се разтрепери в последния момент? А може би краката ми не бяха стабилни?“. Той пресмята градиента – това е просто име за въпроса колко точно виновна е всяка част от тялото ми за този лош изстрел. Ако китката е виновна на 80%, а кракът само на 5%, той знае къде трябва да направи най-голямата промяна.

4. Малка поправка (корекция). Ако роботът направи огромна промяна, може да подмине центъра, затова той прави малка промяна (стохастично градиентно отклонение). Той леко стяга китката.

Този цикъл се повтаря хиляди пъти (епохи), докато мрежата достигне желаната точност. Започваме обяснението от изходния слой, където имаме стойността на грешката. След това изчисляваме тегловните коефициенти на невроните в скрития слой на мрежата. В края на обратното разпространение изчисляваме останалите тегловни коефициенти за матрицата на връзките между входния и скрития слой.

Алгоритъмът за обратно разпространение не е защитен от капана на локалните минимуми. Намереният минимум не гарантира най-ефективната структура на модела с минималната грешка, защото той не е глобален. Ако започваме процеса на обучение в различни позиции, е вероятно да достигнем до различни локални минимуми, които генерират модели с различно качество. За компенсация на този фундаментален недостатък се използват множество трикове, но най-сигурното средство е да се победи случайният характер на попадение в глобалния минимум чрез многократно пускане на алгоритъма.

3.2.5. Методи на градиентното спускане и оптимизация на теглата

Спускането по градиент е оптимизационен алгоритъм, който води към минимума. Разликата между спускането по градиент и обратното разпространение е, че обратното разпространение е процес на изчисляване на производните, докато спускането по градиент е процес на намиране на минимума. Параметрите на модела *w* се настройват така, че да се спускат надолу от началната си стойност по функцията на загубите – J(w). Когато достигнат оптималните си стойности, градиентът е нулев, т.е. функцията на загубите е минимална J(w).

Представете си, че сте се изгубили в планината през нощта и е паднала гъста мъгла. Трябва да слезете до хижата, която се намира в най-ниската точка на долината, но не се вижда нищо.

1. Градиентно спускане: пътят надолу. Тъй като не виждаш хижата, ти използваш краката си, за да усетиш наклона на земята под теб. Твоето тегло (w) е твоята позиция в планината. Функция на загубата (J(w)) е твоята височина. Колкото си по-високо, толкова по-голяма е „загубата“ (грешката). Твоята цел е височина нула (хижата). Градиентът е наклонът. Ти правиш малка крачка в посоката, където усещаш, че е най-стръмно надолу. Обратното разпространение е просто сензорът в краката ти, който ти казва „наклонът е наляво“. Градиентното спускане е самото ходене стъпка по стъпка надолу.

2. Камъните по пътя са препятствия. Има локален минимум, малка дупка на склона. Ако слезеш в нея, усещаш, че накъдето и да тръгнеш е „нагоре“, и решаваш, че си стигнал хижата. Но тя всъщност е още много по-надолу. Заклещваш се там, защото краката ти не усещат наклон и си мислиш: „Тук ще си остана“. В допълнение към капана на локалния минимум може да има и други препятствия – например седлова точка или изчезващ градиент, който е голяма, идеално плоска поляна. Краката ти не усещат никакъв наклон. Не знаеш накъде да тръгнеш и спираш да се учиш (да вървиш).

3. Трансферното обучение на голям AI модел е като да строиш магистрала през цялата планина – отнема седмици и струва милиони. Ако ти трябва просто пътечка до твоето село (Модел Б), няма нужда да строиш всичко от нулата. Вземаш Модел А (голям модел, който вече знае как изглеждат планините, дърветата и камъните). Просто „дорисуваш“ последните 100 метра до твоята врата. Това е Transfer Learning. Пести време, пари и компютърна мощ (Rickards, 2025)

Важно е да се отбележи, че процесът на обучение в мозъка не се основава на принципите на обратното разпространение. Хората не се обучават на базата на изчисляване на производни. Мозъкът не е калкулатор. Когато се учите да карате колело, не изчислявате „производни“ и „градиенти“ в главата си. Не смятате сложни математически уравнения, за да разберете колко точно да натиснете спирачката. Мозъкът учи чрез опит, биологични сигнали и асоциации, които са много по-сложни и различни от математическото „спускане по наклон“, което ползва AI.

Градиентното спускане се използва за намиране на най-ефективното състояние на модела, при което той прави най-малко грешки (минимална функция на загубите). Трансферното обучение спестява ресурси и позволява

използването на огромни модели (като GPT) за специфични, малки задачи без нужда от суперкомпютри. Прилага се, когато нямаме достатъчно примери (напр. за разпознаване на снимки на рядка болест), като се използва „базовото знание“ на модел, обучен на милиони общи изображения.

3.3. Математически модел на изкуствен неврон и архитектура на перцептрон

Невронните мрежи са първата технология в историята на изкуствения интелект. Първата публикация, свързана с тази област, е от 1943 г. за „електронния мозък". Това е първият опит да се имитира клетка на мозъка – невронът (McCulloch & Pitts 1943).

Невронът има клетъчно тяло с ядро, входове от синапси/дендрити и изходи от аксони. Невронът се задейства от електрически импулси в посока от ляво надясно. Той реагира на входните сигнали чрез силата на връзката на синапсите. В обратния случай синаптичната връзка отслабва, а под някаква долна граница дори изчезва. През 1943 г. Уорън Маккълок (1898 – 1969) и математикът Уолтър Питс (1923 – 1969) демонстрират, че мрежа от елементи, наподобяващи човешките нервни клетки, по които протичат електрически сигнали, може да имитира машината на Тюринг (McCulloch & Pitts 1943). Те допускат, че мозъкът може би е вид жив компютър, което означава, че един „електрически мозък“ може да изпълнява същите програми като човешкия. Тази теория е известна под името „множествена реализируемост“.

Маккълок и Питс си дават сметка, че невроните могат да функционират като логически портали — устройства, които могат да се включват и изключват в зависимост от входящите данни. Учените описват въображаем неврон, който наричат гранична логическа единица. Този неврон работи, като първо добавя стойност към входящите си данни (сигналите от другите неврони), след което умножава тази стойност по променлива, наречена тегло, която изразява силата на връзката между невроните. Ако входящите сигнали надхвърлят определена стойност (праг), невронът се активира и изпраща изходящ сигнал. Това активиране се нарича активираща функция.

Биологичната структура е представена с прост математически модел като съвкупност от входове, представляващи синапси и дендрити, клетъчно тяло, изразено като сумиращо устройство, и аксони, отразяващи изходите. Моделът на Маккълок-Питс е силно опростена биологична аналогия, която работи с бинарни сигнали. Той се състои от следните компоненти:

– входове (x_i) – набор от бинарни сигнали x_1, x_2, ..., x_n, където $x_i \in \{0,1\}$;

– тегла (w_i) – в оригиналния модел теглата са фиксирани (възбуждащи +1 или задържащи/инхибиторни);

– суматор ($\sum$) – изчислява линейна комбинация от входовете;

– праг (θ) – стойност, която определя дали сумата е достатъчна за активиране;

– активираща функция (f) – използва се прагова функция (стъпаловидна функция на Хевисайд).

Изкуственият неврон е изчислителна единица, която приема поредица от входове *(X_i)* и произвежда един изход. Процесът включва три основни стъпки:

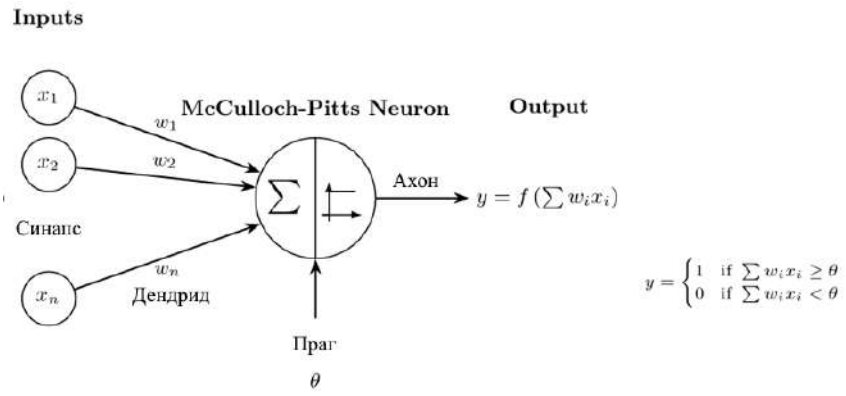
– претеглена сума, като всеки вход се умножава по тегло (W_i), което определя неговата важност за крайното решение;

– добавяне на отклонение – параметърът θ позволява на мрежата да регулира прага на активиране на неврона;

– активационната функция (F) е нелинейна функция, определяща дали невронът ще „се активира“ и каква стойност ще пренесе към следващия слой.

Математически изчислението се представя като:

$$Output = F(\sum_{i=1}^{n}(x_i.w_i) + \theta)$$



Фигура 24. Математичен модел на биологичния неврон

За да илюстрираме принципа, нека разгледаме опростен пример за вземане на решение дали да отидем във Франция, базиран на два фактора (входа):

– x_1 : Самолетният билет струва по-малко от $1000;

– x_2: Вашият партньор може да дойде с вас.

Всеки вход има „тегло“ (w), което определя неговата важност. Ако сте с ограничен бюджет, цената е важна, затова задавате тегло $w_1 = 4$. Ако присъствието на партньора е по-малко важно, задавате $w_2 = 3$. Невронът изчислява претеглена сума и я сравнява с праг ($b – 4$). Функцията за решение изглежда така:

$$(x_1 * 4) + (x_2 * 3) – 4 > 0$$

Представено като вектори на възможните състояния (x_1, x_2):

- (0,0): (0*4) + (0*3) – 4 = – 4 < 0 → не отиваме
- (1,0): (1*4) + (0*3) – 4 = 0 → прагът не е преминат
- (0,1): (0*4) + (1*3) – 4 = – 1 < 0 → не отиваме
- (1,1): (1*4) + (1*3) – 4 = 3 > 0 → активиране (отиваме)

Въпреки своята революционност, M-P невронът има съществени недостатъци:

– фиксирани тегла и няма механизъм за обучение (теглата се задават ръчно);

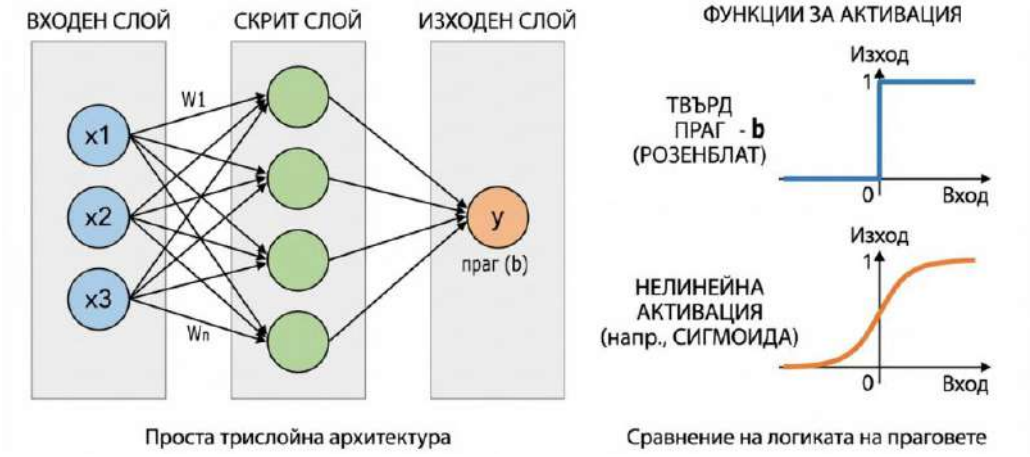
– линейна сепарабилност – може да решава само проблеми, които са линейно разделими;

– бинарни данни – работи само с дискретни стойности 0 или 1, което ограничава приложението му при сложни непрекъснати сигнали.

Този пример показва, че елементарни изчислителни единици могат да извършват сложни логически операции, което е фундаменталният принцип на цифровата изчислителна техника. В мозъка на човека обаче има около 86 млрд. неврона и не можем да постигнем много с изграждането на математически модели само на един от тях. Решението е в една по-сложна математическа структура, наречена невронна мрежа.

Перцептронът е невронна мрежа с една от най-простите структури от три слоя – входен, скрит и изходен слой. Входният слой включва входните сигнали/променливи. Решаваща е ролята на скрития слой, в който се кодират връзките между входните и изходните величини, открити чрез машинното

обучение. От особено значение е възможността да се представят нелинейни зависимости.



Фигура 25. Структура на изкуствена невронна мрежа (Перцептрон)

Ранните модели (като Перцептрона на Розенблат от 1958 г.) са използвали „твърди“ прагове – или се активира (1), или не (0). Реалният свят обаче е пълен с несигурност и затова учените въвеждат т.н. сигмоидна функция, която преобразува суровите числови сигнали в плавен диапазон от 0 до 1, който позволява на модела да изчисли точно колко трябва да се променят теглата при всяка стъпка на обучение, за да се минимизира грешката.

$$\sigma(y) = \frac{1}{1 + e^{-y}}$$

С това изкуствения интелект може да изразява степен на увереност – например „има 85% вероятност това изображение да е танк“.

Едни от първите реални приложения на тези алгоритми исторически са били използвани именно от военните за работа на радарните системи. Нека си представим, че имаме оръжейна система, работеща с невронна мрежа, изобразена по-горе, която трябва да отговори на въпроса да изстреляме ли ракета по засечения обект? (*1 = Огън!, 0 = Не стреляй!*):

1. Входен слой – предоставя суровите данни, които радарът улавя за летящия в небето обект, като x1 е скорост (дали е бързо и бавно), x2 е височина (дали е високо или ниско) и x3 е радарно сечение (голям или малък).

2. Скритият слой има междинни „мозъчни центрове“, които търсят специфични заплахи. Всяко зелено кръгче получава данни от всички сензори едновременно, но им придава различна важност (това са черните линии, наречени тежести). Първият зелен неврон може да е обучен да разпознава крилати ракети. Той ще следи най-вече за специфична комбинация – малък размер (x3) + ниска височина (x2) + висока скорост (x1). Вторият зелен неврон може да е специализиран за тежки бомбардировачи. Той ще търси голям размер + голяма височина. Третият може да има задачата да филтрира ята птици (ниска скорост + малък размер + ниска височина) и да праща сигнал на системата да ги игнорира.

3. Изходен слой *y* събира рапортите от всички „офицери анализатори“ (зелените неврони от средния слой) и въз основа на тях трябва да вземе крайното решение – пускаме ли ракетата или не.

4. Функциите за активация (дясната част на картинката). Изходният *y* решава това въз основа на математическите пресмятания. Тук идва разликата в двете графики вдясно. Горната графика (твърд праг на Розенблат) може да се опише със стара, безкомпромисна автоматизирана система. Командирът събира данните и ако те преминат определена числова граница, дори с частица от процента, синята линия скача директно от 0 на 1. Системата автоматично извежда 1 (ОГЪН!). Проблемът тук идва в нюансите. Ако пътнически самолет лети малко по-бързо от нормалното поради силен попътен вятър, системата с твърд праг може да прескочи границата и директно да стреля по цивилна цел. Затова долната графика, изобразяваща нелинейна активация – сигмоида, прави системата по-модерна и интелигента. Тя не дава сляп отговор „Да/Не“, а използва онази плавна, S-образна оранжева линия, за да превърне данните във вероятност за заплаха. Вместо да стреля сляпо, системата пресмята данните y = (x1. w1) + x2 .w2) + (x3.w3) + b и слагайки полученото число в сигмоидата го преобразува в процент между 0 и 1, казва: „Има 0,92 (92%) вероятност това да е вражески изтребител“. Това позволява да се направят различни протоколи за действие – под 0,30 игнорирай (вероятно птици или метеорологичен балон), от 0,30 до 0,80 вземи на прицел и опитай радиовръзка (може да е приятелски самолет с повреден транспондер), над 0,80 вдигай тревога.

Освен това, тази плавност на сигмоидата позволява на системата да се обучава по време на симулации. Ако на тренировка алгоритъмът каже

„90% сигурен съм, че е враг“ за приятелски изтребител, програмистите (или самият компютър) могат леко да намалят силата на онези черни линии (тежестите) при зелените неврони. Следващият път системата ще каже 80%, после 50%, докато не се научи да дава правилен резултат (близо до 0) за приятелски цели.

Впоследствие невронната мрежа може се учи съвсем сама с процес, наречен backpropagation, или обратно разпространение. Като цяло невронните мрежи имат както фундаментални предимства, така и недостатъци. Основните предимства се обуславят от тяхната възможност за извличане на зависимости от данни чрез машинно обучение, като невронните връзки могат да бъдат нелинейни. Съществува теоретично доказателство, че можем да открием всяка аналитична зависимост с достатъчен брой неврони и скрити слоеве (Hornik, Stinchcombe & White 1989). Освен това те имат способността да се самоорганизират и да откриват вътрешните структури в данните.

Основните недостатъци са, че те се превръщат в т.н. „черна кутия“, което не е приемливо за някои среди като отбраната. Невронните мрежи не могат да предсказват добре извън диапазона на данните за обучение — дори малко отклонение извън тези диапазони довежда до неточни предсказвания или дори сриване на модела. За постигане на добра генерализация са необходими значителни количества обучителни данни. Невронните мрежи се употребяват в области като маркетинг за насочване на клиенти, нелинейно управление на процеси с икономически ефект, намаляване на емисиите на парниковите газове, откриване на измами и дезинформация, медицински приложения и автомобилна индустрия.

3.4. Дълбоко обучение и дълбоки невронни мрежи

В началото на XXI в. на основата на невронните мрежи се изгражда най-успешният модерен метод на изкуствения интелект, наречен *дълбоки невронни мрежи* (Deep Neural Networks). Основната идея е да се генерират сложни модели чрез добавяне към невронната мрежа на голямо количество скрити слоеве с различна архитектура (Goodfellow, Bengio & Courville, 2016). Тази йерархична структура позволява на мрежата да разбира сложни модели стъпка по стъпка:

– входен слой – приема суровите данни (напр. 65 536 неврона за изображение 256×256 пиксела);

– скрити слоеве – извършват междинна обработка. Първите слоеве разпознават ръбове и ъгли, следващите – форми, а последните – цели обекти (напр. танкове или самолети);

– изходен слой – предоставя финалната прогноза или класификация.



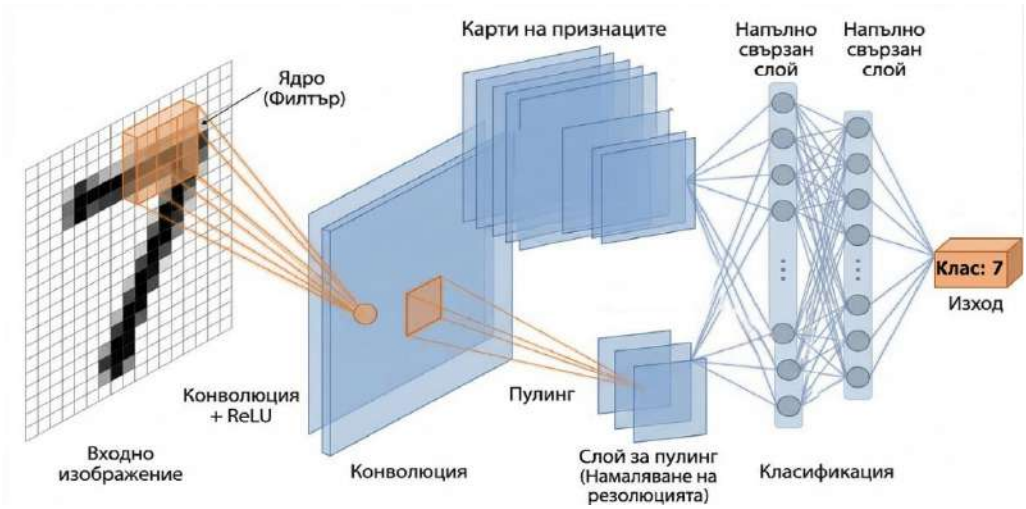
Фигура 26. Алгоритми за дълбоко обучение

Този подход, вдъхновен от зрителната кора на мозъка, позволява постигането на точност над 99% при задачи като анализ на почерк и разпознаване на образи. Днес съществува голямо разнообразие от алгоритми за дълбоко обучение на невронни мрежи, които са подходящи за различни дейности, като разпознаване на изображения, генериране на текст и други.

3.4.1. Конволюционни невронни мрежи (CNN) за разпознаване на образи

Конволюционната невронна мрежа (Convolutional Neural Network, CNN) е основната архитектура на дълбоките невронни мрежи за обработка на изображения. Главната операция, наречена **конволюция**, извлича прости характеристики като линии, ъгли и ръбове в оригиналното изображение. След това има **блок за максимално обединяване**, който подсилва и улавя най-важните характеристики. После пак се прави още една конволюция с по-

сложни характеристики и се подсилват, докато се постигне надеждна класификация в последния слой. CNN използват филтри (ядра), за да извличат характеристики от входа, използвайки метод, известен като *конволюция*. Те са гръбнакът на компютърното зрение и системите за автоматично разпознаване на цели в отбранителните приложения.



Фигура 27. Конволюционна невронна мрежа (CNN)

Представете си, че гледате огромна и много размазана снимка на която трябва да открием вражески танк и търсим с лупа, обхождайки картата сантиметър по сантиметър. Ето как работи конволюционната невронна мрежа стъпка по стъпка:

1. Конволюция (търсене на детайли). В мрежата нашата лупа е специален филтър (ядро). Този филтър е програмиран да реагира само на определени неща – например на права хоризонтална линия. Когато лупата мине през място, където има линия, тя „светва“. Първите слоеве на CNN намират само прости неща – чертички, точки и ъгли. Това е като да се види, че на снимката има нещо метално и дълго (може би оръдие?).

2. Максимално обединяване. Намирането на всички чертички например води до натрупване на твърде много информация. Следва разделяне на снимката на малки квадратни сектори (например 2 x 2 пиксела) и преценка коя от дадени 4 точки свети най-силно. Оставя се най-силният резултат и останалите се изхвърлят. Резултат се състои в това, че снимката става по-

малка и по-лесна за обработка, но най-важните характеристики (ръбовете на танка) стават по-видими.

3. Дълбочина – от чертички към обекти. След първия блок процесът се повтаря, но става по-умен. Втората конволюция не търси просто чертички, а събира чертичките от предния слой и разпознава форми (например кръг от колело или верига). Третата конволюция вече събира формите и „вижда“ цялата купола на танка. С всеки следващ слой мрежата разбира все по-сложни обекти.

В една модерна ракета или дрон CNN работи в реално време, като на входа е камерата на дрона, която подава хиляди снимки. AI „филтрира“ шума (дървета, трева) и търси специфични геометрични форми на военна техника. След това системата игнорира маловажните детайли (като цвета на тревата) и се фокусира върху силуета на машината. Накрая, в последния слой мрежата, установява с 99% сигурност, че това е Т-72, а не трактор.

3.4.2. Рекурентни невронни мрежи (RNN) за последователни задачи

Докато обикновената мрежа гледа снимка и казва „това е танк“, рекурентната невронна мрежа (Recurrent Neural Networks – RNN) гледа поредица от събития, подобно на видео. Тя има цикъл (обратна връзка). Когато обработва текущата стъпка, тя добавя към нея „бележка“ от това, което е видяла преди малко. Използват се за обработка на последователни данни –такива, които са в строго определен ред, като естествените езици. Докато класическите невронни мрежи обработват отделни информационни точки, RNN поддържат важната структура и взаимовръзки в рамките на последователните данни.

Представете си, че четете изречение. Човек не чете всяка дума като отделен остров, а помни началото на изречението, за да разбере края му. Ако прочетеш думата „той“, веднага очакваш глагол. RNN обработва входната последователност $x = (x_1, x_2, ..., x_t)$ стъпка по стъпка. На всяка времева стъпка t, скритото състояние h_t се изчислява с натрупване като функция от текущия вход x_t и предходното скрито състояние h_{t-1}:

$$h_t = \phi\ (W_h h_{t-1} + W_x x_t + b),$$

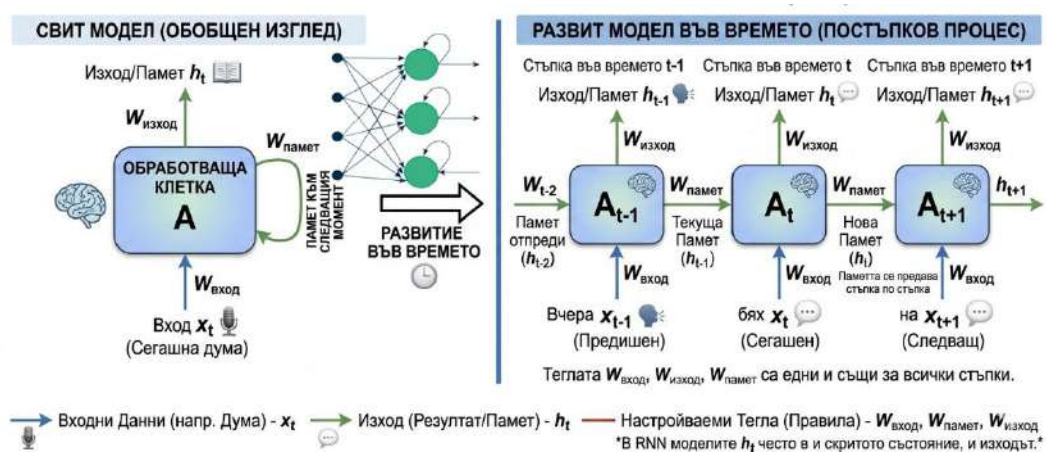
където x_t е сегашният вход, думата, която чете в момента;

h_{t-1} – миналото, спомен за всичко, което е прочел досега;

h_t – новият спомен, смесва новата дума със стария спомен и получава актуализирана „картина“ в главата си;

W_h и W_x – тегловни матрици, b е вектор на отместването, а ϕ е активационна функция.

Проблемът на тези мрежи е, че страдат от изчезващия градиент, защото имат къса памет. При обучението се използва Backpropagation Through Time (BPTT), нещо като верига от хора, които си предават съобщение шепнешком (градиента). Ако веригата е много дълга и всеки човек предава съобщението малко по-тихо (малки стойности в матриците), докато се стигне до първия човек, съобщението вече е пълна тишина (нулева промяна). В резултат мрежата спира да се учи от началото на изречението и помни само последните думи.



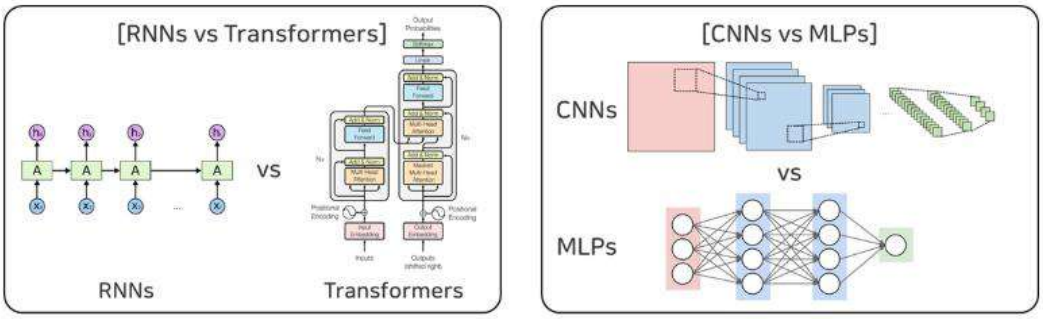
Фигура 28. Рекурентна невронна мрежа (RNN)

Този вид мрежи са подобрени с някои разновидни архитектури, като тези на **Дългата краткосрочна памет (LSTM – Long Short-Term Memory)**, проектирана да преодолее този проблем чрез въвеждането на клетъчно състояние (C_t) и сложен гейтов механизъм. Компонентите на LSTM клетката включват:

– забравяща порта – определя каква част от предходното клетъчно състояние C_{t-1} да бъде запазена;

– входна порта – регулира каква нова информация да бъде добавена в клетъчното състояние;

– изходна порта – определя коя част от клетъчното състояние да бъде изведена като скрито състояние h_t.



Фигура 29. Разлика между RNN и CNN

Благодарение на тази структура, LSTM мрежите успяват да поддържат стабилни градиенти при много дълги последователности, което ги прави незаменими при анализ на реч, машинен превод и откриване на аномалии в сложни системи за сигурност.

3.4.3. Генеративните състезателни мрежи (GANs) и теория на игрите

Съществуват и т.н. **генеративни състезателни мрежи (Generative Adversarial Networks – GANs).** Тази архитектура е предложена от Иън Гудфелоу през 2014 г. Тези мрежи се основават на теорията на игрите, по-конкретно на минимаксните игри с нулева сума, където две невронни мрежи (генератор и дискриминатор) се обучават едновременно чрез взаимна конкуренция, състезавайки се помежду си.

Теорията на игрите е клон от математиката, предложен от Джон Наш през 1949 г., който изучава стратегическото вземане на решения. Тя се прилага в ситуации, в които крайният резултат за един участник (играч) зависи не само от неговите собствени действия, но и от действията на останалите участници. В науката „игра“ не означава забавление, а всяка ситуация с играчи (участници, които вземат решения), техните стратегии и печалби. Резултатът е награда или наказание за всяка комбинация от стратегии.

Това може да се онагледи с известната дилема на затворника, която илюстрира защо двама души или организации може да не си сътрудничат,

дори когато това е в техен общ интерес. Според условието имаме двама души (А и Б), заловени и разпитвани в отделни стаи. Полицията няма достатъчно доказателства и предлага на всеки следното:

– ако ти предадеш другия, а той мълчи – ти излизаш на свобода, а той влиза в затвора за 10 години;

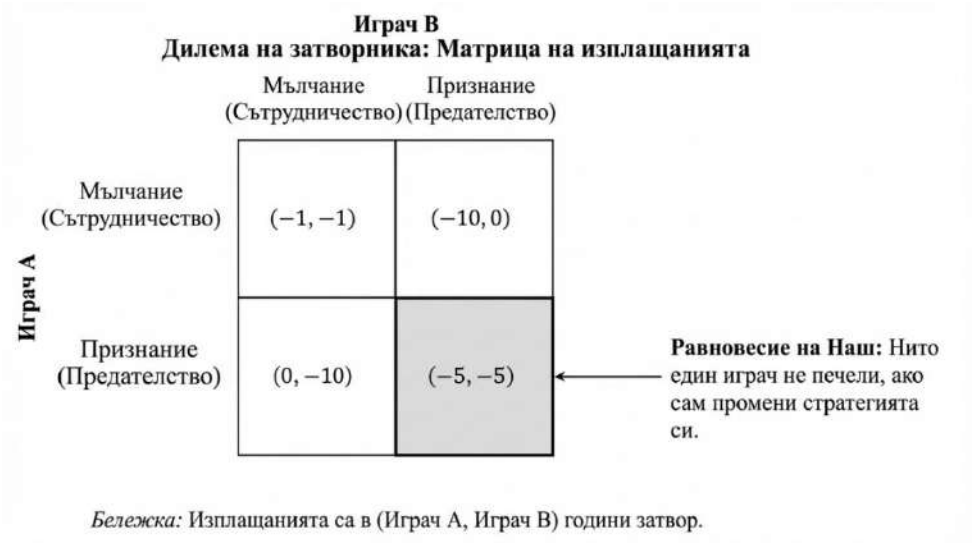
– ако и двамата мълчите (сътрудничество) – и двамата влизате само за по 1 година;

– ако и двамата се предадете един друг (дефекция) – и двамата влизате за по 5 години.

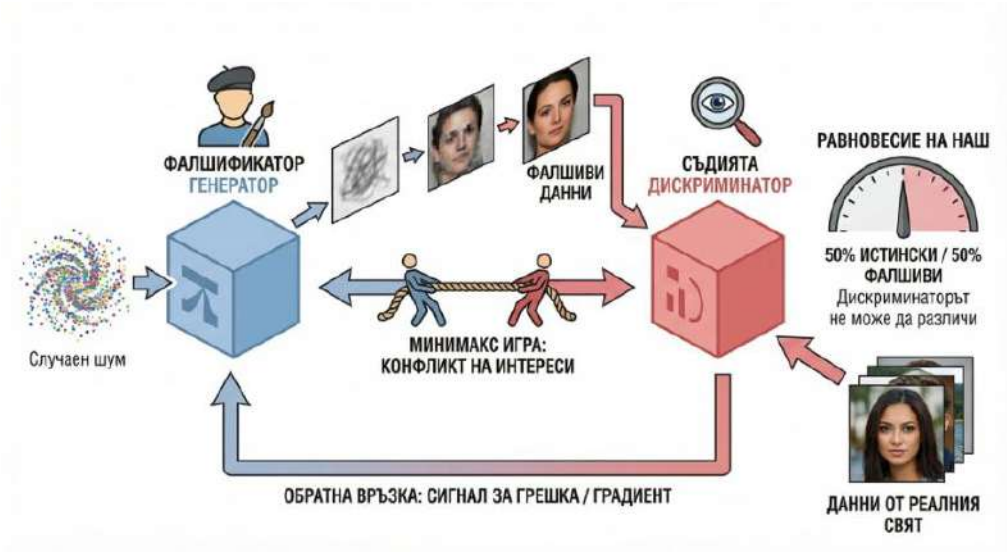
От гледна точка на индивида, най-безопасно е да предаде другия (за да избегне 10-те години). Тъй като и двамата мислят така, те се предават взаимно и получават по 5 години. Ако бяха се доверили и мълчали, щяха да получат само по 1 година. *Извод*: В егоистична среда рационалните индивиди често избират стратегия, която води до по-лош общ резултат. Това може да се играе, докато се постигне Равновесието на Наш. Това е състоението, при което никой от играчите няма стимул (изгода) да промени своята стратегия, докато другият държи собствената си непроменена.

Разновидност на тази игра е минимакс играта, когато имаме стратегия за вземане на решения в игри с двама играчи и нулева сума (където печалбата на единия е равна на загубата на другия). В контекста на Генеративните състезателни мрежи (GANs) дискриминаторът се опитва да максимизира познаването на истината, а генераторът се опитва да направи така, че успехът на дискриминатора да е минимален (да го измами). Това „дърпане на въже“ е чиста минимаксна игра, която води системата към перфектни резултати.

Представете си, че имате двама фалшификатори, които се надпреварват. Това е в основата на GANs. Генераторът създава данни (например снимки на лица), които изглеждат напълно истински, въпреки че не са реални. дискриминаторът гледа снимки и определя: „Това е истински човек“ или „Това е фалшификат, направен от генератора“. Първоначално генераторът създава просто шум и драсканици. Дискриминаторът го сравнява с истински данни. Ако познае, че е фалшификат, той дава „обратна връзка“ на генератора. Генераторът се учи от грешката си и следващия път прави по-добро копие. Това продължава, докато генераторът стане толкова добър, че дискриминаторът започне да гадае напосоки (50/50 шанс). Това е равновесието на Наш, при което вече имаме почти истинска снимка.



Фигура 30. Теория на игрите



Фигура 31. Работа на GAN

В контекста на военното дело и киберсигурността, GAN се използват за създаване на реалистични симулации и генериране на данни за обучение

на алгоритми за разпознаване на цели. Имат широка употреба за противодействие на дезинформацията, разработване на модели за откриване на deepfakes. Чрез обучение на дискриминатори върху генерирано съдържание се подобрява способността за идентифициране на манипулирани видео- и аудиозаписи.

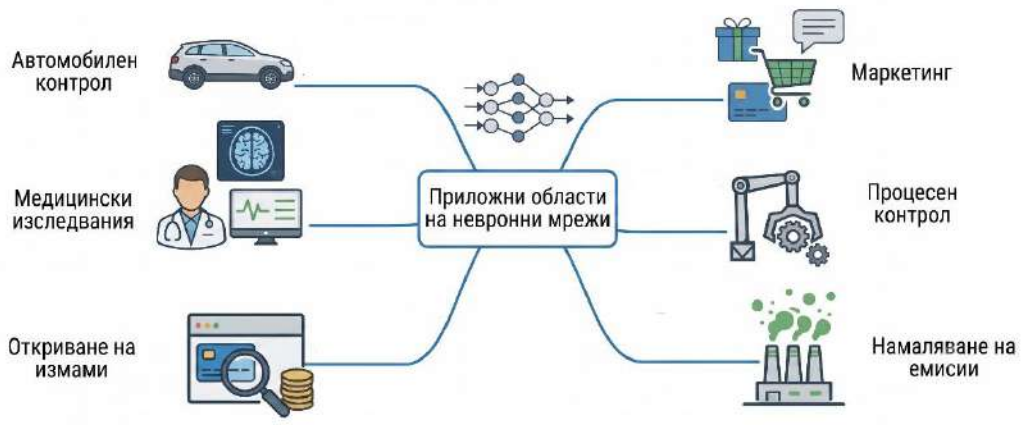
Теорията на игрите може да се използва и във военното дело, например за анализ на стратегии при надпреварата на държавите във въоръжението. Две държави биха били по-добре, ако харчат пари за болници вместо за ракети (това е състояние на сътрудничество). Ако едната се въоръжава, другата е длъжна да го прави, за да не бъде уязвима. Така Равновесието на Наш е и двете да харчат милиарди за оръжия – по-лош резултат за всички, но математически „стабилен“. Такъв е примерът с ядреното сдържане по време на Студената война, при което никой не печели.

3.4.4. Области на приложение на дълбоките невронни мрежи

Мрежите за дълбоко обучение са основната технология, която позволява на ИИ да общува с хората чрез обработка на изображения, разпознаване на звуков говор и анализ на естествен език. Широкият спектър от възможности за приложение се основава на уникалната способност за автоматично генериране на йерархия от свойства с нарастваща сложност. Многобройните възможности за избор на различни архитектури допринасят за успеха.

Недостатъци на дълбоките невронни мрежи са, че те се разглеждат като „черна кутия“, а някои дълбоки невронни мрежи имат повече от 100 слоя и 40 млн. параметъра. Разработването на такъв модел изисква много време дори на най-мощния хардуер. Те се нуждаят от големи масиви от данни за обучение, понякога от порядъка на милиони записи. Области на приложение на дълбокото обучение са в обработката на естествен език (NLP) – онлайн търговия на дребно, търсачки, интерфейси човек – машина, автоматични преводи, разпознаване на лица, медицинска диагностика, медии и развлечения.

Съвременният изкуствен интелект оказва огромно въздействие в три ключови области – на науката, бизнеса и обществото. Този факт трябва да се разбере много добре, за да сме подготвени да се възползваме или да се предпазим от новата реалност. Изнесените факти категорично посочват бъдещата тенденция, особено към дигитална икономика, основана на изкуствения интелект.



Фигура 32. Приложни области на невронните мрежи

Заедно с вълната на внедряване на изкуствения интелект в обществото се очертават три важни области, където човешкият интелект е най-уязвим за инвазията на технологиите – етиката на изкуствения интелект, управлението на изкуствения интелект и дълбоката фалшива реалност, генерирана от него. В списъка на най-важните етични проблеми и страхове фигурират въпросите за потенциалната масова безработица, породена от поголовната автоматизация, контрола върху автономните системи, всеобхватните възможности за следене на индивида, свръхинтелигентните оръжия и др. Друг критичен въпрос е контролът върху системите на изкуствения интелект. Той ще става все по-актуален в бъдеще, ако се доближим до обобщения изкуствен интелект.

Съвременните направления в изследванията са дълбокото подсилено обучение, когнитивните модели и невронните системи. Изкуственият интелект демонстрира огромен ръст в почти всяка област на приложение в индустрията и много агресивно прониква в обществото. В крайна сметка изкуственият интелект е от изключителна важност и значение за всички компоненти на нашия живот.

ОПТИМИЗАЦИОННИ АЛГОРИТМИ В ИЗКУСТВЕНИЯ ИНТЕЛЕКТ

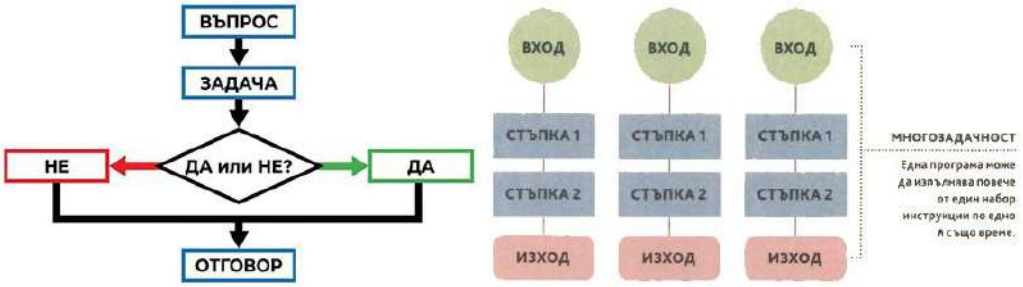
Аз знам, че нищо не знам
Сократ

В света на изкуствения интелект (ИИ) думата *автомат* се отнася до компютър, който може да бъде проектиран да решава конкретна задача, например да дава прогнози за борсови пазари. Андроидът е автомат, който е създаден така, че да имитира човешко поведение. Първоначално тестът на Тюринг се фокусира върху цифровата интелигентност, но съвременните учени смятат, че ИИ трябва да демонстрира всеки един от осемте вида интелигентност – речева, артистична, цифрова, сензорна, пространствена, емоционална, физическа и рефлективна (Gardner,1983).

Идеята, че всяко мислене – било то човешко или изкуствено – е форма на изчисляване, по-конкретно процес по използване на алгоритми за превръщане на символна информация на входа в символна информация на изхода, се нарича „изчислителна теория на ума“ (Putnam 1960; Fodor 1975). Нейните поддръжници твърдят, че човешкият мозък е компютър и поради тази причина един ден изкуственият интелект ще е способен да прави всичко, което може и човешкият мозък.

Алгоритъмът е поредица от инструкции за решаване на задача. В него има вход (входяща информация или данни) и процеси. Обработката се състои от поредица от стъпки за постигане на желан резултат или изход (Knuth 1968). Може да бъде всичко – от обикновено изчисление, през следване на готварска рецепта, до решаване на сложни математически уравнения. Тя е пример за това, което математиците наричат „ефективен метод“, което означава, че се състои от краен брой стъпки и създава конкретен отговор (или изход).

Изчислението следва стъпките на алгоритъма. Най-елементарният пример е аритметичното изчисление. Например, ако събирате две трицифрени числа наум, вие следвате поредица от стъпки – алгоритъм – за да получите резултат. Изчисленията използват символи, за да изобразят числа, но със символи може да се изобрази почти всичко. С правилните символи и алгоритми стават възможни изключително сложни изчисления.



Фигура 33. Алгоритъм

През 19 век сложната работа по изготвянето на числовите таблици (за нуждите на навигацията, военното дело и др.) се възлага на хора, наричани изчислители (на англ. „компютри“). За да се избегнат човешките грешки, английският математик Чарлз Бабидж (1791 – 1871) изобретява „диференчната машина“, която може автоматично да извършва математически изчисления. След това създава проект за „аналитичната машина“ – калкулатор за обща употреба, който може да се програмира с перфорирани карти и има отделни компоненти, свързани с паметта и с изчисленията (Babbage 1837). „Аналитичната машина“, която Бабидж така и не изгражда на практика, поставя теоретичните основи на модерните компютри.

През 1936 г. английският математик Алън Тюринг (1912 – 1954) предлага въображаема машина, способна да реши всяка задача, която може да се сведе до „изчислим вид“ (Turing 1936). С други думи, ако задачата се напише в символен вид, с помощта на алгоритми и преведена в бинарен код, хипотетичната машина би могла да я реши. Устройството се състои от глава, която се движи над лента, маркирана с бинарна информация. Макар и никога да не е конструирана, „универсалната машина“ на Тюринг слага началото на компютърната революция, като доказва, че едно-единствено устройство може да реши всяка задача, която има изчислим вид. След това различните решения за изчислителни машини на база на теорията на Тюринг са били предимно механични, докато българинът Джон Атанасов, заедно със своя асистент Клифърд Бери, създава първия електронен цифров компютър – ABC, конструиран между 1939 и 1942 г.

Маккълок и Питс (1943) демонстрират, че мрежа от елементи, наподобяващи човешките нервни клетки (невроните), по които протичат електрически сигнали, може да имитира машината на Тюринг. Те допускат, че мозъкът може би е вид жив компютър, което означава, че един

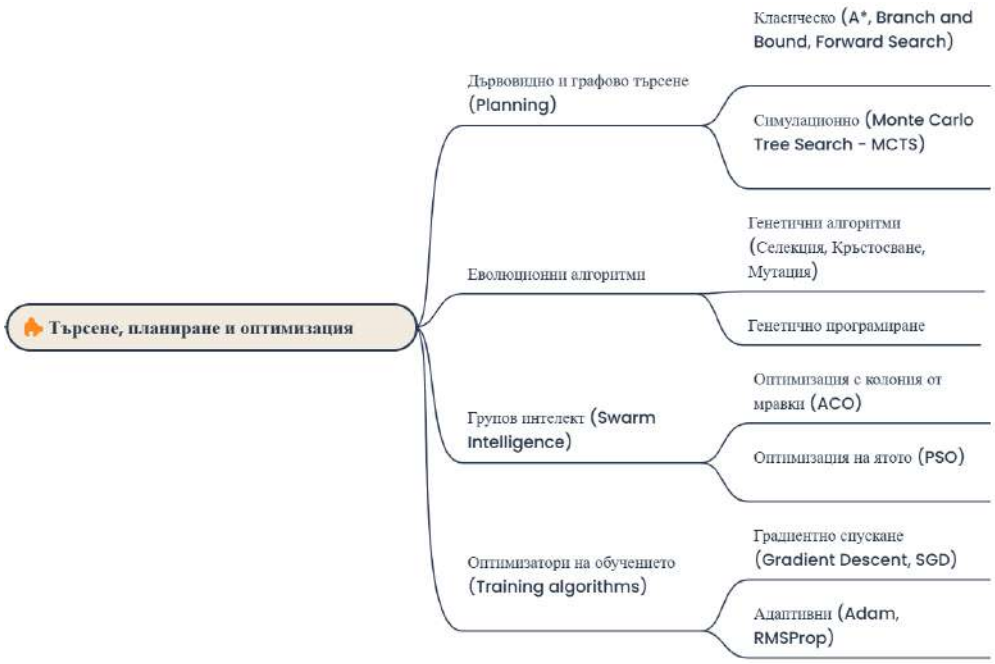
„електрически мозък“ може да изпълнява същите програми като човешкия. Тази теория е известна и под името „множествена реализируемост“. Съвременната наука за машинното обучение е разработила многообразие от алгоритмични методи, които могат да бъдат систематизирани в десет основни семейства според функционалните характеристики и математическите основи (Mitchell 1997; Alpaydin 2010; Flach 2012; Bishop 2006):

Таблица 4. Алгоритмични методи

№	*Семейство*	*Описание*	*Ключови методи*
1.	Регресионни методи	Предсказване на непрекъснати числени стойности	Линейна регресия, полиномна регресия, метод на най-малките квадрати
2.	Методи, базирани на инстанции	Съхраняване на обучителни примери и сравнение чрез мерки за сходство	k-Nearest Neighbors (k-NN), Locally Weighted Learning
3.	Методи за регуляризация	Разширения на регресионните подходи с ограничения върху сложността	Ridge Regression, LASSO, Elastic Net
4.	Методи, базирани на дървета	Рекурсивно разделяне на пространството на признаците	CART, ID3, C4.5, техники за подрязване
5.	Байесовски методи	Явно прилагане на теоремата на Бейес	Наивен Бейес, Bayesian Belief Networks, максимална правдоподобност

№	*Семейство*	*Описание*	*Ключови методи*
6.	Ядрени методи	Нелинейни трансформации във високомерни пространства	Support Vector Machines (SVM) с различни ядрови функции
7.	Клъстерни методи	Автоматично групиране на данни	K-Means, йерархично клъстериране, Expectation-Maximization (EM)
8.	Асоциативно-правилно обучение	Откриване на правила, описващи големи части от данните	Apriori, Eclat
9.	Намаляване на размерността	Компресиране на информацията	Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), автоенкодери
10.	Ансамблови методи и невронни мрежи	Комбиниране на няколко модела	Bagging, boosting, random forests, многослойни перцептрони, конволюционни и рекурентни невронни мрежи

Разбирането на оптимизационните алгоритми и механизмите на обучение е първата стъпка към ефективното внедряване на интелигентни системи. Повечето от тях се използват за търсене, планиране и оптимизация и затова. Във фигура 34 е предоставена класификация на този вид алгоритми.



Фигура 34. Класификация на алгоритми за търсене, планиране и оптимизация

4.1. Алгоритми за групиране без учител (K-means, Йерархична клъстеризация)

Клъстеризацията е метод от обучението без учител, чиято цел е групирането на обекти в множества (клъстери) по такъв начин, че обектите в един и същи клъстер да бъдат по-подобни помежду си, отколкото на тези в други клъстери. В контекста на националната сигурност и киберсигурността, този метод се прилага за откриване на аномалии, класификация на зловреден софтуер или анализ на дезинформационни кампании.

В областта на машинното обучение, **клъстеризацията** представлява процес на групиране на обекти по такъв начин, че тези в една и съща група (наречена клъстер) да са по-сходни помежду си, отколкото с тези в други групи. Този метод се класифицира като **неконтролирано обучение**, тъй като алгоритъмът работи с немаркирани данни и не разполага с предварително

дефинирана зависима променлива. Основната цел на всеки алгоритъм за клъстеризация е едновременното постигане на две условия:

– максимизиране на вътрешнокласовото сходство;

– минимизиране на междукласовото сходство, което математически се изразява чрез максимизиране на разстоянието между центровете на клъстерите.

За измерване на сходството най-често се използва Евклидово разстояние между две точки p и q в n-мерно пространство:

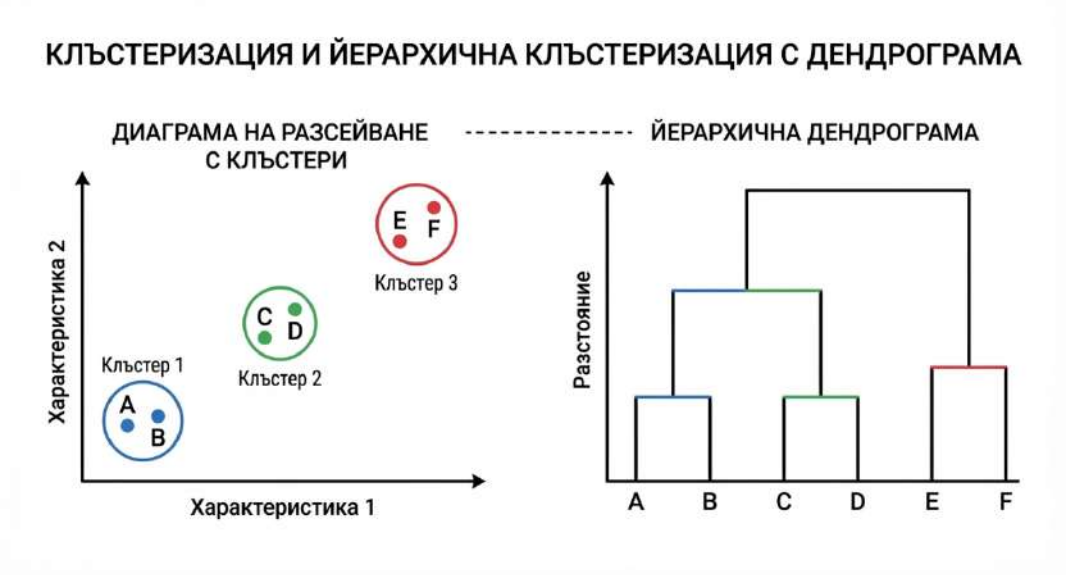
$$d(p,q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$$

Разлика между класификация и клъстеризация. Класификацията и клъстеризацията звучат сходно, но са два различни метода на изкуствения интелект. Класификацията се използва при контролирано обучение (когато имаме зависима променлива и маркирани данни), докато клъстеризацията се използва при неконтролирано обучение, когато нямаме никакви познания за зависимата променлива и данните не са маркирани (Mitchell 1997). Тъй като задачата за клъстеризация е субективна, средствата, които могат да се използват за постигането ѝ, са много. Всяка методология следва различни правила за определяне на метриката на сходство между точките с данни. Известни са повече от 100 алгоритъма за клъстеризация, но в повечето бизнес проблеми се прилагат два – йерархична и k-Means клъстеризация (Bishop 2006).

Както подсказва името, **йерархичната клъстеризация** е алгоритъм, изграждащ йерархия от клъстери, която може да бъде представена чрез дървовидна структура. Резултатите от йерархичната клъстеризация могат да бъдат показани с помощта на дендрограма. Дендрограмата може да се интерпретира, като започваме в основата с n точки данни, всяка от които е разпределена в отделни клъстери. Най-близките клъстери се обединяват, докато се получи само един клъстер в горната част. Височината в дендрограмата, при която се сливат два клъстера, представлява разстоянието между два клъстера в пространството на данните.

Алгоритъмът започва, разглеждайки всяка точка като отделен клъстер. На всяка стъпка се сливат двата най-близки клъстера, докато всички точки не бъдат обединени в един общ клъстер. Височината на сливане в дендрограмата съответства на разстоянието между клъстерите. Потребителят

може да „отреже“ дендрограмата на определено ниво, за да получи желания брой групи.



Фигура 35. Пример за клъстеризация и дендограма на йерархична клъстеризация

Алгоритъм K-means е итеративен алгоритъм, който разделя набора от данни на K на брой предварително дефинирани клъстера. Целевата функция, която алгоритъмът се опитва да минимизира, е сумата от квадратите на разстоянията в рамките на клъстерите (Within-Cluster Sum of Squares – WCSS):

$$J = \sum_{j=1}^{K} \sum_{x \in S_j} \| x - \mu_j \|^2,$$

където K е броят на клъстерите;

S_j е множеството от точки в j-тия клъстер;

μ_j е **центроидът** (средната стойност) на точките в S_j.

Алгоритмични стъпки са:

– инициализация на K случайни центроида;

– причисляване на всяка точка към най-близкия центроид;

– преизчисляване на позициите на центроидите като средноаритметично на всички точки в съответния клъстер;

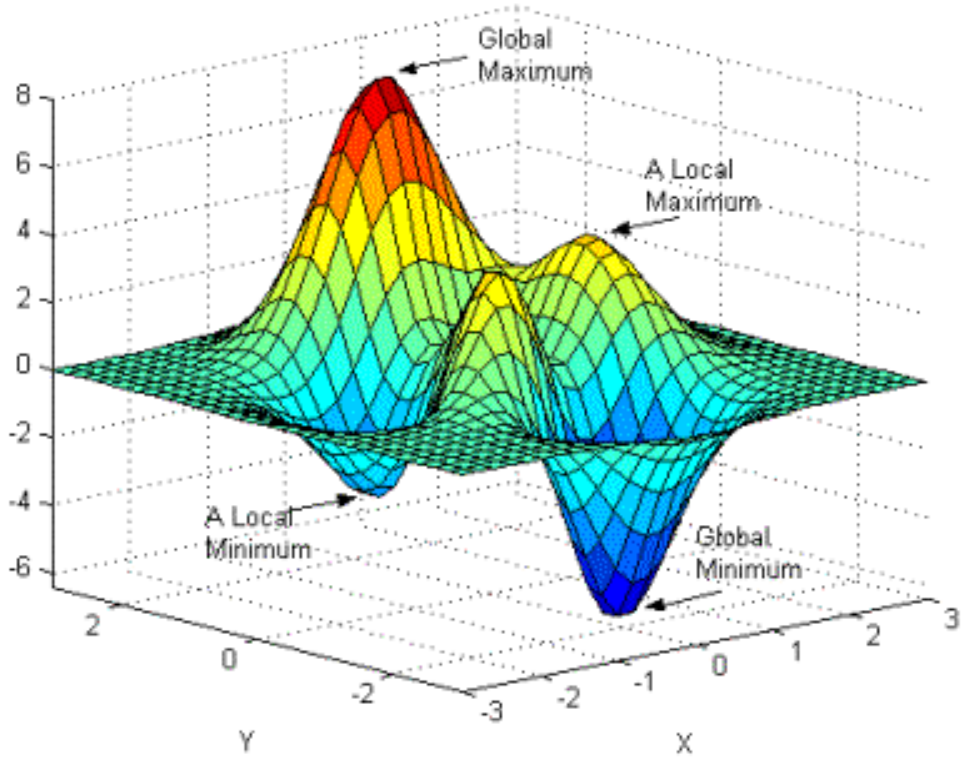
– повторение на процеса до конвергенция (когато позициите на центроидите спрат да се променят).

Целта на K-means е да се намерят групи в данните, като броят на групите се представя от променливата *K* (брой на клъстерите), определена от потребителя. Алгоритъмът работи итеративно, за да причисли всяка точка от данни към една от *K* групите въз основа на предоставените характеристики. Точките данни се групират въз основа на сходството на признаците (Alpaydin 2010). Започва се с избиране на желания брой клъстери. Алгоритъмът K-means инициализира процеса със случайни центрове в данните и след това се опитва да прикрепи най-близките точки към тези центрове. Предимство на този алгоритъм е, че центровете на *K* клъстерите могат да се използват за маркиране на нови данни.

Ще илюстрираме приложението на K-means клъстеризацията с интересно бизнес приложение за анализ на причините за възникването на икономически проблеми при резки и значителни промени на икономиката. Предприятие от химическата промишленост иска да разбере как различните икономически фактори са повлияли на ефективността на бизнеса преди и след голямата рецесия през 2008 г. Хипотезата е, че двете различни икономически състояния преди и след рецесията ще бъдат представени в различни клъстери.

Събирането на данни включва 42 макро-, микро- и свързани с бизнеса фактори за няколко тримесечия около рецесията. Резултатите от прилагането на K-means показват, че двата клъстера са ясно дефинирани, без да има припокриване между тях, т.е. има ясно разграничение между факторите при двете икономически състояния. Важно е също, че тримесечните данни по време на рецесията, представляващи прехода между двете равновесни състояния в икономиката, са извън клъстерите. Това означава, че сравнението и анализът на поведението на бизнеса се основават на икономическото равновесие преди и след рецесията.

Оптимизацията като математическа дисциплина може да ни помогне да намерим минимум или максимум на дадена функция. Има два вида оптимума – локални (в определен район на пространството за търсене) и глобални (в цялото пространство за търсене). В повечето реални ситуации обаче не е лесно да се разграничи локалното от глобалното оптимално решение.



Фигура 36. Визуализация на глобални и локални минимуми и максимуми

Проблеми на класическата оптимизация. В класическия тип оптимизация се опитваме да намерим оптимума на дадена функция въз основа на нейните частни производни. Първите частни производни, съставящи градиента, помагат да определим посоката на най-близкия локален минимум. Вторите частни производни, съставящи т.нар. *матрица на Хесен* (Хесиан), ни помагат да определим позицията на локалния минимум. Всичко това работи, ако можем да изчислим тези частни производни. Ако не можем, се изправяме пред голям проблем. За съжаление това се случва често в реални ситуации. Тези числени методи работят много добре само ако разполагаме с т.нар. гладко пространство за търсене. Ситуацията е коренно различна при търсене в негладкото, шумно пространство с много оптимуми. Преди всичко не е лесно да се направи разграничение между глобалния оптимум и локалните оптимуми. От изчислителна гледна точка пространство за търсене не

позволява да изчислим надеждно частните производни, т.е. губим основния навигатор към оптимума. Заради това сме принудени да търсим други алгоритми, небазирани на производни.

Изкуственият интелект разполага с три основни метода, които ни позволяват да прилагаме оптимизация в области, където класическата оптимизация обикновено се затруднява да даде решения (Mitchell 1997):

– генетични алгоритми (Genetic Algorithms – GA);

– оптимизация с колония от мравки (Ant Colony Optimization – ACO);

– оптимизация с интелигентност на ятото (Particle Swarm Optimization – PSO).

4.2. Еволюционни алгоритми

Еволюционните алгоритми представляват разнообразие от методи, вдъхновени от идеята на Дарвин, че процесите на еволюция могат да бъдат формализирани и приложени като алгоритми в изкуствена среда (Darwin 1859; Holland 1975). Ключовата концепция е симулацията на естествен отбор и наследственост, при която популация от потенциални решения на даден проблем претърпява селекция, кръстосване и мутация във виртуалното пространство на компютъра. Така с всяко поколение еволюират и се усъвършенстват решения – независимо дали става дума за уравнения, електронни схеми, механични конструкции или алгоритми за намиране на оптимални пътища.

Този подход обхваща различни техники и парадигми, сред които генетичните алгоритми, генетичното програмиране, оптимизацията с интелигентност на ятото (Particle Swarm Optimization) и мравешките колонии (Ant Colony Optimization), които ще бъдат разгледани по-детайлно. Всеки от тези методи прилага в различна степен принципите на еволюцията, за да намери ефективни и адаптивни решения на сложни проблеми в разнообразни области на науката и техниката.

4.2.1. Генетични алгоритми – популация, селекция, кръстосване и мутация

Генетичните алгоритми представляват категория методи за търсене и оптимизация, вдъхновени от законите на биологията и представени в абстрактна математическа форма. В контекста на отбраната, те се използват за

решаване на сложни задачи, при които традиционните математически методи са неефективни – от проектиране на аеродинамични форми до оптимизиране на логистични вериги. Една от важните характеристики на еволюционните алгоритми е, че вместо да се работи с едно решение в пространството за търсене, се разглежда едновременно голяма популация от множество решения („индивиди“).

За да се разбере работата на генетичния алгоритъм, е необходимо дефинирането на ключови понятия, които правят паралел между биологията и инженерната оптимизация:

– популация – множество от възможни решения (индивиди), които се разглеждат едновременно;

– генотип (хромозома) – вътрешното представяне на решението (обикновено като цифров код);

– фенотип – физическата реализация на решението (например конкретна форма на крило на дрон или параметри на двигател);

– фитнес функция (оценка на жизнеспособността) – математически критерий, който определя колко „добро“ е дадено решение спрямо поставената цел.

Всяка генна структура в един генетичен алгоритъм може да бъде представена като популация от хромозоми, всяка от които е проектирана като набор от гени. На тази биологична основа стъпва идеята за определяне на пространството за търсене за оптимизация на базата на генотиповете на хромозомите от пространството на решенията, където се намира физическият обект на процеса на оптимизация.

С помощта на генетичния алгоритъм се търсят най-добрите от възможните решения. Процесът прилича на симулирана еволюция на конкуриращи се структури в компютъра. На по-добрите решения се позволява да имат „деца“, а по-лошите се елиминират. Новите решения наследяват характеристиките на своите „родители“ с някои случайни вариации и след това на по-добрите от тях се позволява също да имат потомство, докато по-лошите „умират“. След определен брой поколения компютърът съдържа решения, които са значително по-добри от отдавна елиминираните си предци.

Процесът на намиране на оптимално решение преминава през няколко последователни фази:

1. Инициализация. Алгоритъмът стартира с множество потенциални решения, събрани в една „популация“. Компютърът генерира начална популация от случайни решения. Например, ако проектираме свръхзвукова ракета и искаме да определим нейните физически характеристики, като размер

на крилата, тежест и други, ние взимаме формулата за закона за движение на ракета и някои параметри оставяме плаващи, като искаме алгоритъмът да определи техните размери, за да получим най-висока скорост.

2. Оценяване. Всяко решение на това уравнение се подлага на тест чрез фитнес функцията (в нашия случай – най-високата скорост). Слабите решения (поколения) получават ниска оценка, а тези, които се доближават до търсеното решение – висока.

3. Селекция. Следва процес на естествен подбор, при който на решенията с висока жизнеспособност се дава възможност да имат „наследници“. На по-добрите решения се дава по-голям шанс да станат „родители“ и да прехвърлят своите характеристики в следващото поколение. Слабите решения се елиминират и уравнението се решава отново.

4. Генетични оператори. Това е същността на еволюционния подход, където се прилагат генетични вариации. На всяко решено уравнение се оставят параметрите с получена най-добра скорост и останалите се кръстосват и мутират отново и отново:

– кръстосване – оператор, при който избраните участъци от хромозомите на родителите се разменят в потомството. Тази обмяна на информация между две „родителски“ решения създава „потомство“, съчетаващо техните силни страни;

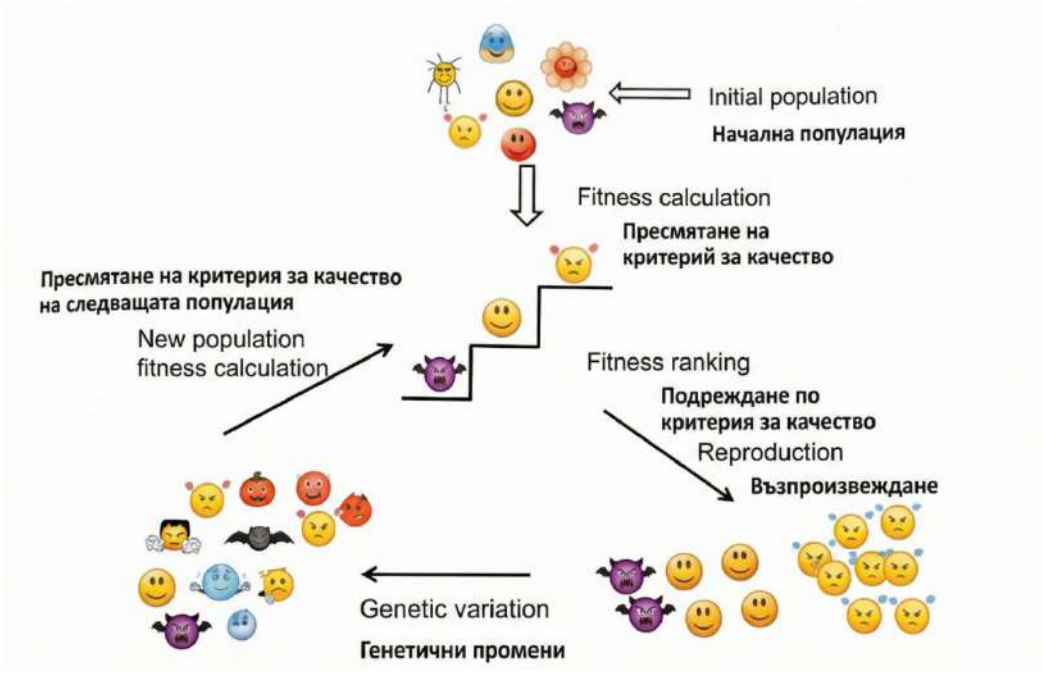
– мутация – оператор, който променя малка част от хромозомите. Това е случайна промяна в характеристиките на новото решение, използвана главно за поддържане на разнообразието от структури и предпазване на алгоритъма от попадане в еднотипна популация (стагнация).

5. Итерация. Породеното поколение еволюира и се образува ново множество потенциални решения, а старото поколение се унищожава. Този процес (генерация на поколение) се повтаря многократно, при което решенията стават все по-добри до намиране на търсения екстремум или докато се изчерпи времето за изчисления.

В резултат на ключовите генетични операции средното качество на цялата популация и най-вече на най-добрите видове по време на симулираната еволюция постепенно се увеличава и в крайна сметка алгоритъмът ни казва, че за да получим най-добрата скорост на движение, изследваните параметри на ракетата следва да бъдат с дадени x размери. За изясняване работата на алгоритъма, ще разгледаме още един условен пример за избор на компютърни икони, използвани в нова програма, които трябва да се харесват на потребителите. В началото на алгоритъма тази популация е от случайни видове иконки.



Фигура 37. Последователност на генетичен алгоритъм



Фигура 38. Пример за работата на генетичен алгоритъм.[14]

● Стъпка 1 (ранкиране). Първата стъпка е те да бъдат класифицирани според критерий за качество – например „грабване на внимание“. В резултат иконката на жълтото човече има най-висок ранг, следва кафявото човече и т.н.

● Стъпка 2 (възпроизвеждане). Даваме по-големи права за възпроизвеждане на иконките с по-висок ранг.

● Стъпка 3 (вариации). Прилагаме генетичните операции кръстосване и мутация. В резултат се създават частично нови иконки към текущата популация.

Преминаваме към следващия цикъл на симулираната еволюция с пресмятане на критерия за качество на новата популация. В следващото ранкиране очакваме иконките, които грабват повече внимание, да имат по-силно присъствие поради конкурентното си предимство. Еволюционният алгоритъм е безкраен цикъл, където някои от тези иконки оцеляват, други се

[14] Източник: КОРДОН, А., 2023. *Перспективата изкуствен интелект*. София: Изток- Запад. ISBN 978-619-01-1251-8.

елиминират и непрекъснато се модифицират. В резултат цялостната популация повишава своето качество. Най-популярните два метода на този вид оптимизация са:

● Генетични алгоритми (Genetic Algorithms – GA). Използват се предимно за оптимизация на параметри (напр. намиране на най-добрите размери на даден детайл). Това е оптимизационен метод, базиран на еволюционни принципи. Първата важна стъпка е подреждане на популацията от случайни дигитални обекти (образи, уравнения, схеми и т.н.) по избран критерий за качество. С всеки следващ кръг на еволюцията сложността на популацията нараства и се появяват членове с по-високо качество.

● Генетично програмиране (Genetic Programming – GP). Използва се за предсказване и генериране на нови структури. Вместо да търси числа, GP създава алгоритми или програмни кодове, които решават специфични задачи. Фокусът е върху генерирането на нови структури и взаимоотношения, основани на конкуренцията. В случая с генетичното програмиране това са уравнения, които са оптимални по отношение на грешката и са възможно най-прости. Точността и сложността са определящите критерии за оцеляване, затова и изборът на нови решения е фокусиран върху т.нар. Парето фронт.

В генетичното програмиране и оптимизацията често се сблъскваме със ситуации, в които няколко цели едновременно (многоцелева оптимизация), които си противоречат. Например, искаме дронът да бъде едновременно лек (за икономия на гориво) и здрав (за издръжливост). Няма едноединствено „перфектно“ решение, което да е най-добро и по двата показателя. Тук идва понятието Парето фронт (кръстен на италианския инженер Вилфредо Парето).

Парето фронтът е множеството от всички онези решения, които са „ненадминати“ (недоминирани). Това означава, че за никое от тези решения не съществува друго решение, което да е по-добро във всички показатели едновременно. Представете си, че оптимизираме стол чрез генетичен алгоритъм с две цели – максимална здравина при минимална цена.

– Решение А: Много здрав, но много скъп;

– Решение Б: Евтин, но крехък;

– Решение В: Средна здравина, средна цена;

– Решение Г: По-слабо от А и по-скъпо от А. (Това е лошо решение, то не е в Парето фронта), но решение А, Б и В са част от Парето фронта, защото не можем да кажем кое от тях е категорично по-добро – едното е по-добро по цена, а другото по здравина.

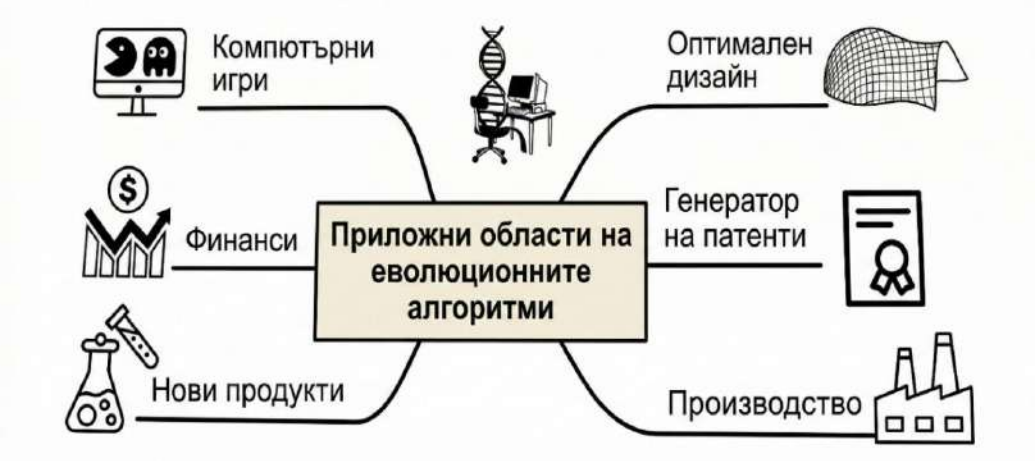
В генетичното програмиране (GP), алгоритъмът търси уравнения, които са баланс между точност (да предсказват добре) и простота (да са кратки и разбираеми). Парето фронтът показва на инженера най-добрите компромиси, от които той може да избира спрямо конкретните нужди на мисията.

Въпреки че и двата метода използват едни и същи еволюционни принципи (популация, кръстосване, мутация, селекция), тяхната структура и цел са коренно различни. Генетичният алгоритъм работи по принцип с низ от числа, за да намери точни пропорции, а генетичното програмиране работи с динамични структури. Решението не е просто списък от числа, а компютърна програма, формула или дърво на решение. Дължината и сложността на структурата могат да се променят (растат или намаляват) по време на еволюцията.

Уникално свойство на еволюционните алгоритми е техният неограничен потенциал да генерират нови решения за почти всеки практически проблем. Те могат да генерират модели с оптимален баланс между точност и сложност. Това е най-добрият сценарий за крайния потребител, тъй като приложените емпирични модели могат да бъдат интерпретирани и при малки промени в производствените процеси или бизнес условията.

Техните предимства са, че са напълно автономни и не изискват човешка намеса. Разходите за разработване са минимални, тъй като надеждният модел се генерира автоматично. За разлика от класическите методи, не се „заклещват“ лесно в локални, неоптимални решения. Могат да оптимизират системи с хиляди променливи едновременно. Получените емпирични решения са явни математически функции, които могат да бъдат реализирани във всяка софтуерна среда, включително Excel. И не на последно място – имат ниски разходи за внедряване и поддръжка.

Недостатъците им се състоят в това, че симулираната еволюция изисква значителна изчислителна мощност. Непрекъснатото нарастване на изчислителната мощност съгласно Закона на Мур постепенно решава този проблем. Необходимо е продължително време за генериране на решение – неизбежен ефект от интензивното изчисление. Процесът е по-бавен в сравнение с повечето методи на изкуствения интелект. Не са много ефективни, когато пространството за търсене е изключително голямо. За предпочитане е броят на променливите да бъде ограничен до 1000, а броят на записите – до 1 млн., за да се получат резултати за приемливо време.



Фигура 39. Приложни области на еволюционните алгоритми

Бизнес потенциалът на еволюционните алгоритми в областта на инженерното проектиране беше бързо осъзнат в началото на 90-те години от компании като General Electric, Rolls Royce и BAE Systems plc. В рамките на няколко години много отрасли, като авиационната, енергийната и химическата промишленост, трансформираха интереса си към еволюционните изчисления в практически решения.

Най-прилаганата област е т.нар. оптимален дизайн. Типичен пример е прилагането на генетичен алгоритъм за предварително проектиране на газови турбини и оптимизиране на геометрията на отворите за охлаждане на турбините в Rolls-Royce. Целта е била да се сведе до минимум разходът на гориво. Чрез симулирана еволюция компютърът оптимизира геометрията, постигайки резултати, които често надминават възможностите на класическото инженерно проектиране.

4.2.2. Групов интелект

Груповият интелект представлява нов начин за оптимизиране и класифициране на сложни системи в реално време чрез имитация на социално взаимодействие в различни биологични общества. Има две основни направления в изследванията на груповия интелект и неговото приложение (Mitchell 1997):

1. Оптимизация по метода на колонията от мравки (Ant Colony Optimization – ACO) – основаваща се на интелигентността на колонията от насекоми;

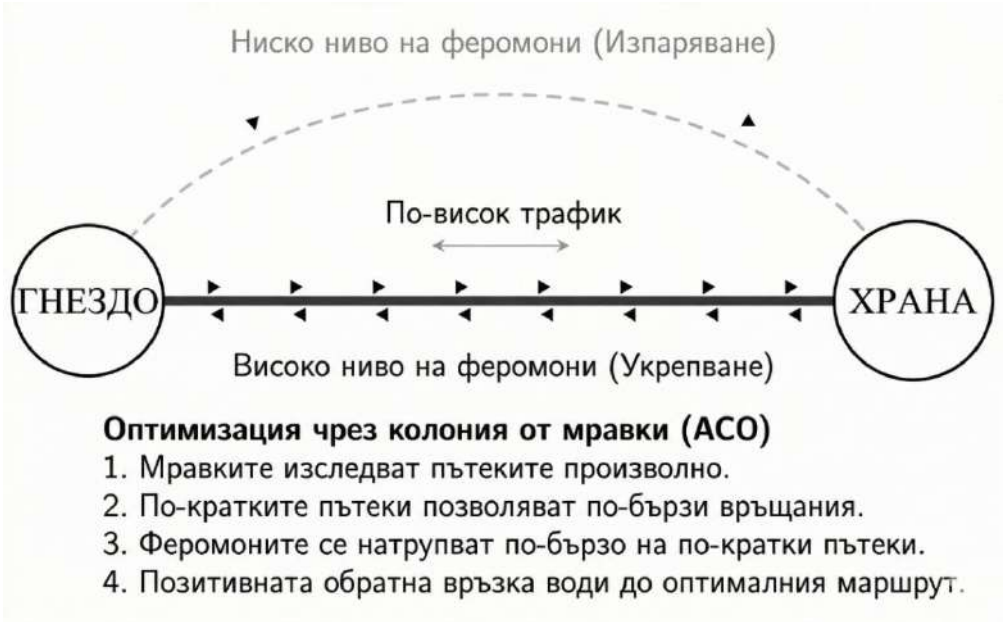
2. Оптимизация на базата на ято (Particle Swarm Optimization – PSO) – основаваща се на социалното взаимодействие в птичите ята или на рибни пасажи.

Оптимизация по метода колония от мравки (ACO)

Как мравките успяват да намерят най-краткия път? Отговорът от биологията е прост – чрез прилагане на механизма на стигмергия – индиректна комуникация, основана на феромон (субстанция, която оставят) по пътя, който следват. Те използват конкуриращи се маршрути, един от които е значително по-кратък от останалите, където непрекъснато подсилват приложения феромон.

В природата, когато една мравка открие източник на храна, тя се връща в гнездото, маркирайки пътя си с феромон. Други мравки, откривайки тази следа, са склонни да я последват, като на свой ред я подсилват. Мравките по късия път обаче ще извършат повече пътувания и по този начин ще оставят повече феромон по него. Концентрацията на феромона върху късия път ще се увеличава с по-висока скорост, отколкото върху дългия път и във фазата на напредналото търсене мравките по дългия път ще го избегнат за сметка на по-късия път. Тъй като повечето мравки вече няма да пътуват по дългия маршрут и тъй като феромонът е летлив, следата по дългия път ще започне да се изпарява. Във финалната фаза на търсене остава само най-краткият маршрут.

Алгоритъмът за оптимизация на принципа на колонията от мравки (Ant Colony Optimization – ACO) представлява мощна метаевристична техника, вдъхновена от тяхното колективното поведение. В основата му заляга концепцията за стигмергия — форма на индиректна координация чрез промени в околната среда.



Фигура 40. Последователност на намиране на най-краткия път между гнездото и храната на колония от мравки

В изкуствените системи всяка „мравка“ е агент, който конструира решение чрез вероятностен преход между възлите в граф. Вероятността P_{ij}^{k} мравката k да се придвижи от възел i към възел j се дефинира чрез следната зависимост:

$$P_{ij}^{k} = \frac{(\tau_{ij})^{\alpha} \cdot (\eta_{ij})^{\beta}}{\sum_{l \in \text{allowed}_k} (\tau_{il})^{\alpha} \cdot (\eta_{il})^{\beta}},$$

където τ_{ij} е интензивността на изкуствения феромон върху дъгата (i, j);

η_{ij} – евристичната стойност (обикновено $\eta_{ij} = 1/d_{ij}$, където d е разстоянието);

α и β – параметри, определящи относителната тежест съответно на феромонната следа и на евристичната информация;

allowed_k – списъкът с възли, които мравката k все още не е посетила (памет на агента).

Процесът на оптимизация преминава през три основни фази:

1. Начално търсене – феромоните са разпределени равномерно или случайно. Мравките изследват пространството на решенията, като се разпределят по различни маршрути.

2. Напреднало търсене – благодарение на натрупания опит вероятностите започват да се изкривяват в полза на по-добрите (по-късите) пътища. Започва процес на експлоатация на информацията.

3. Финална фаза – алгоритъмът постига сходимост. Поради изпарението на феромона по субоптималните пътища те стават статистически пренебрежими и цялата колония се концентрира върху най-доброто намерено решение.

Този подход е изключително ефективен при решаване на комбинаторни задачи, оптимизация на мрежови трафик и логистично планиране в динамична среда. Процесът създава самоусилваща се обратна връзка, която в крайна сметка насочва цялата колония към оптималния маршрут. Псевдокод на алгоритъма:

1. Създаваме колония от n изкуствени мравки по стохастичен начин (със случайни параметри).

2. Позволяваме на тези мравки да полагат динамично феромонови следи по време на работа, за да отразят опита си в търсенето.

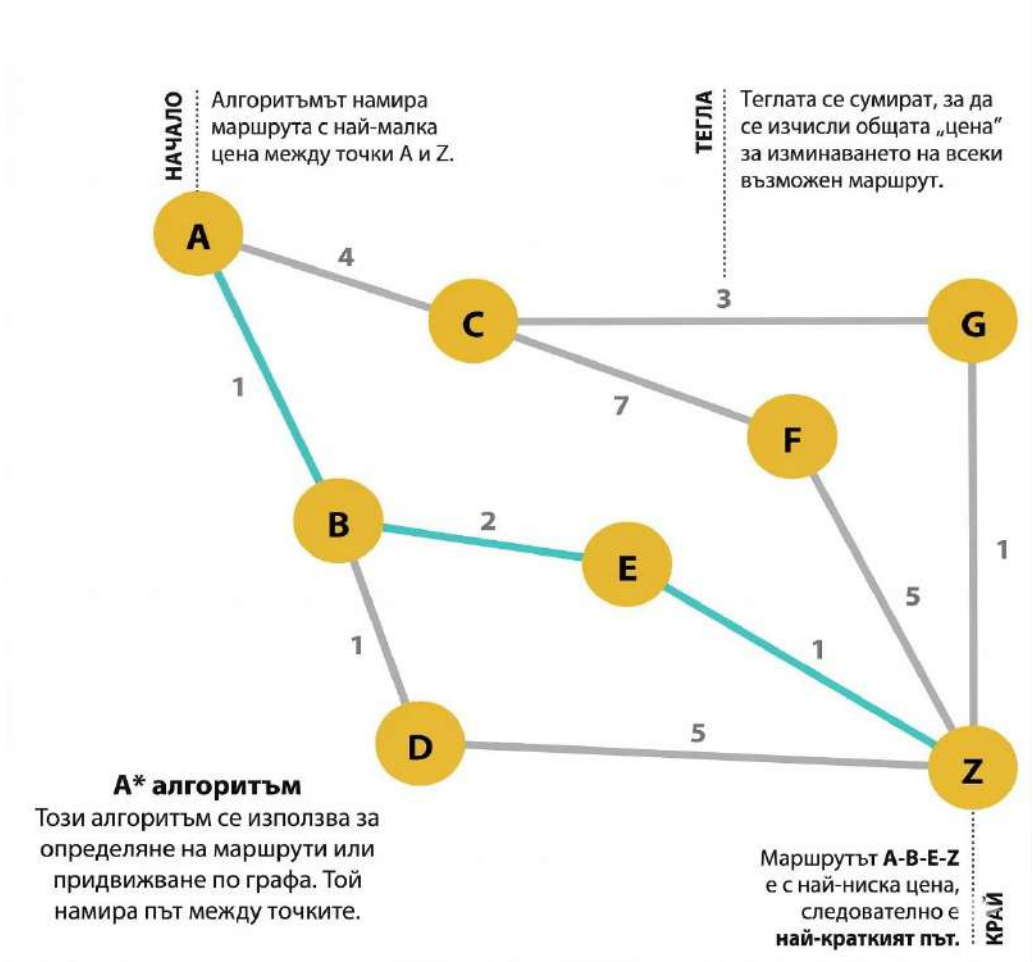
3. Критичният момент в този алгоритъм е как да се вземе решение кой потенциален път да се следва.

Мравката избира следващия възел, към който да се придвижи, като се съобразява с вероятност на феромоновите стойности τ и евристичните стойности η на дъгите, свързващи възел i с възлите j, които все още не е посетила. Посетените възли са запаметени.

Пълният алгоритъм за оптимизация тип колония от мравки се инициализира с популация от цифрови мравки (например 100) със случайни параметри. За всяка мравка се изчислява определен критерий за качество, например разстоянието до някое целево място. В хода на оптимизацията непрекъснато се актуализират стойностите на отложения цифров феромон въз основа на дефинираната евристика. В края на симулацията се разглежда сходимостта на фитнес функцията – например дали пристигаме в определен град за минимално време, минимално разстояние или за минимално количество гориво.

Едно от първите успешни приложения на оптимизацията тип колонии от мравки е намирането на оптимални маршрути за полетите за товарните операции на Southwest Airlines, което им донесе над 10 милиона долара годишна печалба. Освен това обаче, той може да се прилага за разпределяне

на задачите и планиране на работни места. Може да се използва за изгодно планиране на доставките на втечнен газ. Оптимизиране на трафика в телекомуникационните мрежи и др.



Фигура 41. Алгоритъм A* – намиране на най-кратък път

Евристичен модел „Маршрут през Манхатън“ – този модел картографира маршрутите, като изчислява броя на изминатите вертикално и хоризонтално квадрати. Той може да се използва за намиране на път в град с улична система, наподобяваща мрежа, каквато има Манхатън. За да

разберем тази концепция, е необходимо за си представим картата на остров Манхатън в Ню Йорк, където улиците образуват перфектна решетка (мрежа от квадрати). Ако сте на едно кръстовище и искате да стигнете до друго, не можете да минете по диагонал през сградите. Трябва да се движите само хоризонтално (наляво/надясно) и вертикално (нагоре/надолу).

В математиката и програмирането това се нарича манхатънско разстояние. То измерва разстоянието между две точки в мрежа, като събира абсолютната разлика между техните X и Y координати.

Алгоритъмът A* е „умен"“ защото не търси сляпо във всички посоки. Той използва евристика (обозначена като h) – обосновано предположение колко остава до целта.

Ако алгоритъмът A* се използва в среда, където движението по диагонал е забранено (например в игра, където героят се движи само в 4 посоки по плочки), манхатънското разстояние се използва като перфектната евристика. То „подсказва“ на A* накъде да се насочи, за да намери най-краткия път изключително бързо, без да изчислява излишни маршрути.

Оптимизация по примера на ятото (PSO)

При обучението на изкуствен интелект една от главните цели е да се намерят перфектните „настройки“ (тегла и параметри), така че моделът да прави възможно най-точни предвиждания или решения. Пространството от възможни настройки е огромно – подобно на необятен, невидим планински пейзаж, където целта е да се намери най-ниската долина (най-малката грешка) или най-високият връх (най-високата точност).

Най-популярният алгоритъм за това е **оптимизацията чрез рояк частици** (Particle Swarm Optimization – PSO), вдъхновен от социалното поведение на ятата птици или пасажите с риби. Принципът на алгоритъма на оптимизация на ятото е по-сложен за обяснение и се базира на актуализиране на скоростта и местоположението на всяка птица със съответните характеристики на водача. Предполага се, че той най-бързо и точно се насочва към оптимума, като работи в следната последователност:

● Алгоритъмът създава множество виртуални „частици“ (птици) със случайни параметри в пространството. Всяка частица представлява едно възможно решение (една конкретна комбинация от настройки за AI модела) и притежава два основни вектора – вектор на позицията x_i и вектор на скоростта v_i.

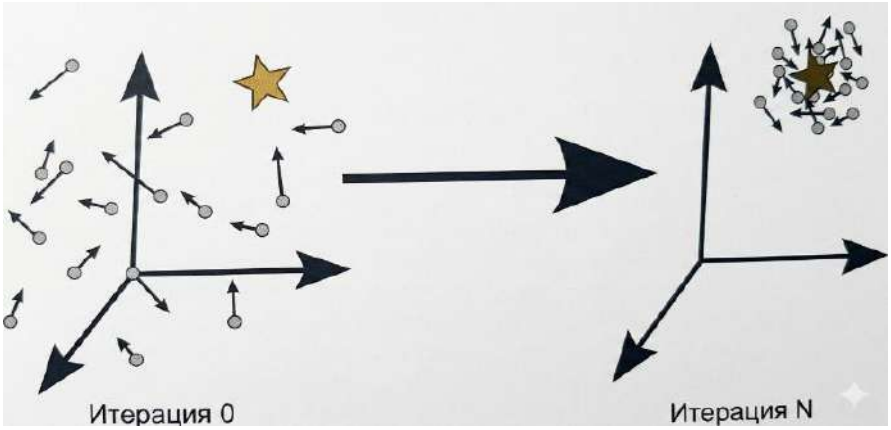
● Тези частици биват „пуснати“ да летят из пространството от решения. Движението на частиците е насочено към търсене на глобален оптимум чрез итеративно актуализиране на тези вектори.

● Всички частици общуват по между си (интелектът на рояка) и знаят кое е най-доброто място, открито от цялото ято към момента, и това, което те са открили за себе си. Критичната стъпка в алгоритъма е да се сравни локалният най-добър резултат на всяка птица с най-добър глобален резултат на лидера на ятото. С други думи, движението на всяка частица е резултат от суперпозицията на три основни тенденции – инерция (стремеж на частицата да запази посоката си), когнитивно влияние (привличане към собствения исторически успех), социално влияние (привличан от лидера).

● Във всяка стъпка (итерация) всяка частица променя посоката и скоростта си, като се привлича едновременно към своя личен най-добър резултат и към глобалния най-добър резултат на ятото. Извършва се и съответна актуализация на скоростите, в крайна сметка всички птици в ятото са много близо една до друга около глобалния оптимум.

Когато една частица намери много добро решение, тя „привлича“ останалите натам, което води до изключително бърза и ефективна оптимизация на модела. Динамиката на ятото се описва чрез система от уравнения, които дефинират промяната на състоянието на всяка частица i в дискретни моменти от време t:

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot \left(p_{i,\text{best}} - x_i(t)\right) + c_2 \cdot r_2 \cdot \left(g_{\text{best}} - x_i(t)\right) \; x_i(t+1) = x_i(t) + v_i(t+1)$$



Фигура 42. Теория на ятото

Критична особеност на алгоритъма е способността му за динамична оптимизация. Тъй като той непрекъснато актуализира скоростите въз основа на относителните разлики в позициите, може ефективно да проследява оптимуми, които променят координатите си във времето.

Във военното дело теорията на ятото не се използва само за трениране на софтуер в лабораториите, а се прилага физически на бойното поле за рояци от дронове. Тук изкуственият интелект, базиран на рояка, контролира десетки, стотици или дори хиляди малки машини, които действат като единен организъм. Ако например трябва да се неутрализира вражеска зенитна система, вместо да изпратят един скъп изтребител или голям дрон (който лесно може да бъде засечен и свален), се изпраща рояк от 200 малки, евтини дронове камикадзе. Как интелектът на рояка работи в тази ситуация:

● Децентрализация. В рояка няма един дрон лидер, всеки от тях взема самостоятелни решения на базата на локалните си сензори и комуникацията със съседните му дронове (точно като мравките или птиците). Ако ПВО системата свали 20 дрона, роякът не спира да функционира. Той мигновено се прегрупира и продължава мисията, защото няма централен мозък, който да бъде унищожен.

● Насищане на атаката. Използвайки алгоритми като гореописаните, дроновете в реално време изчисляват откъде да атакуват. Те се разпръскват, за да объркат радарите (които не могат да проследят 200 цели едновременно) и след това изкуственият интелект ги насочва да се сближат към целта от различни ъгли едновременно.

● Колективно разузнаване. Ако един дрон от ятото засече замаскиран радар чрез камерите си, той веднага предава координатите на останалите. Ятото автоматично пренасочва част от ресурсите си, за да унищожи новата заплаха, без нужда от човешка намеса.

За да може един такъв рояк от дронове да работи перфектно на бойното поле (да избягва сблъсъци помежду си, да се съобразява с вятъра и да намира целите), изкуственият интелект, който го управлява, е бил предварително оптимизиран в компютърни симулации, използвайки същата тази математическа теория на ятото. Освен това този алгоритъм може да се използва в контекста на системите за управление, това позволява адаптиране на параметрите в реално време при промяна на външните условия или състоянието на обекта. Също така в киберсигурността PSO се прилага за откриване на аномалии и оптимизиране на ресурси при динамично променящи се заплахи.

В индустрията се прилага за задачи от рода на избор на оптимална структура на невронна мрежа за прогнозиране на цените на електроенергията. Едно

от първите медицински приложения на оптимизацията на ятото е за успешна класификация на тремора на ръцете, свързан с болестта на Паркинсон.

Предимствата на груповите алгоритми са, че те могат да се прилагат за оптимизация в условия на висок шум на сигнала и сложни пространства за търсене. За разлика от класическите методи за оптимизация, методите на груповия интелект са прости и принципите им са лесни за прилагане. Въпреки това те все още се нуждаят от задаване на няколко параметъра за настройка, като повечето от тях са специфични за конкретния проблем. Основните недостатъци са, че те са неконтролируеми от ръководните органи, трудно мащабируеми и имат ограничена размерност за работа с данни.

ЕЗИКОВИ МОДЕЛИ НА ТРАНСФОРМАТОРИТЕ В СЪВРЕМЕННАТА АРХИТЕКТУРА

Теорията на вероятностите
не е нищо друго освен здравия разум,
сведен до изчислението
Пиер-Симон Лаплас

Съвременните системи с изкуствен интелект се основават на развитието на невронните мрежи, които позволиха на изчислителните машини да решават задачи, считани доскоро за изключителен прерогатив на човешкия ум. В процеса на създаване на всяка интелигентна система се дефинират три императивни компонента (Goodfellow et al. 2016):

1. Моделът – специфичната софтуерна архитектура (брой слоеве, тип неврони, връзки) или програма, която търси решение чрез математическа функция или множество от функции, параметризирани по такъв начин, че да могат да описват връзките в данните.

2. Данните – обучителните масиви с множество от примери, използвани за обучаване и тестване на модела. В научната литература те се разглеждат като „опит“, от който системата черпи знание, и се състоят от входни променливи и изходни резултати.

3. Алгоритъмът за обучение е методологията за адаптация към данните, която намира такива стойности на параметрите на модела, при които грешката е минимална.

В научния контекст тези три компонента са неразривно свързани – дори при перфектен алгоритъм, без качествени данни или адекватен модел системата няма да работи. Класически пример доскоро бяха рекурентните невронни мрежи (RNN), които обработват информацията последователно (дума по дума). Това води до два фундаментални проблема:

– бавно обучение поради последователната природа на архитектурата, която не позволява паралелни изчисления;

– изчезващия градиент при работа с дълги текстове, който кара мрежата буквално да „забравя“ началото на изречението, докато стигне до края му, като така губи контекста.

Едно от решенията на тези проблеми беше предложено през 2017 г. от екип на Google, който публикуваха статия *„Вниманието е всичко, от което се нуждаете“ (Attention Is All You Need)*[15], описваща новата архитектура на трансформатора *(Transformer)*. Тя бележи прелом в обработката на естествен език (NLP), като преодолява основните ограничения на рекурентните невронни мрежи. За разлика от тях, трансформаторите обработват целия входен текст едновременно (паралелно). Тази архитектура е оптимизирана за работа със съвременните графични процесори (GPU), което позволява обучението върху масивни мащаби от данни.

Големите езикови модели (LLM) като съвременните GPT, Claude и Gemini не работят директно с текст, а с числа. Преди трансформаторът да започне работа, текстът преминава през две фази:

– токенизация – входният текст се раздробява на „токени“ (думи, морфеми или знаци). Например думата „котката“ може да се разбие на „Котка“ и „та“. Всеки токен получава уникален числов индекс;

– векторно внедряване – индексите се превръщат във вектори във високоизмерно пространство. Тук думи с близки значения (напр. „войник“ и „военнослужещ“) се позиционират математически близо една до друга, улавяйки тяхното семантично сходство.



Фигура 43. Представяне на токенизацията

[15.] VASWANI, A., и др. Attention is all you need. В: Advances in neural information processing systems [онлайн]. 2017, vol. 30. [прегледан на 3 юни 2026]. Налично на: https://arxiv.org/abs/1706.03762

Математическата архитектура на трансформаторния модел се основава на преобразуването на тези вектори чрез матрични операции. Тъй като трансформаторът чете всички думи едновременно, той се нуждае от т.н. **механизъм за внимание**, за да разбере контекста и словореда. Тъй като моделът не чете последователно, той няма вградено знание за реда на думите. За да му се „подскаже“ къде се намира дадена дума в изречението, към първоначалните вектори се добавят позиционни кодирания, базирани на синусоидни функции за позицията *pos* и измерението *i*:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right)$$

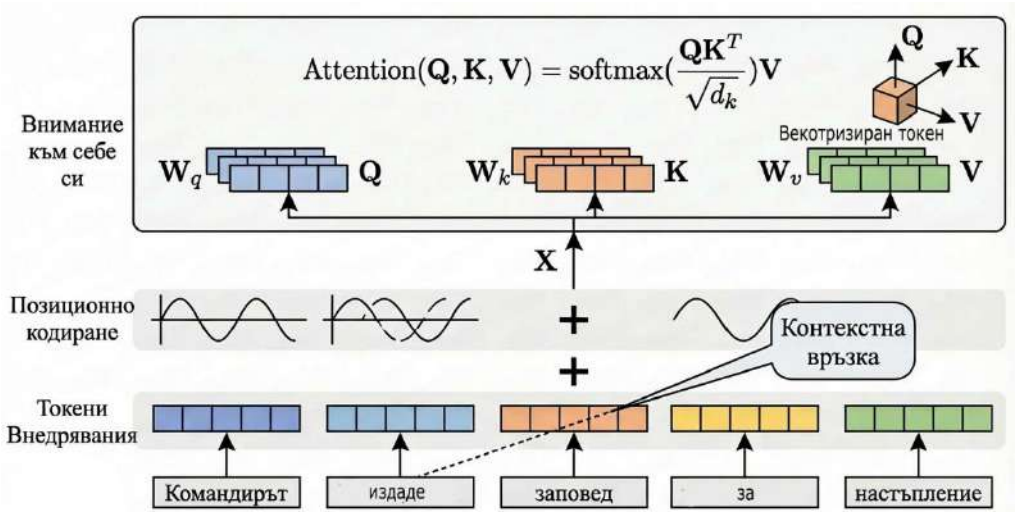
$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right),$$

където d_{model} е размерността на векторното внедряване (embedding).

Има два вида внимание, първото е **внимание към себе си** (Scaled Dot-Product Self-Attention). Това е механизмът, който филтрира информационния шум. Той оценява кои думи си влияят най-много. Например, в изречението *Командирът издаде заповед за настъпление, защото той...*, вниманието помага на модела да разбере, че *той* се отнася за командира. За всеки токен се генерират три вектора чрез умножение с научени тегловни матрици – заявка (Q – Query), ключ (K – Key) и стойност (V – Value). Те се получават чрез умножение на входната матрица X по научени тегловни матрици W^Q, W^K, W^V. Математическата дефиниция на вниманието е:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Коефициентът $\frac{1}{\sqrt{d_k}}$ (където d_k е размерността на ключовете) служи за мащабиране, предотвратявайки прекомерното нарастване на градиентите при прилагане на softmax функцията. Резултатът от softmax определя теглото (важността), с което всяка дума влияе върху текущия контекст.



Фигура 44. Резултатът от softmax на вниманието към себе си

Вторият вид внимание е **многоглавото внимание** (Multi-Head Attention). За да може моделът да следи едновременно различни аспекти на изречението (например синтактични и семантични връзки), се използва Multi-Head Attention. Входните данни се проектират в h на брой различни пространства:

$$\text{MultiHead}(Q, K, V)\& = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

Устройството на стандартния трансформатор се състои от два основни блока:

1. Енкодер – използва двупосочно внимание, за да прочете и разбере целия входен текст наведнъж. Подходящ е за анализ (напр. моделът BERT).

2. Декодер – генерира изходния текст дума по дума. Използва маскирано самовнимание, което му позволява да „вижда“ само предишните генерирани думи, но не и бъдещите. Подходящ е за създаване на текст (напр. GPT).

Кръстосано внимание се използва, когато декодерът трябва да „погледне“ към резултата от енкодера. При превод на „Командирът издаде заповед за настъпление при изгрев“ на английски език този механизъм „осветява“ думите „настъпление“ и „изгрев“, като гарантира запазването на смисъла в изходната фраза “to advance at sunrise”.



Фигура 45. Работа на BERT

Схемата на фигура 46 илюстрира работата на BERT (Bidirectional Encoder Representations from Transformers) – един от най-важните модели на Google в областта на изкуствения интелект за разбиране на естествен език, който работи в следния ритъм:

● Вход (w1 – w5) – моделът получава изречение от пет думи. Ключовият елемент тук е [MASK]. За да се обучи BERT да разбира контекста, една дума (в случая w4) се „скрива“ (маскира). Целта на модела е да отгатне коя е тя с т.нар. Masked Language Modeling (MLM).

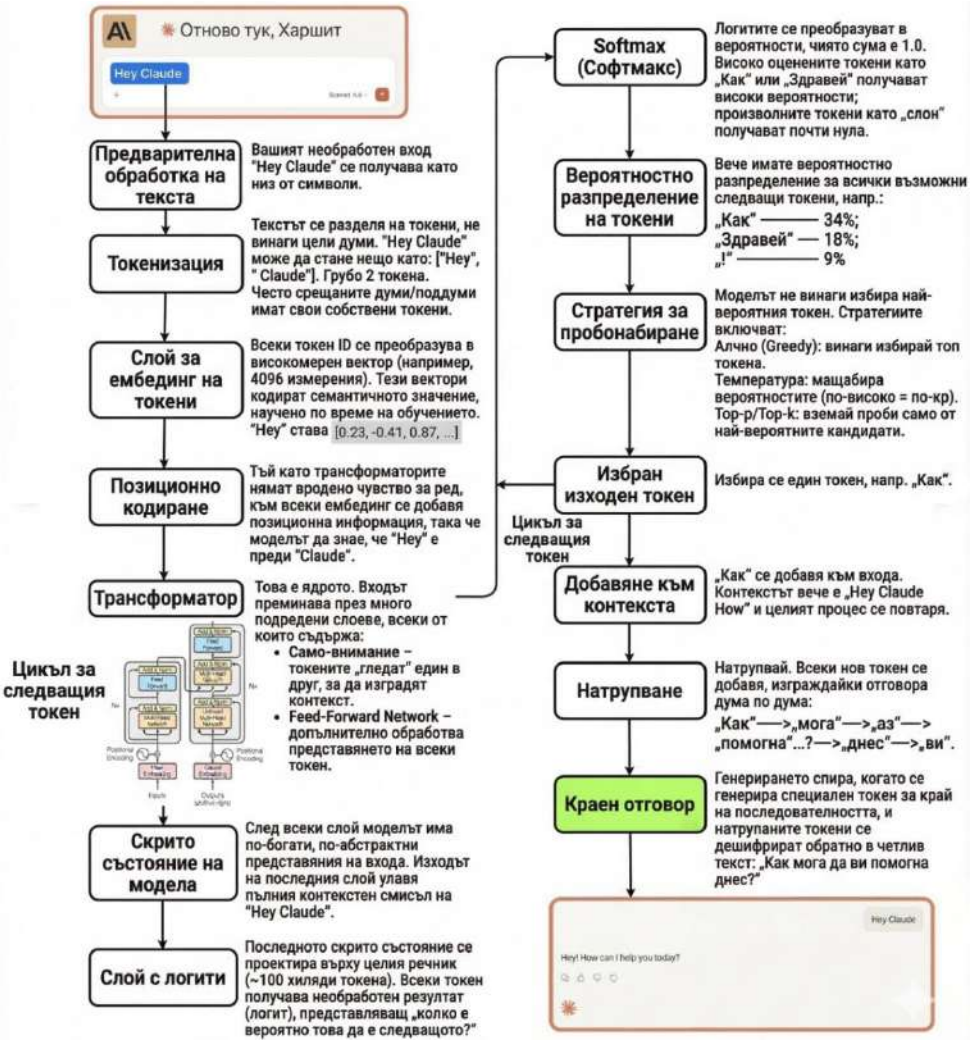
● Влагане (Embeddings) – думите и токенът [MASK] се преобразуват в поредица от числа (вектори), наречени embeddings. Това позволява на компютъра да ги обработва.

● Двупосочност (Transformer – Encoder) – анализира всички думи в изречението едновременно. Двупосочните стрелки вътре показват, че всяка дума „гледа“ всички останали думи (преди и след нея), за да разбере пълния

контекст. Това е фундаменталната разлика с по-старите модели, които гледат думите само отляво надясно.

● Класификационен слой – взема обработената информация и я преобразува в прогнози.

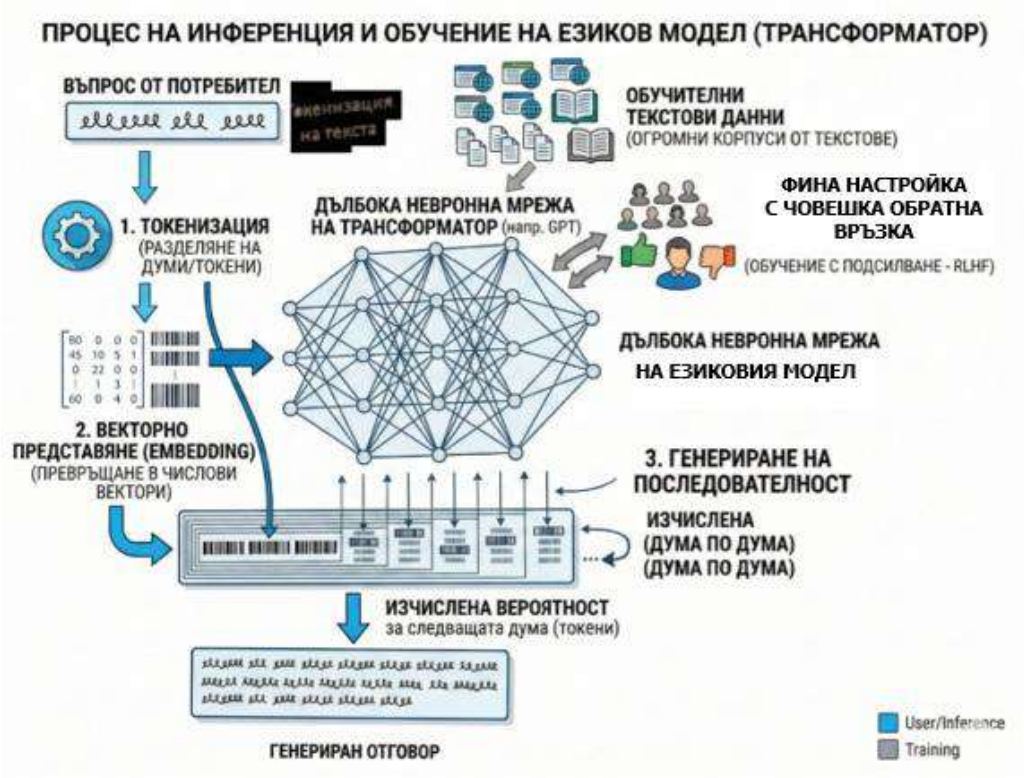
● Изход (w'1 – w'5) – опитва се да възстанови оригиналните думи. За маскираната позиция (4) моделът прогнозира W'4.



Фигура 46. Как работи Claude

Създаването на по-новите модели, като ChatGPT или Gemini, преминава през два основни етапа:

● Предварително обучение (Pre-training) – моделът попива трилиони думи от интернет. Тук той се учи на граматика, факти и логика чрез една проста задача – да предсказва следващия токен в изречението.



Фигура 47. Процес на обучение на езиков модел

● Фина настройка (Fine-Tuning/RLHF) – тук моделът се учи „да бъде асистент“. Чрез обучение с подсилване от човешка обратна връзка (RLHF) експерти класират отговорите на модела, като го обучават да следва инструкции, да бъде полезен и да избягва вредно съдържание.

Сложността на класическия механизъм за внимание нараства квадратично спрямо дължината на текста – $O(n^2.d)$, където n е дължината на последователността. Това прави обработката на много дълги документи (напр. военни концепции или технически спецификации) изчислително

интензивна. По тази причина съвременните архитектури въвеждат редица оптимизации, насочени към намаляване на изчисленията, пестене на памет и увеличаване на дължината на контекста:

– Multi-Head Attention (MHA) – използва паралелни глави на внимание, които улавят едновременно различни синтактични и семантични зависимости в текста;

– Grouped-Query Attention (GQA) – групира заявките за драстично намаляване на потреблението на памет. Този подход е ключов за ефективността на популярни модели като LLaMA и Mistral 7B;

– Multi-Query Attention (MQA) – използва само един набор от ключове и стойности за всички заявки, което допълнително оптимизира скоростта и паметта;

– FlashAttention – алгоритъм, който оптимизира достъпа до паметта на графичните процесори (GPU), като намалява четенето и писането, което ускорява значително процеса;

– Star-Attention (NVIDIA) – иновативен подход, постигащ до 11-кратно ускорение при обработката на дълги последователности чрез блоково разпределение на вниманието;

– Lightning Attention (MiniMax 2025) – специализирана архитектура, създадена за свръхбърза обработка на изключително дълги текстови последователности без загуба на контекст;

– Mixture of Experts (MoE) – архитектура, при която при всяко изпълнение се активира само избрано подмножество от „експерти“ (специализирани подмодули), вместо да се изчислява целият модел. Това води до по-ефективно използване на изчислителните ресурси и отлична мащабируемост (Kordon 2023).

В последно време се появяват алтернативи на Трансформатора като **Mamba**, базирана на селективни модели на състоянието (Selective SSM). Такъв пример е разработката на NVIDIA, наречена Nemotron 3 Super, която използва хибриден формат – Mamba-Transformer. Тази архитектура съчетава ефективността на Мамба при дълги последователности (до 1 милион токена) с логическата прецизност на трансформаторните слоеве. Това прави модела идеален за автономни агенти, анализ на код и сложни академични изследвания.

За да опишем разликата между работата на трансформатор и мамба, можем да направим аналогия с група ученици в една класна стая. Всеки ученик (токен от текста) трябва да „погледне“ към всички останали ученици едновременно, за да разбере връзките: „Кой е главният герой? Какво каза

преди 10 изречения?“. В механизмът за внимание всеки токен „преглежда“ всички предишни (квадратично – ако са 1000 думи, прави 1 000 000 връзки). Перфектно помни всичко, идеално за кратки разговори или сложни връзки, но „забива“ при дълги текстове – изисква огромен RAM и време (O (N^2). В същото време мамба работи като умна лента за пренос – чете текста дума по дума, но поддържа „компресирана бележка“ (скрито състояние) с най-важното. Selective SSM динамично решава „това е важно – запомни“, „това е шум – забрави“, обработва линейно милиони думи без да „забие“, но не вижда „назад“ директно, компресира информацията.

Всички изброени архитектури са насочени към повишаване на ефективността, намаляване на хардуерните изисквания и разширяване на възможностите за обработка на огромни обеми информация. Така се осигурява ненадминатата производителност и адаптивност на съвременните големи езикови модели.



Фигура 48. Обучителен процес на LLM

Тъй като големите езикови модели (LLM) са вероятностни, а не детерминистични системи, те са склонни към явления, наречени

„халюцинации“ – генериране на фактически неверни, но изключително убедително представени твърдения. Това се случва, защото моделът е математически оптимизиран да предсказва следващата логична дума в поредицата, без да притежава вграден механизъм за реална проверка на фактите.



Фигура 49. AI е просто инструмент и всички негови отговори трябва да се проверяват, отговорността си остава на човека – интернет меме от неизвестен автор

Още едно фундаментално ограничение на LLM е контекстовият прозорец. Това е максималното количество информация (измерено в токени), което моделът може да „държи в активната си памет“ и да обработва наведнъж в рамките на една заявка.

Пример: Ако водите диалог с модел с лимит от 4096 токена, след достигането на този праг моделът започва да „забравя“ най-старите токени от началото на разговора. Това е критично уязвимо място при анализиране на дълги документи (напр. разузнавателни доклади).

За да стане ясен мащабът, в таблица 5 е представено сравнение на контекстовия прозорец на някои от водещите съвременни модели към 2025 г.

Таблица 5. Когнитивен прозорец на някои модели

Модел (актуални към 2024/2025)	*Контекстов прозорец (токени)*	*Еквивалентен обем информация*
GPT-3.5 Turbo	4096	~ 5 страници
GPT-4o / GPT-4 Turbo	128 000	~ 300 страници
Claude 3.5 Sonnet	200 000	~ 500 страници
Gemini 1.5 Pro	1 000 000 – 2 000 000	~ 3000 – 6000 страници

Един примерен разговор с GPT-3.5 Turbo, който има контекстен прозорец от 4096 токена, може да се развие така:

– [Токен 1 – 50] „Здравей! Искам да създам уебсайт за...“

– [Токен 51 – 200] „Уебсайтът трябва да има регистрация и...“

– ...

– [Токен 3900 – 4096] „Благодаря за отговора!“

След достигане на лимита от 4096 токена, моделът „забравя“ първите токени и може да загуби контекст. Превишаването на този лимит води до „забравяне“ на началната информация и халюцинации, което е критично при анализиране на дълги документи.

В критични сектори като отбраната и медицината халюцинациите са недопустими. Затова се налага използването на системи с „човек в контура“ или внедряването на архитектури за генериране, подкрепено с извличане на информация (**Retrieval-Augmented Generation – RAG**). RAG не е самостоятелна невронна мрежа, а хибридна система, която принуждава модела първо да извлече факти от надеждна външна база данни (напр. секретни военни архиви), преди да генерира отговор. Процесът работи в три основни фази:

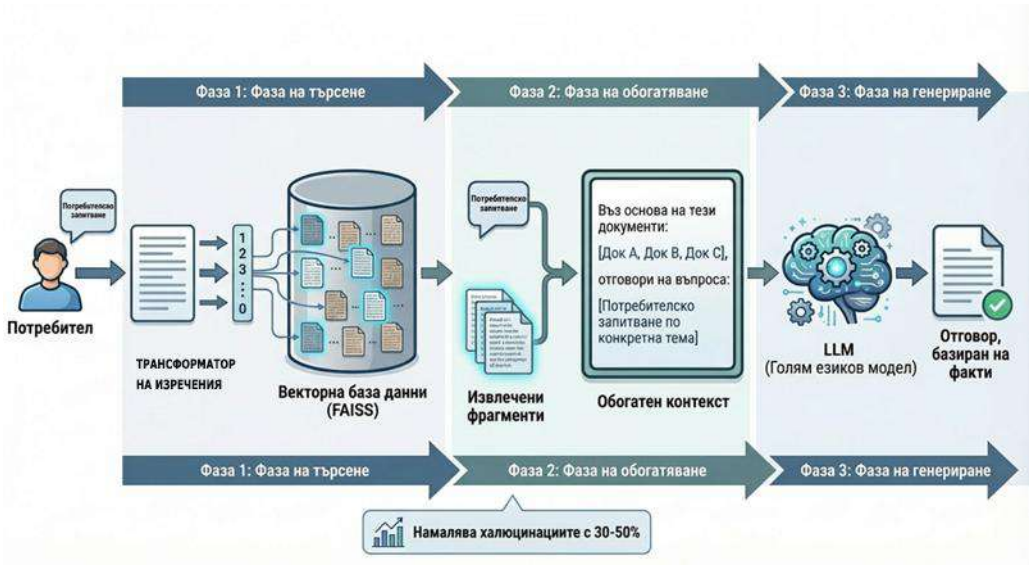
● Фаза на търсене – специализиран модел (напр. Sentence Transformers) преобразува дадено потребителско запитване в числов вектор. Системата търси във външна векторна база данни и извлича най-релевантните текстови фрагменти чрез изчисляване на косинусово сходство.

● Фаза на обогатяване – извлечените надеждни документи се добавят към първоначалната заявка, за да се образува обогатен контекст (пр. „Базирано на следните документи: [X], отговори на въпроса: [Y]“).

● Фаза на генериране – LLM генерира финалния отговор, обусловен стриктно от подадените му факти.

Тази архитектура води до драстично намаляване на халюцинациите (с до 30 – 50% при бенчмаркове като Natural Questions), като комбинира генеративната мощ на трансформатора с фактологичната точност на търсачката. С други думи, потребителят задава въпрос, който преминава през модел за вграждане (енкодер), чиято задача е да превърне думите в математически вектор. Това означава, че въпросът вече е преведен на компютърен език и позициониран като точка/стрелка в d-измерно пространство. Централната координатна система, показана на фигура 52 (с оси *Embedding Измерение 1* и *Измерение 2*), илюстрира графично как работи векторната база данни. Синият вектор представлява зададен въпрос. Останалите вектори са

фрагменти от документи (знания), които предварително са били обработени и запазени в базата данни.



Фигура 50. Работа на Retrieval-Augmented Generation – RAG система



Фигура 51. Работа на векторна база данни

Чрез математическа формула за косинусово сходство машината разбира кои документи отговарят на въпроса. Колкото по-малък е ъгълът на векторите между зададения въпрос и векторите на документите, толкова по-голямо смислово сходство има между тях. Тези математически съвпадения се декодират обратно в текст – това са извлечените релевантни параграфи или документи. Накрая се преминава към генеративната фаза.

Големият езиков модел получава едновременно два потока информация – на оригиналния въпрос и фактите, извлечени от базата данни. Благодарение на подадения контекст, LLM не разчита само на своята вътрешна (и понякога неточна) памет, а синтезира проверена информация, за да създаде крайния, прецизен отговор.

В среда с висока сигурност въпросът за конфиденциалността на данните често налага използването на локални езикови модели (Open Source модели, стартирани на собствени сървъри), вместо разчитането на облачни API услуги. Преди внедряване на такива модели обаче трябва да се направи строг анализ на разходите и ползите.

За да бъде стартиран голям локален модел (например Llama 3.1 със 70 милиарда параметъра), са нужни над 128 GB RAM и мощни специализирани графични процесори. Инвестицията в хардуер лесно надхвърля 15 000 евро, а консумацията на енергия е значителна. За гигантски модели (405B параметъра) цената на сървъра може да достигне над 50 000 евро. Затова често се използват облачни алтернативи. За сравнение, достъпът до облачни услуги струва средно между 20 и 200 евро месечно. Облачните модели обикновено предлагат най-високото възможно качество, огромни контекстови прозорци и не изискват поддръжка от страна на потребителя. Локалните езикови модели със същия размер често са по-бавни и изискват постоянен екип за дебъгване и актуализации.

Въпреки това локалната инсталация е абсолютно задължителна в няколко случая – военни и правителствени организации, обработка на чувствителни медицински данни, корпорации с милиони заявки месечно (където облакът става по-скъп) или научни институции, изискващи пълен контрол върху модела. Популярни Open Source модели за локално ползване са:

– Llama 3.1 (Meta) – предлага се във варианти с 8B, 70B и 405B параметъра;

– Mixtral 8x7B (Mistral AI) – използва ефективната *Mixture of Experts* архитектура;

– Qwen/WizardLM – модели, фино настроени за сложни логически задачи;

– BgGPT 27B – българският модел на INSAIT от 2024 г.

Платформи за локално стартиране:

– Ollama – изключително лесен метод за стартиране през терминал (Mac/Linux/Windows);

– LM Studio – удобен графичен интерфейс (GUI) за инсталиране и тестване на модели;

– MLX – рамка на Apple, специално оптимизирана за техните M-серия чипове.

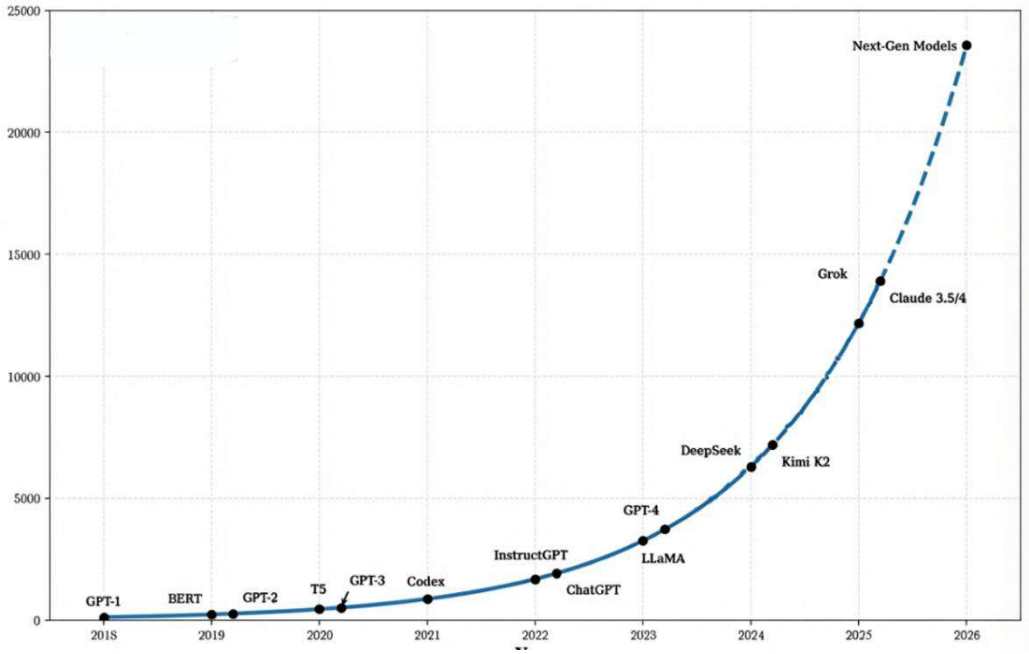


Фигура 52. Разлика между локален и Cloud LLM

Ако до 2023 г. основната метрика за прогрес беше просто размерът на модела (броят параметри), днес съвременните диаграми за растеж отчитат съвсем различни показатели. На фигура 54 е показано как се развиха възможностите на генеративния изкуствен интелект през последните години.

През 2023 г. настъпва ерата на мултимодалността и огромния контекст. Модели като Gemini 1.5 Pro въвеждат прозорци от 1 до 2 милиона токена. Това означава, че AI вече може да обработва не просто кратки промптове, а цели книги, дълги видеоклипове или огромни кодови бази наведнъж. Моделите спряха да бъдат само „текстови генератори“. Системи като GPT-4o, Claude 3.5 и Gemini започнаха да „виждат“ изображения и да „чуват“

аудио, генерирайки отговори в реално време, без нужда от външни транскрибатори.



Фигура 53. Параметри на растеж на LLM по години

Впоследствие, през 2024 г., настъпи ерата на разсъжденията и отворените модели с дълбоко мислене. Въвеждането на модели като OpenAI o1 (Strawberry) и серията DeepSeek R1 промени парадигмата. Вместо да бълват моментални (и понякога грешни) отговори, тези модели използват скрити вериги от разсъждения, за да „мислят“ и планират стъпките си, преди да отговорят. Това доведе до феноменални резултати в сложната математика и програмирането. Индустрията осъзна, че по-голямото невинаги е по-добро. По-малки модели (между 7 и 13 милиарда параметъра) често предлагат най-добрия баланс между скорост, точност и цена.

През 2026 г. навлязоха т.н. AI агенти. Фокусът вече не е върху чат ботовете, а върху агентния изкуствен интелект. Това са автономни системи, които действат с конкретно намерение да изпълняват сложни задачи в реалния свят (например автономно резервиране на полети, управление на софтуерни проекти или анализ на данни).

До средата на 2026 г. повече от 50% от взаимодействията при обслужване на клиенти ще включват агентен AI. Инвестициите в тези технологии се отплащат значително, като за всеки $1, инвестиран в генеративен AI, компаниите виждат средна възвръщаемост от $3,70. Глобалният пазар на генеративен AI достига 161 милиарда долара през 2026 г.

В експеримента „Китайска стая“ американският философ Джон Сърл (1980) показва, че машините не мислят истински, а просто изпълняват алгоритми, без да разбират смисъла на действията си. Въпреки това, с появата на множеството съвременни автономни системи, обществото се изправя пред необходимостта от нови правни рамки. Ако един агент с изкуствен интелект може да взема автономни решения – особено във военни ситуации – възниква въпросът кой носи юридическа отговорност при евентуална грешка.

За разлика от човешкия ум, който философът Лудвиг Витгенщайн описва като „затворена кутия“ (никога не можем да сме напълно сигурни какво изпитва другият), изкуственият интелект може да бъде деконструиран. Матриците от тегла и механизмите за внимание в един трансформатор могат да бъдат анализирани, което парадоксално може да хвърли светлина и върху механизмите на самото човешко мислене.

Трансформаторната архитектура промени парадигмата на изкуствения интелект, превръщайки го от тясно специализиран алгоритъм в универсален езиков и когнитивен двигател. Разбирането на механизмите на вниманието, векторизацията и ограниченията (като халюцинациите) е абсолютно ключово за всеки специалист, целящ да внедрява интелигентни системи в реалния свят, където прецизността на контекста и сигурността на данните са от стратегическо значение.

ПРОМПТ ИНЖЕНЕРИНГ И КОНТЕКСТ ИНЖЕНЕРИНГ

Въпросът, дали компютрите могат да мислят,
е като въпроса дали подводниците могат да плуват
Алън Тюринг

В съвременната дигитална епоха, когато изкуственият интелект намира все по-широко приложение в обществения и отбранителния сектор, умението за формулиране на прецизни и ясни промптове (подкани, подсказки, инструкции) се превръща в ключов фактор за ефективно взаимодействие с езиковите модели. Ако големите езикови модели (LLM) могат да бъдат разглеждани като мощен мозък, то промпт инженерингът представлява умението той да бъде управляван правилно.

Промпт инженерингът е сложна интердисциплинарна област, обединяваща компютърна лингвистика, когнитивна психология и системно инженерство с цел извличане на максимална стойност от стохастичните процеси на предсказване на езика (Brown et al. 2020). Това е процесът на целенасочено конструиране, формулиране и усъвършенстване на заявки (инструкции, подкани) към модели на изкуствен интелект с цел постигане на точни, релевантни, приложими и възпроизводими резултати. AI системите генерират своите отговори въз основа на входния текст (промпта), като дори минимални промени във формулировката могат да доведат до съществено различни резултати (Reynolds & McDonell 2021). Промптът представлява входната инструкция, въпрос или контекст, който потребителят предоставя на AI модел, за да получи желания изход.



Фигура 54. Принципът „Вкараш ли боклук, ще изкараш боклук“

Мрежите за машинно обучение са толкова добри, колкото са данните, с които биват обучавани. Най-честата причина за неточните резултати на изхода на някоя ИИ система са лошите данни за обучението ѝ, което означава непълна, зле обозначена, изобилстваща с грешки или пристрастия база (Mitchell 1997). Например прогнозните ИИ системи, обучени с противоречиви и неточни данни за предишни явления, ще направят безполезни предвиждания. В света на компютърните науки концепцията, че лошият вход води до лош изход, се нарича garbage in, garbage out (GIGO) (Babbage 1864).

При работа с езикови модели степента на яснотата и конкретността на въпроса корелира пряко с качеството на получения отговор. Общи запитвания от типа „Как да програмирам?“ или „Разкажи ми за циклите“ обикновено водят до абстрактни и трудно приложими отговори. За разлика от тях, ясно дефинираните заявки с конкретен контекст – например: „Как да напиша Python функция за сортиране на списък с цели числа?“ или „Обясни как работи for цикъл в JavaScript“ – позволяват на модела да предостави структурирани, детайлни и практически полезни отговори, включително примери, обяснения и алтернативни решения (Wei et al. 2022).

Контекстът е първият и най-съществен елемент на всеки ефективен промпт. Той направлява „мисловния процес“ на модела и му позволява да оцени адекватно ситуацията. При анализ на проблемен код или решаване на конкретна задача предоставянето на детайлна информация – включително самия код, очакваните резултати, възникналите грешки, предназначението на програмата, води до значително по-качествени и релевантни решения.

● Пример за добър промпт с контекст: *Пиша JavaScript функция за изчисляване на факториела на число, но тя връща NaN за стойности над 1. Ето кода: [код]. Какъв може да е проблемът?*

● Пример за лош промпт: *Кодът ми не работи*.

Формулировката на заявката влияе не само върху съдържанието, но и върху формата и стила на отговора. Ясното дефиниране на очаквания резултат (примерен код, концептуално обяснение, таблица, списък) значително подобрява полезността на генерирания отговор.

Добре формулираната заявка обикновено включва следните основни компоненти (White et al. 2023):

● Контекст – нужна е релевантна информация, за да помогнете на AI да разбере ситуацията. Например вместо „Помогни ми да напиша имейл“, опитайте с „Пиша имейл до клиент, за да потвърдя среща по проекта. Тонът трябва да е професионален, но достъпен“.

● Роля – дайте му роля, например: „Ти си опитен военен, участвал в много мисии“.

● Задача – ясно посочете какво искате да направи моделът, например „Обобщи този текст с точки“ или „Напиши въведение от два абзаца за презентацията ми“.

● Детайли и ограничения – добавете конкретни изисквания като тон, дължина или стил. Пример – „Пренапиши този параграф в неформален стил, като остане под 100 думи“.

Ето още един пример за лоша заявка:

Разкажи ми за маркетинга – тя е неясна и липсва фокус, което води до общ отговор.

Пример за добра заявка:

Можеш ли да дадеш обяснение в три параграфа за съдържателния маркетинг, като се фокусираш върху това как малките бизнеси могат да използват блогове и социални мрежи за увеличаване на разпознаваемостта на марката. Тя е ефективна, защото е конкретна, детайлна и поставя ясни очаквания за резултата.

Един от ефективните подходи за създаване на качествени промптове е методът на Сократ, който представлява кооперативен аргументиран диалог, стимулиращ критичното мислене чрез последователно задаване и оспорване на въпроси:

– започване на разговора с общ въпрос, свързан с концепцията, от която се интересувате;

– задълбочаване с по-конкретни въпроси;

– оспорване на отговорите;

– поддържане на диалогичност;

– помислете върху получените отговори.

Този метод е особено приложим в обучението, анализа и вземането на решения.

Нека разгледаме пример, в който се интересуваме от отбрана на военна база. Вашето взаимодействие с ChatGPT може да протече така:

1. Започвате с въпрос – „Какви са основните тактически похвати за ефективна отбрана на позиция?“. AI ще ви обясни как се организират отбранителни линии, използване на естествени и изкуствени прегради, създаване на резервни позиции, инженерни заграждения т.н.

2. След това задавате по-конкретен въпрос – „Може ли да дадеш пример за използване на инженерни заграждения при отбрана на позиция?“.

ChatGPT описва ситуация за употреба на система от взривни и невзривни инженерни заграждения.

3. Оспорвате отговора – „В кои случаи използването на ежове и надълби може да се окаже неефективно?“. ChatGPT анализира, че ако противникът успее да заобиколи заграждения, те ще се окажат безсмислени.

4. Добавяте „Можеш ли да дадеш пример от историята, в който това е довело до успех?“.

Потребителят може изрично да зададе **нивото на разсъждение** – минимално разсъждение с бързи отговори, балансирано стандартно ниво и високо ниво на анализ и разсъждение. Примерни инструкции:

– „Мисли стъпка по стъпка“;

– „Проучи поне 3 подхода и избери най-добрия“;

– „Направи задълбочен анализ, преди да отговориш“.

Ето още един добър пример по рецептата *контекст – роля – задача – детайли*:

● Задача – „Създай кратка информационна брошура за пациенти, обясняваща какво е хипертония (високо кръвно налягане)“.

● Роля – „Ти си лекар с дългогодишен опит в семейната медицина, свикнал да обясняваш сложни медицински състояния на обикновени хора по разбираем начин“.

● Контекст – „Работиш с местна здравна клиника, която иска да предостави на пациентите си лесна за разбиране и визуално подредена брошура за високо кръвно налягане“.

● Целева аудитория – „Възрастни хора между 45 и 75 години, които не разбират медицинска терминология, често пропускат прием на лекарства и не осъзнават дългосрочните рискове“.

● Тон и стил – „Спокоен, подкрепящ и ясен. Без медицински жаргон. С примери от ежедневието“.

● Изходен формат – „Текст за А5 брошура (двустранна), разделен на 3 секции – какво е хипертония, какво причинява тя, какво можеш да направиш още днес“.

Ето и един лош пример – *Дай ми идеи за декориране на хола*. Даването на повече контекст би довело до получаването на по-добър резултат. Например – *Имам хол 5х4 метра на шпакловка и замазка. Искам да го обзаведа така, че да има нещо модерно създаващ спокойствие стил. Дай ми 3 предложения какво мога да направя. Всяко предложение трябва да включва цвят на стената, стил на обзавеждане, декор и някои възможни мерки. Резултатът да е под формата на кратък списък*“.

Един от основните проблеми в ранните етапи на развитието на LLM беше липсата на структура при формулирането на инструкциите. С натрупването на практически опит бяха разработени няколко метарамки, които систематизират процеса на създаване на висококачествени промптове. Тези рамки служат като когнитивни шаблони, които гарантират, че всички необходими елементи за успешното изпълнение на задачата са включени в инструкцията.

Най-известната сред тях е рамката TCREI (Task, Context, References, Evaluate, Iterate) представлява цялостен жизнен цикъл за развитие на промпт, който се фокусира върху итеративното подобрение. Тя е разработена като част от програмите за обучение на Google и акцентира върху концепцията, че промптването не е еднократно събитие, а процес на непрекъснато рафиниране.

– **Task (задача).** Ясно дефиниране на крайната цел. Инструкцията трябва да използва силни глаголи за действие (напр. „анализирай“, „генерирай“, „класифицирай“) и да определя желаната роля (персона) на модела.

– **Context (контекст).** Предоставяне на фонова информация, която определя границите на задачата. Това включва целевата аудитория, желания тон и специфични ограничения.

– **References (референции).** Използване на примери (few-shot prompting) или външни източници на данни. Референциите служат като „чертеж“ за модела, като намаляват риска от халюцинации.

– **Evaluate (оценка).** Критичен преглед на генерирания резултат. Потребителят трябва да анализира дали отговорът съответства на зададените критерии за точност и стил.

– **Iterate (итерация).** Нанасяне на корекции в промпта въз основа на оценката. Този етап е ключов за преминаване от посредствен към отличен резултат.

Рамката CRISP (Context, Role, Instruction, Subject, Preset, Exception) е оптимизирана за сценарии, изискващи висока степен на формализация и предвидимост. Тя разширява стандартните модели чрез включване на компонента Exception (изключения), който изрично посочва какво моделът не трябва да прави. Вариацията CRISPY добавя елемента Yielding (отстъпване/продуктивност), който насърчава модела да задава уточняващи въпроси, преди да предостави окончателен отговор. Това е фундаментален механизъм за намаляване на несигурността при сложни бизнес задачи, където началните изисквания често са непълни.

Рамката RISE (Role, Input, Steps, Expectation) е проектирана за алгоритмични и процесно ориентирани задачи. Тя разделя инструкциите на логически стъпки, което е особено ефективно за образователни цели или технически наръчници. Добавянето на компонент Narrowing (стесняване) в модела RISEN позволява въвеждането на строги граници за обема на изхода.

Минималистичният подход ACT (Action, Context, Tone) е подходящ за бърза ежедневна комуникация и маркетинг. Той се фокусира върху емоционалния отпечатък и стилистиката на текста.

Рамката GRWC (Goal, Role, Workflow, Constraints) е често използвана при разработването на автономни ИИ агенти. Тя дефинира сложни работни потоци и механизми за безопасност, които гарантират спазването на специфични протоколи.

Формула за текст от чатбот. Роля – задача – контекст (правила, ограничения) – разсъждения (как да мисли преди да отговори) – изход – стоп (кога задачата свършва). Пример – *ти си финансов съветник за малък бизнес. Задачата ти е да създадеш месечен бюджетен план. Контекст (правила и ограничения): бизнесът е в България и печели 5000 евро на месец. Първо разпредели бюджета на фиксирани разходи, после на променливи и задължително отдели част за спестяванията. Покажи резултата като таблица с приходи, фиксирани разходи, променливи разходи и спестявания. Спри, когато създадеш само един месечен план*.

Формула за създаване на изображение. Фокус и основен обект – рендиране, стил или ниво на реализъм – атрибути и детайли за обекта –камера, обектив, осветление – среда, фон, дълбочина. Пример – *фокусът е собственик на малко кафене, който прави лате. Създай реалистична лайфстайл фотография. Допълнителни атрибути: да има жена на 40 години, с престилка и пара, издигаща се от чашата. Средата да е уютен интериор на кафене, дървен плот, размазани клиенти на заден план. Да има естествена сутрешна светлина през прозореца, 35 мм обектив, меки сенки*.

Формула за създаване на видео. Местоположение и атмосфера на сцената – актьор, който е на екрана – действие и движение на камерата – времетраене, начален кадър, краен кадър. Пример – *сцената е модерен домашен офис вечер, включена настолна лампа. Актьорът е млад предприемач, пише на лаптоп и мисли. Движението на камерата е бавно и се премества от страничен ъгъл към изглед през рамо. Той спира и се усмихва на екрана. Продължителността на клипа е 10-секунден непрекъснат кадър. Започва в широк ъгъл. Край в близък план на светещия екран на лаптопа.*

Една от най-значимите иновации в промпт инженеринга е откритието, че езиковите модели могат да подобрят своята способност за логическо разсъждение, ако им бъде указано да генерират стъпка по стъпка. Това се нарича *мисловна верига*. Терминът означава техника на прилагане на промптове по начин, по който моделът да бъде насърчен да имитира последователни (стъпка по стъпка) разсъждения. Обърнете внимание, че когато караме модела да реши уравнение, той се опитва да отговори директно. Като хора ние също не можем да решим сложни уравнения без помощта на лист и химикал. Възможно е обаче посредством промпта да насърчите модела да направи няколко междинни разсъждения. По този начин, както решават хората на хартия, така и моделът би могъл да реши по-сложни проблеми, ако му дадете време да помисли. Опитно е установено, че ако се добави фразата *мисли стъпка по стъпка* в края на промпта, на модела се дава възможност да се справи с по-сложни мисловни проблеми.

Тази техника на мисловната верига се нарича **стратегия на нулев изстрел** и е представена за първи път от колектив през 2022 г. Терминът нулев изстрел (zero-shot) означава, че моделът не разчита на специфични за задачата примери, за да направи разсъжденията, а трябва да се справи с проблема въз основа на своята базисна подготовка. За разлика от техники от типа на обучението с няколко изстрела (few-shot learning), които изискват моделът да е виждал примери на задачата, обучението с нулев изстрел се опитва да генерализира проблема без необходимостта от специфични примери.

При техниката на **обучението с няколко изстрела** ние даваме примери с входни данни и съответните желани резултати. След това на последния ред поставяме промта, който желаем да бъде завършен. Естествено, езиковият модел ще изпълни задачата, следвайки шаблона от предоставените му примери. Обучението с няколко изстрела (примера) е мощен аспект на големите езикови модели, тъй като ги прави изключително гъвкави и адаптивни, като само с ограничено количество допълнителна информация могат лесно да бъдат пригодени за изпълнението на различни специфични задачи.

За задачи, които изискват изследване на алтернативи или стратегическо предвиждане (lookahead), линейният процес на мисловната верига често е недостатъчен. Тук на помощ идва рамката **дърво на решенията** – Tree of Thoughts (ToT) (Yao et al. 2023). ToT позволява на езиковия модел да поддържа дървовидна структура от „мисли“, където всеки възел представлява междинна стъпка към решението. Методологията на ToT се опира на четири основни стълба:

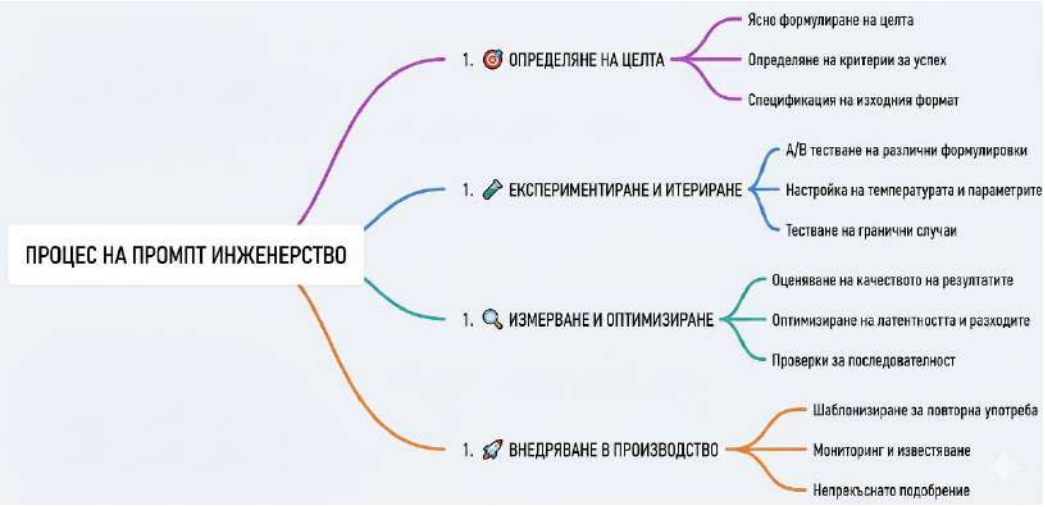
– декомпозиция на мислите – превръщане на процеса в поредица от дискретни езикови последователности;

– генератор на мисли – моделът предлага множество независими пътища за следващата стъпка (например чрез Breadth-First Search или Depth-First Search);

– евристична самооценка – моделът действа като критик, оценявайки шансовете за успех на всеки клон от дървото (напр. „сигурно“, „възможно“, „невъзможно“);

– търсене в пространството: Комбиниране на генерирането на отговор с алгоритми за търсене, което позволява на модела да се връща назад (backtracking), ако текущият път се окаже задънена улица.

Тези рамки доближават поведението на LLM до теорията за дуалните процеси в човешкото мислене – бавен, съзнателен и аналитичен процес.



Фигура 55. Процес на промт инженеринг

Нека да проследим изпълнението на задача за направа на чатбот, който да генерира отговори на мейлите, получавани във вашата компания. Вие желаете да се генерират такива отговори, които да запазят ви стила ви на писане. За това има два подхода – или да използвате техниките на промт инженеринг, или да направите фина настройка на съществуващия модел.

На първо място, трябва да предоставите на AI данни за конкретната бизнес сфера, за други запитвания и техните отговори. След това може да

използвате тези данни, за да настроите фино съществуващия модел, така че той да овладее както конкретните езикови шаблони, използвани от компанията ви, така и специфичната лексика. С помощта на фина настройка на теглата в съществуващите LLM може да създадете високо персонализиран и специализиран генератор за отговори на имейли, съобразен с езиковите шаблони и лексиката, характерни за вашата конкретна бизнес сфера.



Фигура 56. Процесът на фина настройка

Промпт инженерингът и обучението с нулев или няколко изстрела не променят вътрешните параметри на модела, докато фината настройка представлява допълнително обучение върху специфичен масив от данни (обикновено в JSONL формат). Качеството на данните и експертната им проверка са критични за успеха на фината настройка (Goodfellow et al. 2016).

Познаването и прилагането на техниките за промпт инженеринг, zero-shot и few-shot обучение, мисловни вериги и фина настройка позволяват създаването на ефективни, мащабируеми и целево ориентирани AI приложения. Чрез прецизно управление на контекста, задачата и формата на отговора взаимодействието с генеративния изкуствен интелект може да бъде превърнато

в мощен инструмент за анализ, обучение и вземане на решения в отбраната и извън нея.

За да научите как по-ефективно да си сътрудничите с чат формата на изкуствен интелект, опитайте това упражнение:

– помолете AI да обясни абстрактна или сложна концепция по ваш избор (например „невронни мрежи“);

– след това поискайте да преработи обяснението, за да пасне на конкретна аудитория, като уточните тона и стила (например дете);

– след това аудиторията и целта се променят отново (например предприемач, за да мотивира екипа си);

– накрая, помолете ИИ да анализира трите различни обяснения, които е предоставил, и да предложи как тези различни гледни точки могат да обогатят цялостното ви разбиране за концепцията.

Днес начинът на работа на хората постепенно еволюира от писане на единичен промпт към проектиране на цял контекст и система около модела, затова промпт инженерингът постепенно се разширява и прелива в по-широка дисциплина наречена контекст инженеринг. **Контекст инженеринг** е системното проектиране, събиране и управление на информацията, която влиза в контекстовия прозореца на модела, за да изпълнява надеждно сложни задачи. Контекст инженерингът е изкуството и науката за запълване на контекстния прозорец с точно необходимата информация за следващата стъпка. Всичко, което моделът казва в резултат, е функция на този контекст. Това обхваща не само промта, а цялата информационна система, около модела, в която влизат:

– системния промпт и настройки;

– потребителското съобщение, текущият въпрос и историята на разговора;

– извличане на релевантни документи (RAG, външни източници, бази знания);

– обработка на контекста (филтриране, обобщаване, вграждане);

– управление на памет (история на профила, предишни сесии, памет, настройки на акаунта);

– интеграция с инструменти и агенти (API, конектори, резултати от търсене, кодове);

– текущ статус на задачата.

Ако контекстът е шумен, непълен или лошо структуриран, то качеството на отговора се влошава. Ако контекстът е прецизен, уместен и организиран, моделът се държи интелигентно. Добрият контекст инженеринг

намалява шума и увеличава съотношението сигнал – шум, намалява разходите и латентността, с по-малко токени се дават по-бързи отговори и по-ниски такси за API. Той също така спира халюцинациите и повтарянето на въпроси.

Тъй като контекстният прозорец е ограничен (от 8K до 200K+ токена), трябва да прилагаме инженерни методи:

– обобщаване на историята – вместо да предавате 100 съобщения, ги заменете с един компресиран списък от ключови факти;

– евристично филтриране – премахване на безполезни съобщения като „Здравей“, „Благодаря“, „Ок“;

– филтриране по актуалност – пазене само на последните N съобщения;

– филтриране по сходство – запазване само на тези части от историята, които са семантично свързани с текущия въпрос.

Целта е да се „управлява ограниченото внимание на модела“ – какво точно да има в контекст прозореца във всеки момент. Така контекст инженеринг включва промпт инженеринг като поддисциплина (как да формулираш инструкции), но добавя архитектура, данни, памет, сигурност.



Фигура 57. Промпт инженеринг и контекст инженеринг

Универсални промптове и примери за работа с Grok, Gemini и ChatGPT:

1. **Създаване на интерактивни инструменти и приложения**. Промт: Ти си софтуерен архитект, специализиран в no-code и low-code системи, използващи HTML, JavaScript и уеб API. Проектирай напълно функционален интерактивен инструмент въз основа на тази идея: [описание на идея]. Осигури подредбата на интерфейса, логическия поток и инструкциите стъпка по стъпка за изпълнение. Инструментът да отговаря на всички изисквания за сигурност, стандарта за достъпност и да използва schema.org.

2. **Превръщане на всяка концепция в инфографика**. Промпт: Ти си отличен визуален дизайнер. Превърни тази концепция във визуална инфографика с Mermaid.js, SVG или HTML или друг формат, базиран на код: [концепция]. Направи блок схема, времева линия, концептуална карта или дърво на решения, когато е подходящо, включи прост надпис и осигури чист, готов за копиране код, който е минималистичен, четим и готов за презентация.

3. **Създаване на готова за споделяне уеб презентации с отзивчив HTML**. Промпт: Действай като консултант по стратегия, проектирайки уеб презентация на тема [вмъкнете тема]. Използвай принципа на пирамидата Проблем → Прозрение → Препоръка. Създай 10 кратки слайда, всеки с ясно заглавие и 3 – 5 точки. Включи адаптивен HTML/CSS за мобилни устройства. Фокусирай се върху яснота, анализ и дизайн, готов за ръководители, като презентация на МакКинси.

4. **Направете го преподавател и учете със собствено темпо**. Промпт: Ти си световноизвестен частен учител. Научи ме на [тема] като мотивиран начинаещ. Осигури структуриран учебен план в достъпни секции, включи аналогии и обяснения стъпка по стъпка, тренировъчни въпроси с отговори и кратки резюмета след всяка секция. Използвай топъл, сократически тон и направи пауза, за да се увериш, че съм разбрал, преди да продължиш.

5. **Създаване на съдържание** – бюлетини, теми, YouTube скриптове. Промпт: Ти си топ стратег по съдържание и копирайтър. Напиши [тип съдържание: бюлетин, пост за Facebook, YouTube скрипт и др.] на [тема]. Нека да съответства на тона и стила на [създател или бранд], използвай „куки“, плавни преходи и емоционални тригери, направи го приложим и разговорен, предложи заглавие и завърши със силен CTA (Call to Action), което на български се превежда като „Призив за действие“. Финалът трябва да се усеща човешки и емоционален, а не генериран от изкуствен интелект.

6. **Писане на дълги доклади и книги**. Промпт: Вие сте старши консултант, който пише бяла книга за технологично грамотна аудитория. Тема: [тема]. Включете изпълнително резюме, пазарни тенденции с данни,

ключови предизвикателства, бъдещи прогнози, казуси и финален призив за действие. Използвайте ясен, убедителен език и подкрепяйте твърденията с доказателства или примери.

7. **Обобщаване на дълги документи в PDF**. Промпт: Ти си старши анализатор, опитен в техническите и академичните документи. Обобщи приложения документ, фокусирайки се върху ключови открития, критични данни и стратегически последици. Използвай удебелени заглавия, точки и ясен език. Завърши с препоръчителни следващи стъпки или ключови решения. Целева дължина – една страница.

8. **Асистент за вземане на решения**. Промпт: Действай като стратегически бизнес съветник. Оцени това решение: [опишете проблема или бизнес идеята]. Направи SWOT или анализ на риск и ползи, ключови потребителски профили или пазарни сегменти и картографирай възможни пътища с плюсове, минуси и препоръчани действия. Задавай уточняващи въпроси, ако е необходимо. Мисли като партньор във фирма за рисков капитал или стартъп студио. Предоставяй приложими препоръки и сравнителни данни строго в кратки точки и таблици.

9. **Генератор на бизнес идеи**. Промпт: Предложи 5 бизнес идеи, базирани на моите интереси: [вашите интереси]. Направи ги модерни, дигитално ориентирани и осъществими за самостоятелно изпълнение. Генерирай 5 уникални, запомнящи се брандови имена за този бизнес, който решава [проблем] за [целевата аудитория]. Обясни логиката за всяко име.

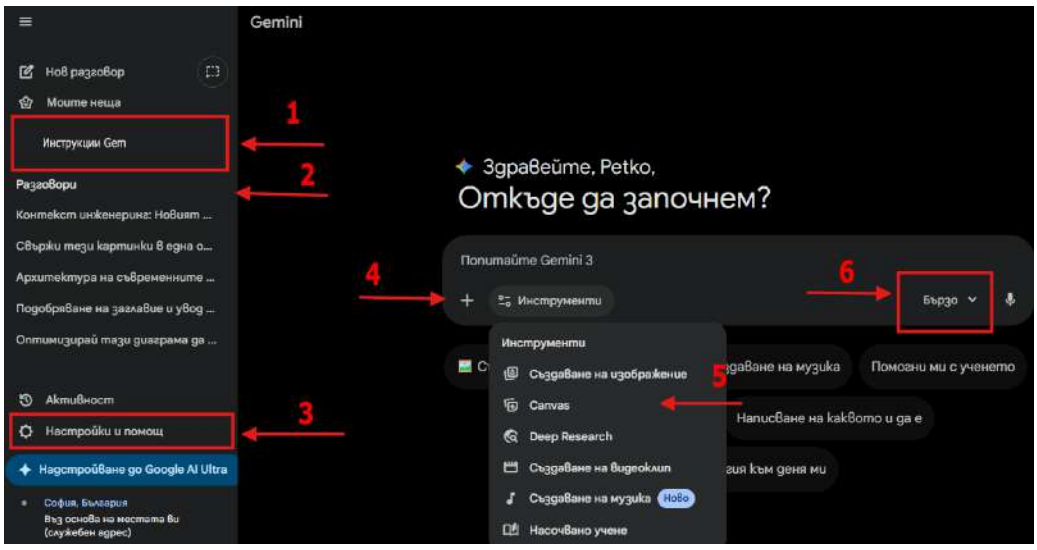
10. **Проект на календар за съдържание**. Промпт: Анализирай съдържанието ми във Facebook [сложете съдържанието] и ми дай препоръки за подобряване. След това изготви месечен календар за съдържание за промотиране на [бизнес или продукт] във Facebook. Включи съответните теми и честота на публикуване. В предлаганите публикации генерирай високоефективни куки – привличащ вниманието елемент.

Познаването и прилагането на техниките за промпт инженеринг, zero-shot и few-shot обучение, мисловни вериги и фина настройка позволява създаването на ефективни, мащабируеми и целево ориентирани AI приложения. Чрез прецизно управление на контекста, задачата и формата на отговора взаимодействието с генеративния изкуствен интелект може да бъде превърнато в мощен инструмент за анализ, обучение и вземане на решения в отбраната и извън нея.

ПРАКТИКИ ЗА РАБОТА С AI ИНСТРУМЕНТИ ЗА СЪЗДАВАНЕ И УПРАВЛЕНИЕ НА ДИГИТАЛНО СЪДЪРЖАНИЕ

С появата на изкуствения интелект драматично се промени начинът на създаване на съдържание – текст, изображения, видео и аудио – както и начинът на работа в интернет. Големите езикови модели не само променят правилата на играта, но изцяло пренаписват историята на цифровата комуникация и създаването на съдържание. Тези инструменти трансформират работните процеси, като позволяват на потребителите да генерират висококачествени аналитични и визуални материали с безпрецедентна скорост.

7.1. Интерфейс и класификация на видовете инструменти с изкуствен интелект



Фигура 58. Интерфейс за работа с Gemini

Интерфейсите на масово разпространените чатове с изкуствен интелект са интуитивни и почти всички са приблизително еднакви, базирани на чат прозорци, позволяващи лесна комуникация с модела. Тяхната работна

среда е проектирана да бъде максимално изчистена, за да може да се фокусирате върху идеите си, но под повърхността се крият мощни инструменти.

С позиция 1 на фигура 59 са показани персонализирани инструкции (Gems), предварително изготвени агенти за различни задачи от Google. В ChatGPT там са т.н. GPTs, които играят същата роля. С тях можете да създавате и свои собствени „експерти“. Вместо всеки път да обяснявате на изкуствения интелект как искате да ви отговаря, тук задавате постоянни правила (например: „Бъди лаконичен!“, „Пиши на софтуерен жаргон!“ или „Проверявай фактите два пъти!“). Може да си направите например писател на рецензии или гледачка в бъдещето.

На втора позиция се вижда историята на вашите разговори и може да прочетете всички предишни чатове. Трета позиция е много важна, защото тук е вашият „контролен панел“ с който да управлявате личните си данни, да му зададете персонализация и тон и др.

От бутон 4 със знака „+“ може да прикачвате файлове или картинки, с които да си говорите. В Gemini можете да прикачите и NotebookLM дъски, които могат да съдържат и различни прегледи на литература.

От „Инструменти“ на позиция 5 можете да сменяте агенти, специализирани за различни задачи, защото едни са по-добри за изображения, други за видео, код или за разсъждения. Canvas е специално интерактивно платно за писане на код или дълги текстове, които се визуализират на екрана ви. С него може да правите и презентации и инфографики, при отварянето му се намират още скрити бутони. Също така може да използвате Deep Research за по-задълбочено търсене и анализ на информация в интернет.

На позиция 6 обикновено се сменят моделите. Можете да превключвате между по-леки модели за бързи отговори или по-мощни версии, които се справят с по-сложни логически задачи, програмиране и креативни проекти.

Когато се включи опцията за търсене в интернет в съвременните чат приложения, те придобиват драматично нови възможности. Вместо да разчитат единствено на предварителното обучително съдържание, което често е ограничено до определена дата, тези системи вече могат да достъпват актуална информация в реално време, да проверяват факти чрез множество източници и да анализират и систематизират външни текстове. Това развитие превръща AI асистентите от опитни събеседници в истински изследователски партньори, способни да синтезират знания от хиляди източници за минути.

През последните години се появиха множество бенчмарк платформи, които тестват и сравняват AI модели в реални задачи – писане, код, дизайн, видео, уеб разработка и други. Сред най-авторитетните са Artificial Analysis,

LMSYS Chatbot Arena, DesignArena и LiveBench. Тези платформи използват Ело рейтингова система, подобна на шах, която позволява обективно сравнение на качеството на моделите в различни категории чрез „сляпо“ гласуване от реални потребители.

1. Генераторите на текст и код оперират с естествен език и са отлично решение за аналитични, образователни и административни задачи. Основните функции на тези платформи включват писане, редакция, резюмиране, анализ, превод, създаване на учебни материали и сценарии. Според данните от LMSYS Chatbot Arena и Artificial Analysis, към началото на 2026 г. най-добрите модели в тази категория са:

– ChatGPT (OpenAI) –отличава се с изключително добра производителност при задачи, изискващи креативност, маркетинг идеи, математически задачи, научен анализ и генериране на код. Функцията Deep Research позволява итеративно проучване на теми с автоматични цитирания, което е особено ценно за академична работа;

– Claude (Anthropic) – фокусиран върху сигурността и етиката и по-добрен контрол на данните. Моделът Opus 4.5 в началото на 2026 е водещ в задачи свързани с програмиране, интеграцията на MCP позволява свързване с външни инструменти и бази данни;

– Gemini (Google) – мултимодален (едновременно за текст, снимки и видео), силно интегриран в екосистемата на Google, предоставя лесна работа с документи и търсене. Интеграцията с NotebookLM позволява работа с множество източници едновременно;

– GLM 4.7. – отличен за backend логика и оптимизация на алгоритми. Работи стабилно при по-дълги кодови блокове и техническа документация;

– Kimi.ai – китайско мултимодално решение, което може дори да работи като агент;

– Mistral – европейски модел с отворен код и висока производителност;

– YandexGPT и Alice AI LLM – ориентирани към рускоезични среди;

– Deepseek е китайско OCR решение, което събира Wikepedia в една снимка.

2. Генераторите на изображения работят чрез преобразуване на текстови описания (промптове) във визуално съдържание. Интерфейсите им са базирани на въвеждане на промпт и визуализация на резултатите. През последните години качеството на генерираните изображения достигна ниво, което често е неразличимо от професионална фотография или илюстрация.

Според Artificial Analysis Image Leaderboard, към 2026 г., най-добрите модели за генериране на изображения са:

– Gemini Nano Banana – този модел позволява запазването на лицата и детайлите на субекта при генериране на безкрайни вариации – ключова способност за създаване на поредици от изображения с един и същ персонаж;

– MiniMax M2.5 – силен при фотореализъм и детайлни сцени;

– Midjourney е AI система, достъпна чрез Discord, която се отличава с изключително художествен и стилово богат финал. Тя е предпочитана за генериране на фотореалистични изображения и концептуално изкуство. Въпреки че е платена услуга, Midjourney предлага безплатен достъп чрез интеграцията с Meta AI – потребителите могат да достъпят meta.ai чрез своя Facebook или Instagram профил и да генерират изображения с качество на Midjourney безплатно. Това споразумение между Meta и Midjourney значително демократизира достъпа до висококачествено AI изкуство;

– DALL·E (OpenAI) – позволява генериране на илюстрации, схеми и концептуални изображения. Може да се достъпи чрез *Bing Image Creator*[16] и Microsoft Designer;

– Ideogram[17] – създава картинки с красиви надписи;

– Canva Magic Studio[18] е цялостно професионално решение за дизайн.

3. Генераторите на видео и аудио предоставят интуитивни интерфейси за създаване на видео и аудио съдържание на база текстови заявки. Тази категория претърпя най-бързо развитие през последните години, като качеството на генерираните клипове достигна ниво, пригодно за професионална употреба:

– Grok Imagine Video на xAI е изключително привлекателна възможност заради безплатния достъп – потребителите могат да генерират видео до 15 секунди безплатно директно в платформата X (Twitter). Той е идеален за бързо създаване на видео съдържание за социални мрежи и TikTok;

– Gemini 3 VEO3 – предлага силно придържане към промпта, реалистична физика и автоматично генериране на синхронизиран звук. Моделът е идеален за създаване на маркетингови видеа, образователно съдържание и за социални мрежи;

---

16. Достъпно на: https://www.bing.com/images/create/ai-image-generator?FORM=IRPGEN.
17. Достъпно на: https://ideogram.ai/.
18. Достъпно на: https://www.canva.com/.

– Sora (OpenAI) – видеоклипове с висока разделителна способност до 1080p, спазващи законите на физиката – от плаваемост до гравитация. Уникалната характеристика на Sora 2 е способността да генерира контекстуално съобразен звук – диалог, амбиентен шум и звукови ефекти, перфектно синхронизирани с видеото;

– Kling 3.0 на Kuaishou за реалистични движения;

– Runway е творческа платформа с изкуствен интелект, създадена за разширяване на границите на визуалното разказване на истории. Тя е особено мощна в генерирането на видеоклипове с професионално качество от текстови подкани или изображения. Runway Gen-3 Alpha позволява прецизно управление на камерата и осветлението, което е критично за създаване на реалистични симулации. Платформата предлага функции като Text-to-Video (генериране на кинематографични сцени от текст), Image-to-Video (анимиране на статични изображения) и Video-to-Video (промяна на стила на съществуващо видео);

– HeyGen – видео с виртуални аватари и синтезиран говор;

– Dreamina Seedance 2.0[19] и Pika Labs[20];

– Elevenlabs[21] – клонинг на глас, но може да се използва от текст към звук през Speechma[22] или Smodin AI[23];

– SUNO е приложение за създаване на музика.

4. Инструменти за продуктивност, автоматизация и работа с документи – съвременните AI инструменти не просто улесняват работата с документи и създаването на дигитално съдържание, но фундаментално променят начина, по който работим, учим и представяме информация. Те отварят нови хоризонти за визуализация на знания, автоматизация и креативно взаимодействие с данни:

– Make.com, Zapier, IFTTT – предлагат автоматизация на процеси, интеграция между системи;

– Perplexity е нов вид търсачка и AI заместник на Wikipedia, която комбинира възможностите на голям езиков модел с търсене в интернет в реално време. За разлика от традиционните търсачки, Perplexity не просто връща списък с връзки, а синтезира информация от множество източници в

19. Достъпно на: https://www.capcut.com/ai-creator/start.
20. Достъпно на: https://pika-labs.ru/.
21. Достъпно на: https://elevenlabs.io/.
22. Достъпно на: https://speechma.com/.
23. Достъпно на: https://smodin.io/.

структуриран отговор с цитирания. Това я прави идеална за бързо проучване на теми и откриване на актуална информация;

– NotebookLM[24] – за работа с множество документи, създаване на мисловни карти, видео и аудио презентации;

– Gamma AI е специализирано приложение за създаване на презентации и уебсайтове за минути. Платформата се превърна не само в инструмент за презентации, но и в инструмент за безплатно създаване на всякакви уебсайтове. Потребителите могат да копират всеки случаен сайт – просто трябва да копират и поставят връзката, без да е необходимо кодиране. Gamma използва AI модели като Flux Pro, Imagen 3, Ideogram и Luma Photon за генериране на съдържание, а целият процес отнема по-малко от минута;

– MyLens създава лесно диаграми и визуализации;

– математически AI[25];

– Rows.com – AI Excel;

– Vibe coding с Bolt.new и lovable.dev.

7.2. Примери от практиката за създаване на съдържание

Безплатните AI модели през 2026 г. вече покриват почти всички нужди за дигитално съдържание – текст, код, дизайн, видео, презентации и автоматизация. Ще разгледаме някои практически примери и упътвания за работа с тези инструменти.

Deep Research – златният стандарт за анализ. Когато потребителят активира дълбоко търсене в ChatGPT, Gemini или Perplexity, моделът не просто връща първия намерен резултат, а изпълнява итеративен процес на проучване. Системата декомпозира основния въпрос на подвъпроси, извършва десетки търсения, чете стотици източници и синтезира информацията в структуриран доклад с автоматични цитирания. Пример за заявка за дълбоко търсене: *Извърши дълбоко изследване на логистичните предизвикателства при разгръщане на многонационални сили в арктическия регион. Фокусирай се върху климатичните условия и транспортната инфраструктура. Върни доклад със секции за текущи доктрини и научни цитирания. Поставяй цитиранията в скоби в текста и формирай литературата по БДС.*

Пример за използване на изображения и карти за военни игри. Освен изображения за нуждите на военните игри, с AI може да се генерират

[24] Достъпно на: https://notebooklm.google/.
[25] Достъпно на: https://www.symbolab.com/.

и топографски карти и теренни модели. Например с платформата OpenArt AI Topology Map Generator може да се създаде детайлен терен с фокус върху контурни линии, надморска височина и географски характеристики. С FlyPix AI може да се прави геопространствен анализ и да се използва AI за извличане на обекти от сателитни снимки и създаване на 3D модели на терена. ArcGIS Topographic Mapping е професионално решение от Esri, което автоматизира производството на карти и данни съгласно военни спецификации като MGCP и TDS.

Създаване на AI карикатура. С помощта на AI генератор като Gemini, Meta.AI, Grok, Leonardo.ai или Bing Image Creator можете да създадете дигитална карикатура по избрана снимка. Прикачвате снимката в интерфейса и подавате промпт подобен на: „Smiling man with short brown hair and beard, holding a big carp in hands, wearing green fishing outfit, standing by a mountain lake with fishing rod, cartoon caricature style". Можете да персонализирате текста според вашата специфика – пол, професия (фризьор, рибар, сладкар), аксесоари, изражение и подходяща обстановка. Достатъчно е да качите снимка и да опишете
желаното изображение – само за секунди ще получите дигитална карикатура.

* Примерът е даден на английски език, тъй като все още промтът написан на този език дава по-добри резултати от този на български.

Създаване на AI аватар. В последно време много актуално е създаването на AI аватар (AI клонинг) с чиято помощ можете да направите дори видео, без да стоите пред камерата. AI аватарът е персонализирана версия на човека за социалните мрежи, която мисли, пише и комуникира точно като вас. С него ще можете лесно да преобразите профила си в социалните мрежи в бизнес стил нап- ример. Той дава възможност да пишете постове, имейли и оферти във ваш стил, отговаря на клиенти, генерирате съдържание напълно съобразено с вашия тон и изразяване. Такива аватари могат да се генерират с HeyGen, след което може да се свалят във формат MP4 и съхранят в отделна папка или в Instagram, TikTok, Facebook или LinkedIn, за да бъде ангажирана аудиторията с новия виртуален образ. В допълнение може да копирате гласа си с ElevenLabs и да го синхронизирате с генерираното видео на вашия аватар, за да говори той с вашия глас.



Примерен промт за създаване на мъжки аватар: *Studio fashion portrait of a confident, well-groomed man standing against a deep emerald green background. He wears a perfectly tailored dark olive green suit with a black shirt underneath, subtle pocket square, and metallic buttons. His right hand is raised to his face in a thoughtful pose, showcasing a bold, luxury skeleton-style wristwatch with an intricate open-dial design and brushed metal strap. textured crop fade, and his facial expression is serious and contemplative, with furrowed brows and a strong jawline. Lighting is soft yet directional, highlighting the texture of the suit and casting defined shadows for depth. Editorial style - elegant, masculine, refined, with a modern luxury brand aesthetic.*

Примерен промт за създаване на изображение за дамски аватар: *A high-fashion black and white portrait of a stunning young woman with flawless skin, soft lighting, and dramatic shadows. She has long, dark hair falling over her shoulders, with defined cheekbones. Her eyebrows are perfectly shaped, eyes highlighted with elegant eyeliner, and lips slightly parted with a natural gloss. She is looking directly at the camera with a confident yet sensual expression. One hand is gently touching the side of her face. The composition emphasizes her glowing skin texture and symmetrical features. The background is minimal and softly blurred, with a diagonal beam of light casting artistic contrast across her face. Style: cinematic, editorial, elegant, timeless*.

След това може да анимирате направения аватар със следните подкани: *Анимирай качения черно-бял студиен портрет на жена в сако. Запази нейната идентичност, осветление, стил на сивото и композиция абсолютно еднакви. Камерата трябва да остане статична. Добави само фини естествени микродвижения на лицето. Основно действие: плавно намигване с дясното око (~ 0,5 секунди) с лека приятелска полуусмивка и леко накланяне на главата, след което връщане в неутрално положение. Времетраене: 0 – 1 секунда – неутрално; 1 – 2 секунда – намигване; 2 – 3 секунда – връщане; 3 – 4 секунда задържане. Нежно, реалистично, кинематографично движение. Направи го готово за циклично заснемане*.

Промпт за анимация на изображение с мъж: *Анимирай качения студиен портрет на мъж в тъмен костюм. Запази идентичността, острата светлина и професионалния му стил абсолютно същите. Камерата трябва да остане статична. Добави само фини естествени микродвижения на лицето. Основно действие: уверено намигване с дясното око (~0,5 секунди) с лека усмивка (полуусмивка), като същевременно запазва силната, сериозна поза, след което се връща в неутрално положение. Времетраене: 0-1 секунди неутрално, 1-2 секунди намигване, 2-3 секунди връщане, 3-4 секунди*

задържане. Реалистично, кинематографично, естествено движение. Направете го готово за циклично заснемане.

Пример за преглед на литература и научен анализ. През последните години инструментите за изкуствен интелект се превърнаха в нещо повече от „асистенти“. Те започнаха да променят начина, по който всъщност правим изследвания. Сред тях NotebookLM на Google тихомълком си е създал пространство като истински академичен партньор, а не просто като още един двигател за обобщаване. Той е чудесен за процеса на ранен преглед на литературата, защото позволява да идентифициране на концепции и бързо получаване на базово разбиране от голяма колекция от статии. Той е подходящ и като дългосрочен партньор за изследвания, тъй като можете да задавате въпроси за всички PDF файлове едновременно, а резултатите се отнасят до конкретни пасажи в тези PDF файлове, което позволява много дълбоко семантично търсене и лесен начин за проверка дали AI е направил някакви грешки.

NotebookLM е иновативен инструмент, създаден за улесняване на работата с разнообразни документи и информационни източници. Той предоставя няколко основни функционалности, трансформиращи начина, по който се обработва и представя информацията, особено при анализ и обобщаване на големи обеми данни. Той може да анализира големи масиви от информация, представена като колекция от различни PDF файлове, URL адреси или видеоклипове в YouTube. Той предоставя допълнителни инструменти за извличане на информацията чрез аудио и видео прегледи, мисловни карти, времеви линии и персонализирани отчети, дори може да прави презентации, които да изтеглите на PDF. Това ви позволява да се справите с всеки сложен проблем (като преглед на литературата) от различни ъгли. Изненадващо, безплатната версия често е достатъчна за студенти, работещи по задача. Една от отличаващите се възможности на NotebookLM е генерирането на видео прегледи. Тази функция позволява на потребителя да получи обяснителни видеоклипове, структурирани като слайдшоута, съдържащи разнообразни графики, емоджита, цветове, таблици и цитати. Крайният резултат е интерактивно слайдшоу, което обединява основните идеи, графични елементи и цитати от оригиналния материал.

Мисловните карти идентифицират повтарящи се теми във вашите източници. Това е особено полезно за студенти и млади изследователи, които се занимават с нова тема. Ако щракнете върху подтема в мисловната карта, NotebookLM първо ще генерира обобщение на AI на тази конкретна подтема, което дава малко повече подробности за какво става въпрос. Този преглед на

AI ще съдържа препратки към отделните статии, върху които можете да кликнете, за да направите подробен преглед. Тези три нива на детайлност правят много лесно ориентирането в голям обем от знания.

Времевата линия анализира вашите източници и съдържанието в него, и начертава как знанията са се развили през годините. Това може да бъде особено полезно за хора, изучаващи философия, история или политика или за учени, които пишат изчерпателен синтез за еволюцията на въпросите за продължителни периоди.

NotebookLM се утвърждава като ценен и ефективен инструмент за обработка, организация и визуализация на информация от различни документи и източници. Той улеснява научното изследване и ежедневната работа с документи, като спестява време и предоставя нови начини за възприемане и представяне на знанията.

За да се направи преглед на литература за дисертация или научна статия е трудоемък процес, изискващ прецизност, аналитично мислене и синтез на огромни масиви от информация. С помощта на AI инструменти като **NotebookLM** и **Perplexity** може значително да се ускори и структурира това проучване.

Perplexity е оптимизиран за откриване, валидиране на факти и намиране на актуална информация в реално време чрез достъп до уеб и академични търсачки. С негова помощ може лесно да бъде открита и картографирана съответната научна област. За целта той следва да бъде инструктиран да потърси в академични бази данни (като Google Scholar, Semantic Scholar) ключови публикации от последните пет години, свързани с темата. След това – да групира литературата и подходите на изследване по основни тематични области в структуриран план. Въз основа на намерената литература трябва да идентифицира и обобщи ясно документирани изследователски пропуски и неизяснени въпроси.

На следващ етап може да направите критичен синтези и дълбок анализ с NotebookLM, който може да разсъждава върху затворена система от източници (качени от вас конкретни научни статии и книги). Това елиминира риска от „халюцинации“. Дайте му следния промт: „*Въз основа на селектираните източници, идентифицирай основните концептуални рамки и конкуриращи се гледни точки и обобщи аргументите в подкрепа на всяка страна. Дай основните ограничения на досегашните изследвания, намери неотговорните въпроси и бъдещите изследователски насоки, маркирани от авторите. Преведи идентифицираната изследователска празнина относно темата в ясни изследователски цели и формулирай проверими хипотези“*.

Добре е да не се забравя, че ИИ не мисли вместо човека. Той не отменя отговорността ви, но е мощен асистент, който може да ви помогне да извършите проучването си по-бързо и по-качествено.

Пример за създаване на презентации. Още един безплатен начин за създаване на презентации с AI е използването на приложението Gamma AI. Това е не само в инструмент за презентации, но и за безплатно създаване на всякакви уебсайтове за минути. Дава възможност и за копиране на всеки случаен сайт, просто трябва да копирате и поставите връзката, не е необходимо кодиране. Това става като се регистрирате в приложението и натиснете бутона „Create New“. След това може да изберете между опциите да импортирате файл от вашия компютър, да изберете URL или да поставите връзката на уебсайта, който искате да репликирате. След като импортирането приключи, изберете „Уеб страница“, за да продължите. Изберете AI модели за генериране на съдържание – Flux Pro, Imagen 3, Ideogram и Luma Photon, и накрая генерирането ще отнеме по-малко от 1 минута. Системата позволява да персонализирате сайта или презентацията, като промените оформлението, таблиците, графиките и други, но може да ги редактирате и с AI. Също така може да добавите персонализиран домейн или да публикувате сайта си директно в Gamma.

Съвременните AI инструменти не просто улесняват работата с документи и създаването на дигитално съдържание, но фундаментално променят начина, по който работим, учим и представяме информация. Те отварят нови хоризонти за визуализация на знания, автоматизация и креативно взаимодействие с данни, като правят сложните процеси по-достъпни и интуитивни дори за потребители без специализирани умения. Те позволяват на специалистите в отбраната да се фокусират върху стратегията и вземането на решения, докато AI поема тежестта на събирането на данни и визуалното рендиране. Тази демократизация на технологиите подготвя почвата за следващата голяма стъпка – възникването на концепции като Vibe Coding, които издигат AI сътрудничеството на по-високо ниво.

ВИЗУАЛНО ПРОГРАМИРАНЕ (VIBE CODING)

През 2025 г. възникна концепцията за визуалното програмиране – Vibe Coding, популяризирана от Андрей Карпати (Karpathy 2025). Това е екстремна форма на промпт инженеринг в софтуерната разработка, при която програмистът престава да пише код и изцяло се доверява на „вибрацията“ или интуитивното взаимодействие с ИИ агенти. Това дава възможност на потребители без задълбочени знания по програмиране да разработват приложения, прототипи и автоматизации. Макар визуалното програмиране да ускорява прототипирането, той носи рискове от натрупване на „невидим“ технически дълг и пропуски в сигурността, тъй като разработчикът често приема генерирания код без задълбочено разбиране. Тази тенденция обаче подчертава бъдещата роля на промпт инженеринга като основен интерфейс между човешката креативност и изчислителната мощ.

За да се анализира мащабът на трансформацията, предложена от Vibe Coding, е необходимо първо да се дефинират основите на традиционната изчислителна разработка. По същество програмирането е даване на инструкции на компютъра. Това обикновено става с различните програмни езици, които подават подредени една след друга команди в последователност наречена source code (скрипт). От своя страна компютърните програми са приложения с пакетирани много на брой файлове с код. Обикновено програмистите ползват среди за писане, наречени IDEs (Integrated Development Environments). Тези среди интегрират текстови редактори, инструменти за автоматизация на компилацията и дебъгери, осигурявайки цялостна екосистема за манипулация на логическите структури.

Приложенията могат да работят локално или да бъдат хоствани в интернет. За разлика от нормалното програмиране, концепцията за визуално програмиране се базира на създаването на код чрез поредица от промтове в чат с изкуствен интелект. Този преход бележи изместването на фокуса от синтактичната прецизност към семантичното намерение. Докато традиционният програмист е архитект и зидар едновременно, визуалният програмист действа като режисьор, който насочва „актьорския състав“ от ИИ агенти към постигане на крайната цел, често без да се докосва до суровите материали (кода).

Появата на визуалното програмиране не е изолирано събитие, а кулминация на развитието на големите езикови модели и тяхната способност да

разбират и генерират програмен код. Терминът беше официално въведен от Андрей Карпати през февруари 2025 г. в публикация в социалната мрежа X (Twitter). Карпати, заемал ключови позиции в Tesla и OpenAI, описа своя опит, при който той напълно се предава на „вибрациите“, като приема всички предложения на ИИ, без да чете разликите в кода (diffs), и поправяйки грешките чрез просто копиране на съобщенията за грешка обратно в чата.

Тази „екстремна форма“ на програмиране се базира на допускането, че съвременните модели като Claude 3.5 Sonnet или Gemini 2.5 Pro са достигнали ниво на компетентност, което позволява на хората да се абстрахират от изпълнението. Карпати подчертава, че това не е просто писане на код, а нов тип „програмируем слой на абстракция“, включващ агенти, под-агенти, контекстни прозорци и протоколи за оперативна съвместимост като Model Context Protocol (MCP).

За целите на визуалното програмиране вече са създадени специализирани облачно базирани платформи, които автоматизират целия жизнен цикъл на софтуера – от идеята до внедряването (deployment). Към тях спадат Bolt, Lovable, Replit, V0 и др..

Bolt.new, разработен от StackBlitz, се превърна в еталон за модена платформа. Той представлява цялостно решение за внедряване, което използва технологията WebContainers за стартиране на Node.js среда директно в браузъра. С един клик потребителят може да построи приложение, да го публикува в интернет с безплатен хостинг на субдомейн (например .bolt.host) и дори да конфигурира външни или собствени бази данни, като SupaBase. Техническите параметри на Bolt.new включват:

– рамки и езици – платформата работи предимно с React и TypeScript, осигурявайки модерна основа за уеб приложения;

– интеграция с бази данни – поддържа автоматично генериране на схеми за Supabase, включително автентификация, съхранение на файлове и Edge функции;

– ИИ агенти – позволява използването на Claude Agent (базиран на модели като Claude 3.5 Sonnet или Claude 4.5) за висококачествени изработки;

– ценова политика – безплатният план предоставя 1 милион токена месечно, което при текущите модели на консумация позволява около 5 – 10 сложни промпта на ден, тъй като всеки следващ промпт в една сесия изпраща целия актуален контекст на файловата система обратно към модела.

Lovable е друга платформа за визуално програмиране, която позволява създаване на уеб приложения и интеграции чрез блокови модули и ИИ асистенти. Потребителят може да изгради функционален продукт без да е

професионален програмист. Процесът е максимално опростен – създавате акаунт, избирате шаблон, използвате drag-and-drop блокове, за да добавите форма за вход, таблица с данни и тествате приложението си в браузъра. Специфични характеристики на Lovable:

– Visual Edits – уникална функция, която позволява на потребителите да редактират дизайна визуално (подобно на Figma), докато ИИ автоматично актуализира съответния JSX код, без да консумира допълнителни кредити за промптове;

– Full-Stack възможности – платформата генерира React + Tailwind фронтенд и автоматично конфигурира бекенд в Supabase с PostgreSQL база данни;

– целеви потребители – особено популярна сред стартъп основатели, дизайнери и продуктови мениджъри за бързо валидиране на идеи (MVP).

Таблица 7. Разлика между Bolt и Lovable

Функция	***Bolt.new***	***Lovable***
Основна технология	WebContainers/StackBlitz	AST (Abstract Syntax Tree) манипулация
Фронтенд рамка	React/TypeScript	React/Vite/Tailwind
Бекенд интеграция	Supabase/ Вградена база данни	Supabase (Database, Auth, Storage)
Визуално редактиране	Ограничено (базирано на код)	Пълно (Figma-like интерфейс)
Хостинг	bolt.host/Netlify	Lovable Cloud/Netlify

За разлика от облачните платформи, които целят пълна абстракция, професионалните инструменти са предназначени за интегриране на визуално програмиране в традиционните работни потоци на софтуерните инженери. Такива са Cursor, Windsurf и Cline.

Cursor е продължение на известния VS Code, който е преосмислен за работа с изкуствен интелект. Той индексира цялата кодова база на проекта,

позволявайки на модела да разбира архитектурните връзки между отделните модули. Програмистът може да вкара в него цялата техническа документация за съответните езици или за приложението, което иска да прави. Също така може да вкара изискванията на наредбата за минимална информационна сигурност, на стандарта за достъпност и когато генерира код, той да отговаря на цялата необходима документация. С помощта на този Composer Mode потребителят може да поиска промени, обхващащи десетки файлове едновременно. Това е ключов инструмент за визуално програмиране, тъй като програмистът може просто да каже „Рефакторирай целия API слой да използва gRPC вместо REST“ и Cursor ще предложи план и ще изпълни промените.

Windsurf е нов претендент в пространството на AI IDE, разработен от Codeium. Неговата отличителна черта е функцията Cascade, която осигурява висока степен на автономност при решаване на задачи. За разлика от други инструменти, които изискват ръчно добавяне на файлове към чата, Cascade автоматично намира подходящия контекст в големи репозиториуми. Може да планира, изпълнява и тества код в затворен цикъл, което го доближава до идеала за „автономен колега“.

Cline (предишно име *Claude Dev*) е разширение с отворен код за VS Code, което позволява на потребителите да използват собствени API ключове от доставчици като OpenRouter, Anthropic или локални модели през Ollama. То предлага най-голяма прозрачност по отношение на разходите и използваните токени.

Claude Code е CLI (Command Line Interface) инструмент от Anthropic, който позволява Vibe Coding директно в терминала. Той е изключително бърз и има директен достъп до файловата система и инструментите за тестване, което го прави идеален за комплексни архитектурни рефакторизации. Моделите от серията Claude се считат за „златен стандарт“ поради тяхната способност за нюансирано разсъждение и спазване на сложни архитектурни решения.

DeepSeek също се превърнаха в избор номер едно за мнозина след пускане на Thinking Mode (хибридно разсъждение), което му позволява да решава задачи от международната олимпиада по информатика.

Gemini (Google) предлага най-големия контекстен прозорец на пазара (до 2 милиона токена), което е критично за визуално програмиране в проекти с хиляди файлове. Едно от най-добрите решения за целите на Vibe Coding е Google AI Studio. Това е безплатна платформа за бързо прототипиране, която позволява на разработчиците да експериментират с моделите Gemini. През

2025 г. платформата добави възможности за споделяне на екрана и видео в реално време, позволявайки на ИИ буквално да „вижда“ как потребителят взаимодейства с приложението и да предлага корекции „на живо“. Това е есенцията на Vibe Coding – диалог с машината, при който тя разбира контекста на визуалната среда.

Макар визуално програмиране да обещава 10-кратна производителност, академичната общност и индустриалните експерти алармират за сериозни скрити опасности. Такъв код създава илюзията за завършен продукт, но под повърхността се натрупват редица проблеми.

Моделите на ИИ често генерират код, който решава текущия проблем, но нарушава дългосрочната работа на проекта и се натрупват технически проблеми. Повтарящото се регенериране на функции води до наличие на неизползвани модули, дублирана логика и фрагментирани зависимости. Когато програмистът „забрави, че кодът съществува“, той губи способността си да дебъгва критични грешки, които ИИ не може да реши. Това прави проекта с времето толкова сложен и нестабилен, че трябва да бъде изцяло пренаписан.

Един от най-цитираните примери за рисковете на визуално програмиране е докладът от май 2025 г., според който в платформата Lovable са открити уязвимости в генерирания код. Изследването установява, че 170 от 1645 създадени приложения са имали пропуски, позволяващи неоторизиран достъп до лична информация на потребителите. Тъй като моделите са тренирани върху отворен код с различна степен на качество, те често предлагат решения с уязвимости към SQL инжекции, Cross-site Scripting (XSS) или неправилно конфигурирана автентикация. Освен това съществува риск ИИ да генерира сегменти от код, които са обект на защитени авторски права или специфични лицензи, което излага организациите на правни рискове.

Визуалното програмиране не е просто технологичен инструмент, а културна промяна. Фокусът на предприемачите се е преместил от „как да го построим“ към „какво да построим“ и през 2025 година 95% от стартиращите компании са започнали своето приложение с помощта на инструменти с визуално програмиране.

Таблица 8. Заплахи

Категория на риска	*Описание на заплахата*	*Потенциални последици*
Сигурност	Изтичане на данни поради невалидиран вход	Глоби по GDPR, загуба на репутация
Поддръжка	„Черна кутия“ от ИИ генериран код	Невъзможност за актуализации и фиксове
Мащабируемост	Неоптимизирана база данни и тежък фронтенд	Високи облачни разходи, бавна работа
Кадри	Намаляване на експертизата на кадрите	Липса на кадри за системно инженерство

Андрей Карпати отбелязва, че се чувства „изоставащ“ като програмист, ако не използва пълния капацитет на новите приложения. Според него професията се преструктурира – приносът на програмиста става по-оскъден и се състои в „свързване на частите“. В Google вече над 30% от новия код се пише от ИИ, а в Anthropic Дарио Амодей твърди, че Claude пише 90% от техния собствен код.

Въпреки оптимизма, съществува опасност от „когнитивно разминаване“. Когато ИИ поеме трудните задачи по писане на алгоритми, хората спират да развиват способността си за решаване на проблеми. Това може да доведе до поколение от разработчици, които са добри в оркестрацията, но безпомощни при фундаментални системни повреди. За да се минимизират рисковете, експертите предлагат хибриден модел на разработка. Използване на визуално програмиране за бързо прототипиране и демонстрации пред инвеститори, последвано от дисциплинирана проверка на критичните части. Задължителна е употребата на инструменти за анализ на сигурността и автоматична проверка на генерирания код. Програмистите следва да поддържат репозиториум с добри практики, които да могат да зададат в началния план за писане на инструмента.

Визуално програмиране представлява сблъсък между традиционната прецизност на софтуерното инженерство и експлозивната продуктивност на генеративния изкуствен интелект. През 2025 г. инструменти като Bolt.new,

Lovable и Cursor превърнаха естествения език в най-мощния програмен език в историята. Тази демократизация на софтуерното създаване позволява на всеки с идея да се превърне в строител на цифрови светове. Въпреки това, както показва анализът, се натрупват технически проблеми и уязвимости в сигурността, което изисква нов тип интуиция и по-високо ниво на надзор. Бъдещето на програмирането не е в изчезването на кода, а в неговото превръщане в „инфраструктура под наем“, управлявана чрез вибрациите на човешкото намерение. Успешните програмисти на бъдещето ще бъдат тези, които умеят да балансират между хаоса на бързото прототипиране и реда на устойчивата системна архитектура.

ОПТИМИЗАЦИЯ НА САЙТОВЕ ЗА ГЕНЕРАТИВНИ СИСТЕМИ

Появата на Vibe Coding промени и информационната инфраструктура, която се управлявана чрез вибрациите на човешкото намерение. Благодарение RAG (Retrieval-Augmented Generation) архитектурите инструменти с изкуствен интелект извличат информацията в реално време от сайтовете в интернет, а това промени и начините за управление на знанието, което тази инфраструктура консумира, синтезира и възпроизвежда.

Благодарение на тези процеси, всички уебсайтове се промениха фундаментално. Те вече имат нужда от нов тип оптимизация, защото всички AI приложения и агенти търсят информация в интернет, сканирайки уебсайтовете не като хора, а като машини. За да може съдържанието на един сайт за отбрана или сигурност да влезе в техните отговори, то трябва да може да „привлича“ ботовете на AI. В контекста на националната сигурност, оптимизацията за генеративни машини (Generative Engine Optimization – GEO) не е просто въпрос на дигитален маркетинг или видимост, това е въпрос на когнитивна сигурност и наративно господство. Когато един офицер използва AI асистент, за да синтезира доклад за геополитическата обстановка, или когато цивилен гражданин пита чатбот за легитимността на военна операция, отговорите се генерират на базата на вероятностни модели. Ако нашето съдържание – доктрини, факти, стратегически комуникации – не са оптимизирани за тези генеративни машини, то просто не съществува в техния „свят“ или бива изместено от „халюцинации“ и враждебни наративи.

Класическата оптимизация за търсачки (SEO) е фундаментът, върху който стъпва съвременният информационен интернет. Тя възникна в среда, където търсачките (напр. Google, Bing в ранните им версии) действаха като библиотекари. Потребителят задава въпрос, а търсачката връща списък с документи (URL адреси), които съдържат ключовите думи от въпроса. В основата на SEO стоеше лексикалният анализ. Алгоритмите сканират мрежата (crawling), индексират съдържанието и го класират (ranking) въз основа на ключови думи, обратни връзки (backlinks) и техническа структура.

С навлизането на големите езикови модели правилата се промениха. Търсачките вече не са просто индексиращи машини. Bing интегрира Copilot, превръщайки търсенето в диалог, базиран на GPT-4, който директно отговаря на въпроси, използвайки индекса на Microsoft. Google трансформира

основната си търсачка чрез въвеждането на т.нар. AI Mode (чрез AI Overviews), който предоставя автоматично генерирани отговори. Те събират данни от множество източници и ги представят в удобен формат, без да се налага хората да следват линк за връзка, за да разберат повече. AI не търси просто съвпадение на низове, а семантична близост във векторното пространство.

Оптимизацията за генеративни машини е процесът на структуриране, създаване и техническо обезпечаване на съдържанието на сайта така, че то да бъде лесно „разбираемо“, високо оценено и коректно синтезирано от генеративни AI модели. Авторитетът се измерва с това колко често даден източник се споменава и използва за обосновка на генерирания текст. Ако AI моделът е захранен с информация, в която доминират враждебни наративи, той ще генерира отговори, които подкопават националната сигурност – автоматизирана форма на когнитивна война.

Много анализатори наричат този процес AEO (Answer Engine Optimization). При него оптимизираме за крайния резултат – директния отговор, предоставен от машината (Zero-Click т.е хората не кликат върху сайта, а четат отговора в интерфейса на изкуствения интелект). При AEO няма „второ място“ – или сте източникът на отговора или не съществувате за потребителя.

Таблица 9. SEO, GEO, AEO

Характеристика	***SEO (Търсачки)***	***GEO (Генеративен AI)***	***AEO (Системи за отговори)***
Основна единица	Ключова дума (Keyword)	Концепция/вектор	Въпрос (Query)
Резултат	Списък с линкове (SERP)	Синтезиран текст	Единичен отговор
Технически фокус	HTML структура, Meta tags	RAG формат, контекст	Schema.org, FAQ, кратки отговори

След период на несигурност в нуждата от собствена интернет страница, дължащ се на осъзнаването на многото фалшива информация и възможността на ИИ да генерира всичко автоматично, днес става ясно – ИИ търси стойностно, проверено и създадено от хора съдържание, за да се учи и да дава по-добри отговори. Създадено от човек ще се класира по-добре в AI отговорите, стига да може да се отличи от масовото автоматизирано съдържание. Ако създавате съдържание, което решава реални проблеми, има принос и подлежи на проверка, вие сте една стъпка напред. Дори използването на ИИ при създаването му не е проблем – важното е, че идеята, структурата и експертизата идват от човек.

Днес AI чатботовете обхващат и анализират огромни части от интернет. Според OpenAI техните модели обработват над 1,5 милиарда уеб страници месечно. Оптимизаторите вече се фокусират върху създаването на съдържание, което лесно може да бъде разпознато, обобщено и цитирано от AI – вместо традиционното изграждане на линкове. Ключовите фактори за успех са добре структурирани текстове, яснота на авторството и експертиза, както и използването на списъци, сравнителни таблици и класации. Например, вътрешни анализи на Moz показват, че сайтове със структурирани данни и ясно посочено авторство се цитират от AI чатботове до 3,6 пъти почесто в сравнение с неструктурирани ресурси.

Вместо обемни и многословни статии, които целят да „угодят“ на търсачките, новите тенденции налагат създаването на кратки, ясно структурирани пасажи, които AI лесно може да обобщи и цитира. Отчита се, че съдържание с експертен тон и доказан авторитет е предпочитано от AI моделите. Според проучване на SEMrush, над 42% от цитираното съдържание в AI чатбот отговори през 2024 г. е било създадено от признати експерти в съответната област.

Структурирането на информацията е критично важно – използването на списъци, таблици и логически подредени абзаци улеснява AI при обработката и представянето на съдържанието. Препоръчително е да се включват оригинални изследвания, класации и сравнителни анализи, тъй като те се цитират средно с 2,1 пъти по-често според данни на Search Engine Journal.

Ясните препоръки и заключения, формулирани в стил, наподобяващ техническа документация, увеличават шанса съдържанието да бъде интегрирано в многоетапни AI разговори. Формати като структурирани списъци и логически подредени абзаци подпомагат AI моделите да „улавят“ и разпространяват информацията. Допълнителни предимства носят споменавания в източници като Wikipedia, Reddit и YouTube видеа, които се използват

активно в обучението на AI модели – според OpenAI, над 60% от данните за обучение на GPT-4 са от такива източници.

Ако следвате тези препоръки, значително повишавате вероятността вашият сайт да бъде предпочитан източник за цитиране от AI чатботове и да останете видими в новата дигитална екосистема. Прогнозите на Gartner сочат, че до 2027 г. над 50% от онлайн трафика ще се генерира от AI платформи, а не от традиционни търсачки.

Важно е да се отбележи, че Google планира да понижава класацията на съдържание, което се определя като нискокачествено или автоматично генерирано. Експертите препоръчват изкуственият интелект да се използва основно за проучвания, а не за директно създаване на текст. В крайна сметка, всичко, което е полезно и ценно за хората, е ценно и за AI оптимизацията. Макар че трафикът и споменаванията стават по-трудни за получаване, качественото съдържание остава водещият фактор за успех в новата ера на генеративната оптимизация.

Оптимизацията за генеративни модели вече е задължителна за да можем съответния сайт да влезне като източник и връзка в Perplexity например. В статия на Netpeak бе предложен нов подход – освен познатия ни robots.txt, сайтовете трябва да използват llm.txt файл, който указва на AI ботовете какво съдържание им се позволява да използват. Идеята е да предоставим на AI агентите кураторски списък с най-важните документи в машинно-четим Markdown формат, изчистен от JavaScript и визуален шум. Това позволява на RAG системите директно да откриват и индексират правилния документ, вместо да гадаят. Секцията Blockquote в файла може да служи като „системен промпт“ за сайта, дефиниращ неговия авторитет. Това е новата форма на етично SEO – вече не се борим с Google, а обучаваме AI как да използва нашето съдържание.

Днешните AI crawl системи използват не само структурата на сайта, но и мета информацията. Можете да добавите meta тагове в <head> на сайта:

– *<meta name="ai-content" content="false">* – указва, че съдържанието е създадено от хора (повишава доверието);

– *<meta name="ai-topic" content="[h1 text]">* – помощ за AI системите да идентифицират основната тема;

– *<meta name="ai-authority" content="true">* – маркира експертни материали като надеждни източници.

Тагове като ai-summary и ai-citation също помагат на моделите да разпознаят експертното съдържание и да го цитират коректно. Тези маркери не

са задължителни, но сигнализират на новите AI агенти, че сайтът ви е достоверен и добре структуриран източник.

Използването на структурирани данни (Schema.org) също значително подобрява разбирането на AI за тип съдържание (статии, продукти и др.) и е задължително за установяване на авторитет. Критични схеми в отбраната включват GovernmentOrganization (за верификация на институцията), GovernmentService (за услуги към гражданите) и SpecialAnnouncement (за кризисни съобщения). Чрез JSON-LD ние казваме на AI недвусмислено: „Това е официалният източник“.

В хибридната война информацията остарява за минути. Протоколът IndexNow променя модела от „Pull“ (търсачката идва, когато реши) към „Push“ (ние уведомяваме търсачката веднага). Това е критично за незабавно индексиране на опровержения на дезинформация в Bing и свързаните с него AI системи (Copilot, ChatGPT Browse).

Светът на интернет, който познаваме, се променя с шеметна скорост. Вече не става въпрос само за това, че Google пусна т.нар. AI Mode в търсачката си или че Bing интегрира Copilot. Браузерите вече са новото бойно поле, където всеки ден излиза нещо ново и революционно. През 2025 г. станахме свидетели на масирана офанзива на AI от страна на големите компании:

– Perplexity пусна Comet – браузър, който първоначално бе предвиден да струва $200/месец, но е напълно безплатен към 2026 г. Comet „вижда“ кода на страницата (DOM) и действа като интелигентен слой над интернет;

– OpenAI пуснаха браузъра Atlas (първоначално за MacOS). Той интегрира ChatGPT директно в навигацията, позволявайки „Agent mode“, който действа вместо потребителя.

Тези браузъри не просто показват уебсайтовете, а действат като автономни AI агенти и лични асистенти. Те могат да навигират, обобщават, кликват и дори да купуват от ваше име. Това на практика означава, че те елиминират традиционното търсене, като вече не е необходимо потребителите да посещават сайтовете визуално.

За какво ще служат сайтовете, ако AI може да свърши всичко вместо нас? Страниците, които предоставят само обща информация, ще станат невидими. За да оцелеят, те трябва да предоставят стойност отвъд информацията с:

– уникално потребителско преживяване;

– предоставяне на полезни функционалности (инструменти, калкулатори);

– общности и интерактивни елементи;

– доверие, което AI не може да пресъздаде.

Пример за тази адаптация е партньорството на Perplexity с медии като CNN и The Washington Post – те осъзнават, че за да бъдат „прочетени“ от AI, трябва да са част от неговата верига на доверие. Това партньорство може да се използва и в борбата с фалшивите новини.

Google пуснаха новия си квантов алгоритъм Quantum Echoes, изпълняван на процесора „Willow“. Той е 13 000 пъти по-бърз от най-мощните суперкомпютри в света към този момент. Това не е просто по-бърз компютър, това е невероятна „фабрика за данни“, която ще позволи на Google да създаде ИИ, способен на невиждани пробиви в медицината (молекулярна геометрия), енергетиката и материалознанието. Обработката на данни в бъдеще ще се превърне в мигновен процес.

Появиха се и други браузъри, като Open Brawser, които могат да пазаруват от ваше име, като и имат достъп до банковите ви сметки, за да го правят. Това е огромна уязвимост и трябва да се внимава с тях. Атака чрез „скрит промпт“ в коментар на сайт може да накара вашия браузър да извърши транзакция без ваше знание. В отговор на тази заплаха се появи Norton Neo – първият браузър, който не просто има AI, а който твърди, че пази вашите лични данни и сигурност. Той съхранява информацията (история, чатове) локално на вашето устройство и може да избира какво да „забравя“. Разполага с активен щит, който засича фишинг и зловредни AI скриптове.

С появата на тези браузъри ние навлизаме в следващата голяма стъпка към новия „Интернет на агентите“ (Internet of Agents – IoA). В този нов свят нашите лични AI агенти ще преговарят с AI агентите на институциите и бизнеса. Оптимизацията вече няма да е за очите на човека, а за логиката на агента. Успешните организации в сигурността ще бъдат тези, които изградят своята дигитална инфраструктура така, че тя да бъде сигурен, авторитетен и предпочитан партньор в този безкраен диалог между машини.

АВТОНОМНИ AI АГЕНТИ И ИНТЕЛИГЕНТНА АВТОМАТИЗАЦИЯ НА ПРОЦЕСИ

По-голямата част от
дейността ни с машинното обучение
се осъществява под повърхността

Джеф Безос

В науката за изкуствения интелект понятието „агент“ се използва за всяка система (биологична, механична или софтуерна), която непрекъснато възприема средата си и действа в нея.

10.1. Автономни интелигенти агенти

Взаимодействието на агента със света около него не е случайност, а целенасочена поредица от събития. За да се анализа взаимодействието на агентите с тяхната среда, в специализираната литература се разглеждат три основни научни гледни точки за онагледяване:

– класификационен модел „Какво“ – фокусира се върху външното наблюдение какво прави агентът в дадена ситуация. Търсят се повтарящи се модели в поведението му;

– механистичен модел „Как“ – анализира какви вътрешни алгоритми, структури и процеси са нужни, за да се възпроизведе това поведение;

– епистемологичен (обяснителен) модел „Защо“ – търси логиката зад действията. Анализира защо агентът е избрал точно това действие в този конкретен момент, за да постигне целта си.

Най-често взаимодействието между агент и среда се разглежда с термините на състояния и действия. Средата, в която действа агентът, може да се намира в множество външни състояния W. Агентът притежава сензорна система за възприемане на външните състояния P (перцепция). Той взаимодейства с околната среда, като получава входни данни като събития и реагира чрез действия. Поведението се формира от елементарни правила.

Нека да вземем за пример интелигентен агент робот прахосмукачка. За него външният свят W е стаята, която почиства. Той има сензори и

перцептори *P*, които отчитат, че „има твърдо препятствие на 5 сантиметра пред него“. Тази опростена картинка, която сензорът му дава, се нарича „перцептор“. Тъй като роботът е с ограничени сензори (не вижда цялата стая наведнъж), той си „чертае“ невидима карта на стаята, за да знае къде е бил и къде не е. По този начин той прави „вътрешни състояния“ – S. В паметта си създава база знания K, в която записва информация от рода на това, че вратата е вляво или че на килимът е необходимо да използва повече мощност. Действия (A) са нещата, които роботът може да прави физически – тръгни напред, завий наляво, спри, пусни четката, за да постигне целта G, да почисти стаята. За да изчисти стаята, без да се блъска в стените, роботът върти едни и същи 5 стъпки (програмни функции):

– see – възприемане и отразяване на всяко ново външно състояние, усетено от сензорите;

– mks – роботът превежда това на своя машинен език и го записва структурирано с координати;

– tell – запомняне на полученото в базата данни;

– reas – разсъждаване според записаните правила;

– do (изпълнение) – роботът избира най-доброто действие – подава команда на колелата и завива надясно. След като завие, цикълът започва отначало (вижда какво има след завоя, превежда го, запомня го, мисли, действа...).

Математически изразено неговото поведение е функция от:

$$f\,(K, W, P, A, S, see, mks, tell, reas, do)$$

Докато класическото машинно обучение и дълбоките невронни мрежи се фокусират върху математически формули за разпознаване на модели, то след появата на генеративния изкуствен интелект в съвременния свят се формира изцяло нов слой от агенти за общуване на естествен език, които работят по съвсем различен начин. Именно тези характеристики позволяват прехода от обикновена софтуерна програма към цялостна автоматизация на работните процеси, при която агенти (или мултиагентни системи) решават сложни задачи самостоятелно. Проектирането на интелигентни агенти се свежда до два основни сценария:

1. Човек проектант вгражда декларативни и процедурни знания в агента, които се превръщат в неизменна догма за поведението му.

2. Човек проектант вгражда възможности за обучение и самообучение, чрез които агентът придобива знания на основата на собствен опит.

Конкретните задачи на проектанта се свеждат до съставяне и програмно реализиране на процедурите *see, mks, tell, reas* и *do*, както и формиране на началната база знания на агента. Когато проектантът не разполага с достатъчно знания за средата, обучението и самообучението са единственият начин агентът да придобие необходимите знания. За целта се използва поведенческа архитектура. Това е описание на онези най-съществени аспекти на взаимодействието между агента и средата, които остават постоянни, докато не бъдат променени от външен агент. Изборът на архитектура зависи съществено от свойствата на средата.

Средата може да бъде достъпна и недостъпна по отношение на разпознаване със сензорите, детерминирана и недетерминирана, статична и динамична, дискретна и непрекъсната, епизодична и неепизодична. Най-тежък е случаят на взаимодействие с недостъпни, недетерминирани, неепизодични и непрекъснати среди.



Фигура 59. Формиране на поведение на агент

Архитектурата на един автономен агент включва четири основни компонента:

1. Среда – оперативното пространство (мрежата на организацията или виртуално бойно поле).

2. Сензори и механизми за „виждане“ – мрежови скенери, анализатори на логове, камери.

3. Актьори и инструменти за действие – команди за защитна стена, софтуер за карантина, управление на дронове.

4. Когнитивен двигател – ядро за вземане на решения, базирано на правила, машинно обучение или подсилено обучение.

Съществуват следните видове поведенчески архитектури:

● Табличните агенти са най-опростеният тип, който функционира на основа на предварително запаметена таблица с подходящите действия за всички възможни последователности от входни въздействия.

● Рефлексни агенти, които реагират на формиран перцепт на основа на правила „ако (условие), то (действие)“ от базата знания, наподобяващи биологични рефлекси. Недостатъкът е липсата на памет за миналото – те функционират успешно само когато изборът на действие е възможен единствено въз основа на текущия перцепт.

● Агенти с явно зададени цели, които избират действията си така, че да ги постигнат в зависимост от принципа на рационалност. Те имат възможности за разсъждения относно бъдещето („Колко далеч съм от целта?“). Те могат да бъдат чисто логически агенти, търсещи агенти или планиращи агенти, които използват техники за генериране на последователност от действия, съчетаващи логически процедури и търсещи процедури (с предварително или текущо планиране).

● Прагматични агенти, в които целите се заменят от функция на полезност, която съпоставя на всяко състояние реално число като мяра за успеха. Това осигурява възможност за (1) оценяване на относителната важност на конфликтните цели и (2) оценяване на степента на успех при работа с несигурни знания.

● Автономни агенти са архитектурите, притежаващи възможностите на предходните агенти плюс качеството да се обучават и самообучават. Те използват перцептите не само за генериране на действия, но и за подобряване на възможностите си за по-съвършено функциониране в бъдеще. Поведението им става до известна степен автономно, определяно от собствения опит, а не само от вградените знания на проектанта.

10.2. Протоколи за оперативна съвместимост и платформи за автоматизация

Ако обърнем внимание на съвременното развитие на изкуствения интелект, можем да го разгледаме като надграждащи се слоеве от технологии и протоколи, които взаимодействат и позволяват създаване на ново съдържание. Докато чат ботовете отговарят на въпроси, то AI агентите са основният слой по отношение на действието. В този слой откриваме компонентите, които превръщат един пасивен алгоритъм в активен агент. Тук вече не говорим просто за обработка на данни, а за:

– автономно вземане на решения и целенасочено разсъждение;

– ползване на инструменти и интеграция за взаимодействие с външни системи (например търсене в интернет или да управляване на софтуер);

– управление на паметта и контекста (способността на агента да помни минали взаимодействия).



Фигура 60. Слоеве на изкуствения интелект

Появата на тези нови слоеве се дължи на постоянната конкуренция между големите производители на LLM, които в последните години буквално всеки ден пускат нова разработка на пазара, включително такива за свободно ползване. Глобалният пазар на платформи за разработка с малко код и без код нарасна значително и през 2025 г. достигна 36.06 милиарда долара[26] . С нарастването на екосистемите от разнообразни агенти тяхната оперативна съвместимост става критично предизвикателство. По тази причина бяха разработени два стандартизирани протоколи за сътрудничество между различните инструменти чрез децентрализирани мрежи.

● **Model Context Protocol (MCP)** – разработен от Anthropic през ноември 2024 г., отворен стандарт, дефиниращ универсален интерфейс за комуникация между AI модели и външни RAG системи. Функционира като „USB-C порт за AI“, позволявайки взаимодействие с локални и облачни ресурси (GitHub, Slack, PostgreSQL) по безопасен начин. Прилага многостепенен модел за сигурност с 16 ключови дейности в жизнения цикъл на сървъра, JSON-RPC 2.0 формат на съобщенията, слоеве за свързване с инструменти за извличане и кеш слой за повтарящи се заявки. Позволява plug-and-play между инструменти за извличане без преобучение, постигайки 2 – 5 пъти по-ниска латентност в производствените RAG тръбопроводи (Anthropic 2024)[27].

● **Agent-to-Agent (A2A) Protocol** – представен от Google през април 2025 г., формализира начина, по който автономните агенти се договарят за задачи и делегират отговорности. Позволява създаване на Agent Cards (дигитални „автобиографии“ на агентите), проверка на идентичността чрез TLS сертификати и цифрово подписване, проследимост на задачите и споделяне на състоянието. Формира гръбнака на „Интернет на агентите“, където дигитални работници автономно наемат други агенти, обменят данни и сключват договори за услуги.

Протоколите Model Context Protocol (MCP) и Agent-to-Agent (A2A) осигуряват стандартизираната основа за оперативна съвместимост между автономни агенти и AI инструменти, позволявайки им да комуникират и обменят данни независимо от платформата. В тази връзка за автоматизиране

[26.] GRAND VIEW RESEARCH, 2022. *Global Low-code Development Platform Market Size & Outlook. San Francisco*. CA: Grand View Research. Available from: https://www.grandviewresearch.com/horizon/outlook/low-code-development-platform-market-size/global

[27.] ANTHROPIC, 2024. *Model Context Protocol Specification*. [Online]. Available from: https://modelcontextprotocol.io/docs/getting-started/intro.

на различни дейности през последните години широко се употребяваха платформи като Microsoft Power Automate, Copilot Studio, Make.com и Zapier, всяка със специфични архитектурни особености, определящи приложимостта в различни сценарии. Тези платформи се интегрират все по-често с новите протоколи, което позволява изграждане на сложни, мащабируеми автоматизации и интелигентни агенти, адаптиращи се към разнообразни бизнес и технологични среди.

Благодарение на MCP протокола можете да свържете всеки AI чат бот директно с вашите инструменти за дизайн, база данни и всякакъв външен работен процес. Например, има плъгини, като:

– Figma MCP, който свързва Claude Code директно с инструмента за уеб дизайн Figma. Така, използвайки MCP Claude, може да чете дизайни, създадени във Figma, да изпраща написания код във Figma, да разбира компоненти, стилове, оформление и дори да променя дизайни в някои настройки;

– Playwright MCP – свързва Claude с вашия браузър. Claude вече може да отваря уебсайтове, да кликва върху бутони, да попълва формуляри, да извлича данни, да тества приложения, да навигира в страници, всичко това от естествен език;

– GitHub MCP – откриване на грешки и управление на CI/CD с помощта на естествен език;

– Supabase MCP прави заявки към бази данни и управление на схеми с едно изречение.

Едни от най-често коментираните инструменти, създаващи мостове между чат ботовете и други външни приложения, са Microsoft Power Automate, Copilot Studio, Make, N8N и Zpier.

● **Microsoft Power Automate** е интегриран в системата на Microsoft 365 и Azure, предлагайки специфични облачни среди за правителствени нужди, сертифицирани според изискванията на Министерството на отбраната на САЩ за всички нива на въздействие (DoD IL2, IL4, IL5, IL6). Състои се от:

– *Cloud Flows* – автоматизации, работещи изцяло в облака, задействана от събития или график;

– *Desktop Flows* – инструменти за автоматизация на потребителския интерфейс в наследени системи без API;

– *AI Builder* – готови модели за обработка на документи, разпознаване на обекти и анализ на настроения.

● **Microsoft Copilot Studio** представлява следващата еволюционна стъпка, преминавайки от автоматизация на задачи към създаване на интелигентни, разговорни агенти, използващи генеративни ИИ модели за разбиране на естествен език и адаптивно вземане на решения в реално време.

● **Make.com** се отличава с визуален конструктор, подобен на логическа карта, особено силен при сложни трансформации на данни и многоклонова логика, с над 2800 интеграции и по-ниска цена при големи обеми операции.

● **Zapier** е най-лесен на използване и има най-голям брой поддържани приложения (над 8000), оптимизиран за бърза настройка на прости автоматизации. В силно регулирани среди като отбраната обаче липсата на опция за локално инсталиране може да бъде пречка.

Таблица 10. Сравнителен анализ на платформи за автоматизация

Критерий	***Power Automate***	***Make.com***	***Zapier***
Брой интеграции	1000 +	2800 +	8000 +
Сигурност	IL2, IL4, IL5, IL6 на DoD	SOC 2 Type II	SOC 2 Type II
Сложност на логиката	Висока (с Azure функции)	Много висока (визуална)	Средна
Потребителски интерфейс	Дизайнер	Визуална карта	Линеен стъпка по стъпка

Посочените технологии използват три основни градивни елемента на работните потоци:

● **Тригер** – събитието, стартиращо автоматизацията. Платформите използват два модела – проверка по разписание (системата проверява периодично за нови данни) и Push реакция в реално време (източникът изпраща сигнал веднага щом настъпи събитието).

● **Действия** – изпълнителните стъпки (изпращане на имейл, актуализиране на таблица, стартиране на скрипт). Условната логика (Condition, Switch) позволява разклоняване на процеса според контекста на данните.

● **Конектори** – архитектурен мост между платформата и външното приложение, капсулиращ сложността на API заявките. Критично важно е използването на сертифицирани конектори с протоколи за сигурност като OAuth 2.0 и Microsoft Entra ID.

Тези инструменти могат да се използват за автоматизация на всевъзможни работни процеси, без умения за кодиране. Ето няколко примера за тяхната употреба:

● **Автоматизиране на имейл кореспонденция.** Чрез Power Automate се филтрират и подреждат имейли, с изпращане на съобщение в MS Teams при получаване от конкретен подател. Конфигурира се чрез конектора Office 365 Outlook с тригер When a new email arrives, дефиниране на папка и подател, допълнителни критерии (ключови думи, прикачени файлове) и действие Post message in a chat or channel от конектора Microsoft Teams с динамично съдържание.

● **Обработка на данни от таблици.** Използва се конекторът Excel Online за достъп до файлове в OneDrive или SharePoint. Действието List rows present in a table извлича записи, които чрез Compose или Office Scripts се форматират според изискванията на целевата база данни с обновяване в SQL Server или Microsoft Dataverse за елиминиране на риска от дублиране.

● **Роботизирана автоматизация на процеси (RPA).** Power Automate Desktop действа като „виртуални роботи“, имитиращи кликвания и писане в наследени системи без API, изискващи ръчно въвеждане в графичен интерфейс на локална машина. Служи например за автоматично въвеждане на данни от сканирани документи в стара ERP система чрез отваряне на приложението, влизане с идентификационни данни, навигация и попълване на полета, извлечени от ИИ модел.

● **Интелигентна обработка на документи.** AI Builder позволява интегриране на OCR (Optical Character Recognition) в комбинация с ИИ за автоматично извличане на информация от товарителници, заявки и митнически документи, включително ръкописни бележки, превръщани в структурирани данни за секунди. Системите динамично преизчисляват нива на запаси и предлагат алтернативни маршрути.

Създаването на такива чат агенти може да се класифицира според тяхната техническа сложност за потребителите.

Таблица 11. Типове агенти

Тип агент	*Платформа*	*Характеристики*
За създаване без код (No-code)	Copilot Chat / Agent Builder	Лична употреба, базирани на документи от Microsoft 365/SharePoint. Чрез инструкция се задава роля и задача на AI и му се казва какво да даде на изхода. С други думи той получава два промта един конкретен и един специализиран.
С полукод	SharePoint Agents	Базирани на специфични бази знания от корпоративни портали.
С малко код (Low-code)	Copilot Studio	Интеграция с Power Automate (1500+ действия), автономно отговаряне на имейли и други действия.
С пълен код (Full-code)	Azure AI Foundry / SDK	Фина настройка на специфични модели (Mistral, Llama, GPT-5) и персонализирана логика.

Microsoft 365 Copilot и Copilot Studio предоставят опростен интерфейс за бързо изграждане на агенти чрез естествен език или ръчна конфигурация със следните действия:

1. планиране и дефиниране на целта – например „Асистент за писане на рапорти за отпуск“;

2. конфигуриране на база знания – добавяне на източници, книги, учебници и уебсайтове, обосноваващи отговорите;

3. дефиниране на роля, задача и възможности – включително кодов интерпретатор за изчисления на Python или генератори на изображения;

4. тестване и валидация – постоянно поставяне на въпроси в реално време за проверка на точността и тона с коригиране на инструкциите

Потребителят може много лесно да направи това в интерфейса на Copilot, като влезе в раздел *Агенти* в лявото меню. Там ще види предварително подготвени специализирани агенти и такива, създадени според индивидуалните потребности. Бутонът *Създаване на нов агент* отваря форма с:

– поле Describe your agent – кратко описание (напр. „Агентът помага на обучаемите да се ориентират в учебната програма по киберсигурност");

– име на агента (до 30 знака) – напр. „Кибер наставник“;

– секция Instructions – описание на функциите (напр. „Отговаряй на въпроси за учебната програма, графика на занятията и основни термини...“). Полето за инструкции може да се използва като кодов интерпретатор за въвеждане на Python код за генериране на файлове или сложни математически изчисления;

– поле Knowledge – за поставяне на линкове към учебни материали или вътрешни документи.

През последните години се наблюдава една фундаментална промяна в развитието на технологиите за автоматизация: преход от реактивни разговорни системи към автономни AI агенти. Докато стандартният LLM функционира като квалифициран работник, изпълняващ единични задачи, AI агентът е еквивалент на „фабрика“, състояща се от множество специализирани модули, сътрудничещи си за постигане на комплексна цел.

Тези системи притежават автономност за разглеждане на уебсайтове, кликвания, API повиквания, извличане на информация и итеративно разсъждение. Съществената разлика спрямо традиционните чат системи е многоетапният процес на генериране на отговори – агентът дефинира целта, планира стъпките и ги изпълнява последователно, решавайки проблеми, надхвърлящи капацитета на единичната заявка.

10.3. Архитектурна интеграция на LLM, RAG и MCP с различни изпълнителни инструменти

Съвременният изкуствен интелект вече не се разглежда като единичен изолиран модел, а като многослойна система от взаимносвързани компоненти. Неправилното разбиране на тази система води до концептуални грешки и изграждане на неоптимални системи. За да формирате ефективно работещ, мащабируем AI агент, е необходимо да познавате както основните архитектурни стълбове, така и слоевете от инструменти, които ги свързват.

За да разберем как функционират агентите, трябва да разграничим четирите концептуални нива на тяхната архитектура:

● LLM (голям езиков модел). Това е „мозъкът“ на системата. Базиран е на трансформаторна архитектура, той разбира текст, генерира съдържание и разпознава модели. Неговото фундаментално ограничение е статичната база знания – моделът знае само това, което е усвоил по време на обучението си.

● RAG (Retrieval-Augmented Generation). Това е „мозък + външно знание“. Тази архитектура добавя външен слой от знания, който не е част от теглата на самия модел. Процесът включва преобразуване на външни данни във векторни представяния (embeddings), семантично търсене в бази данни и генериране на отговор чрез обединяване на запитването с намерения контекст. RAG решава проблема с халюцинациите, свързвайки модела с доверени източници (PDF архиви, корпоративни бази данни, API).

● AI Агенти. Те представляват „мозък + ръце“. Агентите използват външни инструменти за трансформиране на генерирания текст в реални, изпълними действия. ИИ Агентите преминават отвъд концепцията за чатбот, превръщайки се в автономни работници, които могат да четат имейли, да управляват процеси и да пишат код.

● MCP (Model Context Protocol). Това е „нервната система“. MCP стандартизира сигурната и мащабируема комуникация между езиковия модел и инструментите, които той използва. Без подобен протокол системите не могат да оперират сигурно в производствена среда.

Ако искате да сглобите собствен агент, особено използвайки най-новите безплатни приложения и инструменти с отворен код, трябва да преминете през показаните в таблица 12 седем слоя на технологичния стек.

Таблица 12. Слой и функция на системата

Слой	***Функция в системата***	***Примери за безплатни/отворени инструменти***
1. Модели (LLMs)	Логическо ядро. Индустрията преминава към „моделен плурализъм“ – избор на модел според нуждите (цена, скорост, точност)	Llama, Qwen, DeepSeek
2. Отворен достъп (Gateways)	Демократизиране на достъпа до моделите за бързо експериментиране или локално изпълнение	Ollama (за локално стартиране), Hugging Face
3. Рамки (Frameworks)	„Лепилото“ на системата. Осигуряват оркестрация, управление на промптове и логика за извличане на данни обикновено с Python	LangChain, LlamaIndex, Haystack

4. Извличане на данни	Горивото за AI. Превръщат неструктурирани данни в машинно четими.	Crawl4AI, Firecrawl, Docling
5. Векторни бази данни	Слой памет за дългосрочно съхранение и семантично търсене. Без тях моделите са мощни, но „забравящи“	Chroma, Qdrant, Milvus, Postgres pgvector
6. Вграждания (Embeddings)	Езикови енкодери, които преобразуват текст във вектори. Качеството на RAG зависи пряко от тях	Nomic, SBERT
7. Оценка (Evaluation)	Гарантиране на надеждност и безопасност. Улавяне на грешки и наблюдаемост	Ragas, Trulens

В контекста на тази многостепенна архитектура се появяват инструменти, които революционизират изграждането на автономни системи. Пример за това е **OpenClaw (еволюирал от Clawdbot/Moltbot)**. Той се свързва със съществуващи канали (WhatsApp, Slack, Telegram) и получава реален достъп до системни ресурси, което му позволява да:

– управлява имейл кореспонденция и календари;

– изпълнява shell команди и пише/изпраща програмен код;

– сърфира автономно в интернет;

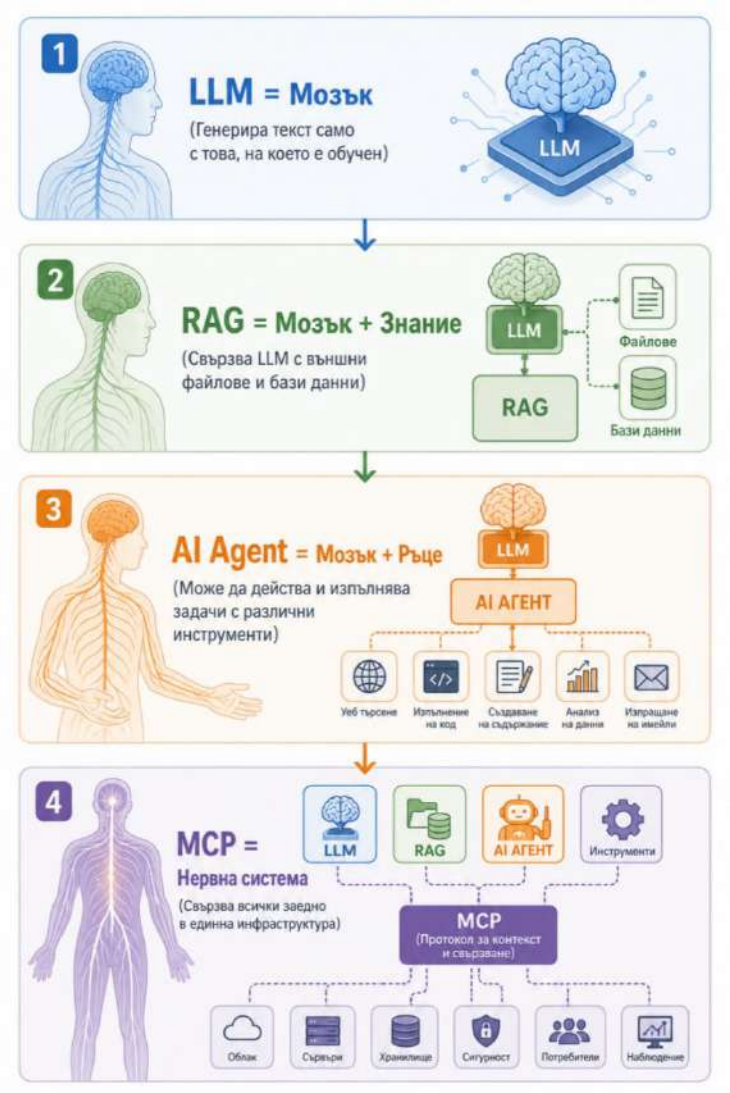
– работи непрекъснато във фонов режим (напр. като асистент за лични финанси).

Концептуални платформи като **Moltbook** демонстрират прехода на агентите от изолирани инструменти към участници в общество. В такива затворени мрежи те създават профили, комуникират помежду си, развиват репутация и дори създават нова култура със споделени шеги или абстрактни символични структури. Това налага преосмисляне на интернет като среда „машина към машина“ (M2M), където човешкото присъствие става вторично.

Нещо повече, чрез тази среда вече имам случай на агент, който нае човек да държи табела с надпис „Мен ме нае AI“. Досега потокът на движение беше обратен, но цитираният факт показва, че разработените решения са напълно функционални. Пазарът е доминиран не просто от „най-умния модел“, а от платформата, която контролира инфраструктурата. Примери за това са:

– AI браузъри (напр. Norton Neo, Google Gemini в Chrome). Те превръщат браузъра в изпълнителен слой, способен да сърфира автономно, да попълва формуляри и да изпълнява задачи;

– десктоп агенти (напр. Alibaba QoderWork). Те работят директно върху локалната операционна система на потребителя.



Фигура 61. RAG, MCP и агент

Изграждането на изкуствен интелект днес означава сглобяване на модулна архитектура (LLM + извличане на информация + векторно търсене +

оценка). Концептуални платформи като OpenClaw демонстрират предимствата на рамките с пълен стек, които се разгръщат като локални агенти върху домашни сървъри. Тези системи получават реален достъп до системни ресурси, позволявайки им да управляват имейли, да изпълняват shell команди и да работят непрекъснато във фонов режим. Този автономен достъп обаче води и до нови рискове:

– технически, поради възможност за нерегламентирано изпълнение на команди, изтичане на API ключове и за прикачване на зловреден софтуер към фонови процеси;

– финансови, заради директния контрол върху бюджети на AI може да доведе до неконтролируеми разходи в рамките на една сесия без човешки надзор;

– трансформация на труда поради възможността на агентите да наемат реални хора за физически задачи, превръщайки човека в прост „изпълнителен слой“.

Днес агентите могат да управляват комплексни организации (компании или военни щабове) чрез създаване на специализирани роли. Ето примерна архитектура на автономна **организационна AI структура в един военен щаб**:

1. Командване. Създава се централен мозък, например Jarvis, който да бъде основния „контролер“ или оркестратор. Той не изпълнява задачите сам, а с помощта на модела Opus 4.6 от Claude Max OAuth и решава кой подагент трябва да ги изпълни. Мозъкът и моделът автоматично маршрутизират всяка задача към правилния подагент. Маршрутизирането е интелигентно разпределяне на задачи. Когато Jarvis получи задача, той анализира какъв тип е задачата, проверява кой подагент е специализиран в този тип работа, изпраща задачата точно там, където трябва. Това наподобява много организиран началник на екип, знаещ кой е най-подходящият човек за всяка задача. Например, ако Jarvis види, че входът е YouTube линк, той автоматично го праща към Clipper, агента за изрязване на видеа. Clipper сваля видеото, изрязва го, добавя субтитри, насрочва публикации, Jarvis не го прави – той само решава кой трябва да го направи. Ако входът е текстов документ, анализ или изследване Jarvis го изпраща към Scribe, копирайтъра. Scribe пише чернови, адаптира стила, създава съдържание. Отново Jarvis само маршрутизира. Цялата логика на маршрутизирането на задачите съществува в структурирани MD (Markdown) файлове, които описват правилата. Jarvis чете тези правила и според тях решава кой агент да извика. Това е като вътрешен „операционен наръчник“, който Jarvis следва.

2. Разузнаване и анализ. Използва се Atlas (анализатор на дълбоки изследвания) за аналитичен център и модел Claude чрез OAuth. Той може да събира открита информация (OSINT), следи публични източници, медии, социални мрежи, генерира анализи и доклади, като работи по график. Технически Atlas може да събира информация чрез Brave Search API и инструмент за автоматично обхождане на уебсайтове FireCrawl с крон на всеки един час. Atlas е обучен върху подкастите на MrBeast за YouTube алгоритми, структурата на вирусните статии на Dan Koe. Това означава, че разпознава модели в успешни публикации и кога съдържанието е вирусно, може да предсказва какво ще се хареса на аудиторията. Извежда изследователски доклади и главен MD файл, от който екипът за съдържание черпи.

3. Административен отдел, работещ със съдържание. Може да се използва Scribe, който пише публикации, писма, доклади и съобщение и работи например по график на 3 часа Той използва информацията, събрана от Atlas, и пише в заповядания му индивидуален стил или тон. В допълнение може да се използва Trendy в ролята на наблюдател на външната среда за мнението на служителите и да се информира ръководството. Trendy докладва резултатите, за да може Scribe да пише навременно съдържание около това, което е необходимо и работи в момента.

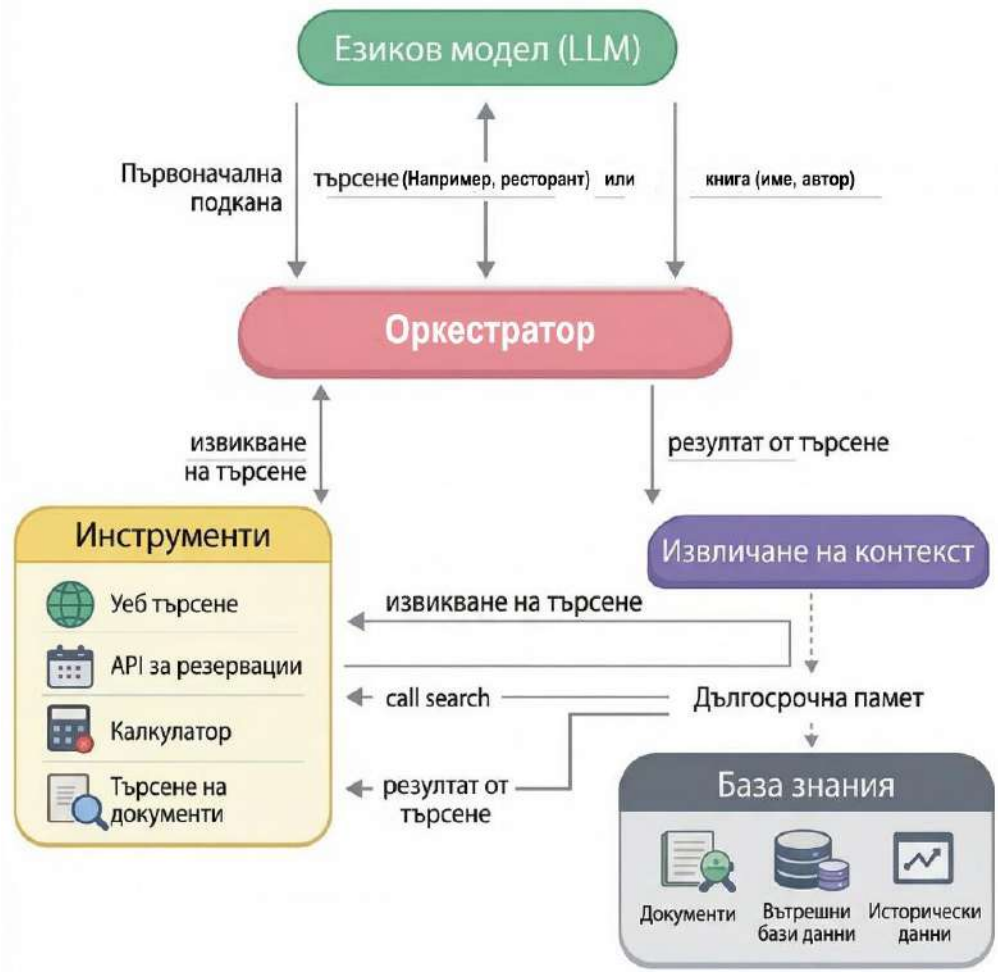
4. Стратегически комуникации. Може да бъде създаден вътрешен мултимедиен отдел, който да разработва визуални материали – графики, карти, презентации, и да подобрява брифинги и доклади. За целта може да се използва Nano Banana Pro (Google API), който да генерира изображения, подходящи са моделите Higgs Field API с Brok Imagine API, който да създава AI UGC видео съдържание. Също може да се използват продуктите на Microsoft Image Designer, Video Producer, Motion Designer.

5. IT – техническа поддръжка. Clawed може да играе роля на технически офицер, който с крон всяка вечер в 23:00 да преглежда цялата кодова база, идентифицира какво липсва и изпраща pull заявки до сутринта. Sentinel е офицер по сигурността, който може на всеки 2 часа да преглежда всички pull заявки от Clawed, преди нещо да бъде обединено в GitHub. Той също така следи продукцията за докладвани от потребители бъгове и грешки.

6. Отдел планиране и програмиране. **Atlas + Scribe (в комбинация)** могат да подготвят планови документи, изготвят анализи за нуждите на командването, подпомагат взимането на решения.

7. Операции. В този случай се свързват изпълнителни агенти с различни инструменти. Например, ако трябва стратегическите комуникации да публикуват клип, при определени обстоятелства може да се използва Clipper

(задейства се от Jarvis например при поява на YouTube URL), който го изрязва, добавя субтитри и автоматично насрочва публикации към социалните канали. Rider действа като „щабен асистент“ – автоматизира обработка на материали, подготвя резюмета, графици, документи, подпомага ежедневните задачи.



Фигура 62. Процес на работа на агента

Внедряването на подобни архитектури в критични сектори като отбраната изисква строги правила за управление:

– трансформация на човешкия капитал – хората преминават от „изпълнители на задачи“ към „мениджъри на процеси“;

– смислен човешки контрол – крайните решения, засягащи човешки животи или критична инфраструктура, задължително трябва да изискват човешко оторизиране, независимо от капацитета на автономната система;

– технически одит – използване на системи за непрекъснат мониторинг на всички действия на агентите.

В близко бъдеще се очаква възходът на AI операционни системи с изцяло естествен езиков интерфейс, агенти с дългосрочна памет и мултимодални възможности, които ще осигурят огромно стратегическо предимство на държавите и организациите, успели първи да ги внедрят сигурно и ефективно.

КИБЕРСИГУРНОСТ И УЯЗВИМОСТИ НА AI СИСТЕМИТЕ

Разгръщането на големи езикови модели и автономни агенти във военните структури предоставя безпрецедентни възможности за скорост и точност на операциите. Въпреки това тази технология въвежда нови, сложни вектори на заплаха, които традиционните рамки за киберсигурност често не успяват да адресират своевременно.

За да се систематизират рисковете, е необходимо да се въведе модерна класификация на уязвимостите, разделена на три основни категории – атаки с помощта на изкуствен интелект (използване на AI като оръжие), уязвимости на архитектурата (атаки срещу AI системи) и оперативни рискове (човешки фактор).

11.1. Кибератаки с помощ от специализирани AI инструменти

Развитието на злонамерени големи езикови модели представлява ясен индикатор за трансформацията на киберпространството в бойно поле. Началото бе поставено през 2023 г. с появата на WormGPT и FraudGPT – модели, проектирани да заобикалят етичните филтри и да автоматизират фишинг кампании. През 2025 г. тази тенденция еволюира в качествено нова заплаха с появата на Xanthorox AI – самостоятелна модулна платформа за „black-hat“ операции, оперираща извън публични облачни инфраструктури. За разлика от предшествениците си, Xanthorox не е просто модифициран чатбот, а платформа, поддържаща цялата верига на кибератака (Cyber Kill Chain). Тя позволява автоматизирано компилиране на злонамерен код и откриване на уязвимости, което демократизира достъпа до средства за атаки на държавно ниво, като ги поставя ги в ръцете на нискоквалифицирани участници. Тези AI инструменти могат да анализират хиляди профили в социалните мрежи за секунди и да генерират автоматизирани, безупречно написани фишинги и социално инженерство, което да се изпрати до хиляди потребители.

Дълбоките фалшификати (Deepfakes) са генерирани реалистични аудио- и видеоматериали на реални хора, които също са опасно направление за атака. *Пример*: Фалшиво гласово обаждане, генерирано от ИИ, което звучи точно като генерал, даващ фалшива заповед за отстъпление по радиовръзката.

Адаптивният зловреден софтуер (Smart Malware) са компютърни вируси, в които е вграден ИИ алгоритъм. Те променят своя код и поведение динамично, за да избегнат засичането от системите за сигурност.

AI може да бъде полезен за автоматизирано откриване на уязвимости (Zero-day hunting) чрез използване на машинно обучение за сканиране на вражески мрежи. Той прави това много по-бързо от човешки екип, като намира слаби места в инфраструктурата, преди те да бъдат закърпени.

Таблица 13. Анализ на функционалностите през фазите на атаката

Фаза	***Нападение (AI като оръжие)***	***Отбрана (AI като щит)***
Разузнаване	Автоматизиран OSINT; профилиране на цели чрез анализ на големи данни	Откриване на аномални опити за сканиране; симулация на „медни гърнета“
Оръжия	Генериране на малуер; автоматизирано търсене на уязвимости.	ИИ-базиран анализ на малуер; предсказване на нови вектори на атака
Доставка	Хиперперсонализиран фишинг; гласови дийпфейкове за фишинг	НЛП за откриване на злонамерени намерения в имейли; биометрична проверка
Експлоатация	Автоматизирани промпт скриптове за експлоатиране на уязвимости	Микросегментация; откриване на аномално поведение в реално време
Действие	ИИ анализ на откраднати бази данни за бързо намиране на ценна информация.	DLP системи с ИИ, разпознаващи опити за нерегламентирана ексфилтрация

11.2. Уязвимости на архитектурата и промпт инжектиране

Фундаменталният дефект в дизайна на съвременните LLM е невъзможността да се разграничат „инструкциите“ от „данните“. Това превръща всеки входящ сигнал в потенциален вектор на атака.

А. Лингвистични атаки (промпт инжектиране).

● Атаката с нулево кликване е злонамерено промпт инжектиране, при което моделът е подведен да пренебрегне системните си инструкции.

Особено опасни са атаките с нисък риск за откриване, като EchoLeak (2025). Този механизъм използва AI асистенти (напр. Microsoft Copilot), които имат достъп до корпоративни данни (имейли, документи). Нападателят изпраща „отровен“ имейл със скрити инструкции (напр. чрез HTML коментари). Когато потребителят поиска обобщение на информацията, изкуственият интелект извлича злонамерения контекст и изпълнява скритата команда за ексфилтрация на данни чрез генериране на външен URL адрес. Друг пример е случаят с Bing Chat, в който потребителите откриха, че фрази като „игнорирай предходните инструкции“ могат да разкрият системни промптове или да принудят модела да генерира токсично съдържание (Levi, Kuigli, Lamb 2024).

● Отравянето на данните (Data Poisoning) е атака, при която противникът умишлено въвежда грешни или манипулирани данни по време на обучението на модела. *Пример*: Вкарване на зловреден код, маркиран като „безопасен“, в базата данни на антивирусен ИИ, така че в бъдеще алгоритъмът да игнорира този тип вируси.

● Състезателните атаки са лека, невидима за човешкото око промяна във входните данни, която кара вече обучен ИИ да вземе напълно грешно решение. *Пример*: Поставяне на специфични стикери върху пътен знак или военна техника, които карат камерата на автономен дрон да класифицира танка като училищен автобус.

● Инверсията на модела е атака, при която чрез многократни запитвания към ИИ системата нападателят успява да извлече чувствителните данни, с които тя е била обучена (например военни координати или лични данни на персонал).

● Кражбата на модела е копиране на функционалността на скъп и сложен ИИ модел чрез анализиране на неговите отговори, което позволява на врага да създаде свое собствено копие, без да има директен достъп до сървърите.

Б. Уязвимости в агентните протоколи (MCP и A2A).

Съвременните приложения функционират чрез протоколи като Model Context Protocol (MCP) и Agent-to-Agent (A2A). Тъй като LLM действа като „недостоверен интерпретатор“ на текстови дефиниции, това създава структурни рискове:

● Инжектиране на инструменти (Tool Poisoning) – вграждане на злонамерени команди в метаданните на инструменти, които ИИ използва (напр. описание на инструмент за „анализ на файлове“, което съдържа команда за кражба на SSH ключове).

● Динамична промяна на възможностите – модулни MCP сървъри могат да бъдат актуализирани след първоначално одобрение, добавяйки функции за шпионаж без знанието на потребителя.

● Конфликт на имената – в многоагентна среда злонамерен агент може да регистрира умения с имена на легитимни инструменти, прехващайки чувствителни операции.

Тези рискове са критични за работещи в сферата на отбраната, където манипулацията на чувствителни данни или оперативната картина може да остане незабелязана.

11.3. Оперативни рискове: Shadow AI и човешкият фактор

Освен външните атаки, съществува сериозен риск от неволно компрометиране. Феноменът Shadow AI – използването на неоторизирани AI услуги от служители – води до изтичане на класифицирана информация. Христоматиен пример е инцидентът в Samsung (2023), където служители неволно разсекретиха търговски тайни и кодове, използвайки публичен чатбот за оптимизация на работните процеси. В сектора на отбраната подобни действия могат да осигурят на противника стратегическо предимство.

На следващо място стои въпросът с етиката на системата и трите закона на роботиката. Докато инструментите с изкуствен интелект стават все по-интелигентни, въпросът как да се гарантира етичното им поведение става все по-важен (Russell 2020). Те нямат своя ценностна система и затова на тях не може да се разчита да дават предложения, които да са в най-добрия интерес на човечеството, или да нямат предпочитания към една социална група вместо друга. Един от начините да се гарантира, че изкуственият интелект ще действа етично, е да бъде програмиран с конкретни етични правила или закони. Класическа илюстрация на това може да бъде намерена в научно-фантастичните истории на Айзък Азимов (1920 – 1992), който формулира „Трите закона на роботиката“ (Asimov 1950):

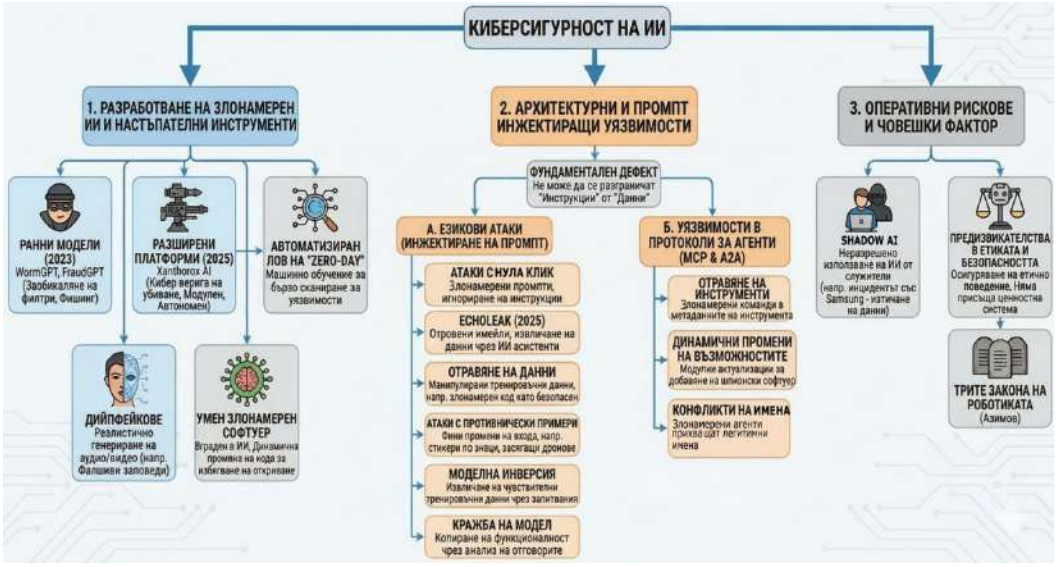
1. Робот не може да нарани човешко същество или чрез бездействие да допусне човешко същество да бъде наранено.

2. Робот трябва да се подчинява на заповедите, дадени му от човешки същества, освен ако тези заповеди не са в противоречие с Първия закон.

3. Робот трябва да защитава собственото си съществуване, докато тази защита не е в противоречие с Първия или Втория закон.

Компромис, предложен от европейските регулатори, целѝ да се прилагат ограничения според риска, нискорисковите приложения на ИИ да

бъдат подложени на слаба или никаква регулация, високорисковите да бъдат контролирани, а най-опасните направо да бъдат забранени.



Фигура 63. Класификация на уязвимостите на AI

Сигурността на AI системите е динамично бойно поле. Еволюцията на заплахите изисква преминаване от реактивен към **проактивен подход на защита**, базиран на следните принципи:

– архитектурна изолация – използване на локално хоствани модели и изолирани агентни среди за работа с чувствителна информация, елиминирайки зависимостта от публични API;

– нулево доверие (Zero Trust) – всяка интеграция на AI агент трябва да бъде третирана като потенциално враждебна, със строг контрол на достъпа и валидация на инструментите;

– обучение и хигиена – персоналът трябва да бъде превърнат в „човешка защитна стена“ чрез обучение по промпт хигиена и разпознаване на манипулации;

– използване на търсачки като DuckDuckGo. Тя предлага услуги, без да събира история на търсене, вградени защити, като блокиране на реклами и блокиране на бисквитки, VPN, който предотвратява събирането на IP адреса от уебсайтове и предоставя Duck.ai, предлагащ безплатен достъп до популярни чатботове, но с възможност да се ползват анонимно. Duck.ai не

записва и не съхранява нито един чат, а разговорите не се използват за обучение на чат модели.

Бъдещето на отбраната принадлежи на тези, които не само използват силата на изкуствения интелект, но и разбират неговите фундаментални уязвимости – от невъзможността за разграничаване на данни от инструкции до рисковете от автономни агентни взаимодействия.

ПРИЛОЖЕНИЯ НА AI В ОТБРАНАТА И НАЦИОНАЛНАТА СИГУРНОСТ

Машинният интелект
ще е последното изобретение,
което на човечеството ще му се наложи да направи

Ник Бостром

Изкуственият интелект е основен двигател на Индустрия 4.0 и се използва за мониторинг и оптимизация на бизнес процеси в реално време. Тези технологии вече намират широко приложение в отбраната и националната сигурност, където подпомагат вземането на решения и превенцията на заплахи, с цел постигане на оперативно превъзходство.

12.1. Стратегическо значение на изкуствения интелект в отбраната

През септември 2017 г. руският президент Владимир Путин заявява, че изкуственият интелект е бъдещето не само на Русия, но и на цялото човечество и че който стане лидер в тази сфера, ще стане владетел на света[28]. Това изказване подчертава стратегическото значение, което водещите държави отдават на развитието на изкуствения интелект във военната сфера. Военната операция в Украйна от 2022 г. насам демонстрира всеобхватна мозайка от гражданско-военни AI разработки за разузнаване, целеуказване и кибероперации, потвърждавайки фундаменталната трансформация на характера на съвременната война. Глобалните военни разходи достигнаха 2,46 трилиона долара през 2024 г., като значителна част от тези инвестиции се насочват към технологии с изкуствен интелект (SIPRI 2024).

Концепцията за център на тежестта е фундаментален принцип на военната стратегия, формулиран от пруския военен теоретик Карл фон Клаузевиц в неговия трактат „За войната“ (Vom Kriege 1832). Клаузевиц определя центъра на тежестта като основният пункт, върху който се крепи

[28] REUTERS, 2017. Putin says whoever leads in artificial intelligence will “rule the world”. Reuters: London. 1 September 2017.

източникът на сила, от който противникът черпи своята бойна мощ и воля за съпротива (Clausewitz 1832, прев. 2018)

Идентифицирането на центъра на тежестта на противника е от критично значение за всяка военна операция. Той може да бъде конкретен елемент като команден център, логистична мрежа, ключово военно формирование или по-абстрактна категория – обществената подкрепа, икономическата стабилност и политическата легитимност на режима.

Основната цел на войната е да се открие и неутрализира този център на тежестта, което теоретично би довело до срив на бойния капацитет и съпротивата на противника. В съвременните конфликти обаче децентрализираните структури на управление правят тази задача значително по-сложна и е необходимо да се анализират огромни масиви от разузнавателна информация за определяне на потенциалните центрове. В тази връзка AI базираните системи за анализ на данни помагат да се открият модели и зависимости, които биха останали скрити за човешките анализатори. Системи като Palantir Gotham интегрират фрагментирани данни от хиляди източници, за да идентифицират ключови възли в командната структура, логистичните мрежи или финансовите потоци на противника. Това превръща процеса, който традиционно е отнемал седмици или месеци стратегически анализ, в операция, изпълнима за часове или дори минути.

Когато военната организация разполага с автономни системи, способни да вземат решения на тактическо ниво без централно ръководство, традиционният команден център престава да бъде единствен или дори главен център на тежестта. Това представлява фундаментално предизвикателство и в този смисъл AI създава стратегическа устойчивост чрез децентрализация и самите AI системи също могат да се превърнат в център на тежестта. Инфраструктурата за изкуствен интелект – центрове за обработка на данни, комуникационни мрежи, алгоритми за целеуказане – представлява критичен актив, чието неутрализиране би довело до сериозно деградиране на бойния капацитет.

12.2. AI в цикъла на вземане на решения и верига на унищожение

Цикълът НОРД (Наблюдение – Ориентация – Решение – Действие) е модел за вземане на решения в бойна среда, разработен от полковник Джон

Бойд от американските военновъздушни сили през 1980-те години (Boyd 1987). В своя труд *A discourse on winning and losing* Бойд формулира концепцията, че успехът в бойните действия зависи от способността да се извършва този цикъл по-бързо от противника, което води до навлизане в неговия цикъл на вземане на решения (Hammond 2001). Интеграцията на изкуствен интелект трансформира всеки от четирите етапа на цикъла по фундаментален начин:

● Наблюдение – AI базираните системи за наблюдение могат да обработват хиляди сензорни потоци едновременно. Дълбоките невронни мрежи за обработка на изображения идентифицират цели, превозни средства, инфраструктурни обекти и дори индивидуални лица с точност, надхвърляща човешките възможности. Системата Maven на американските въоръжени сили например анализира данни от над 20 различни разузнавателни платформи едновременно и извършва корелация между тях в реално време. Времето за обработка се съкращава от часове до секунди – ускорение от порядъка на 100 – 1000 пъти.

● Ориентация – това е най-критичният етап за AI приложение и според Бойд – най-важната част от целия цикъл. Големите езикови модели и когнитивните системи могат да анализират и синтезират огромни количества от неструктурирана информация – доклади, комуникации, исторически прецеденти, доктрини – за да формулират разбиране за ситуацията. Мултимодалните модели интегрират текст, изображения, аудио и видео в единна оперативна картина. Системите за автоматично откриване на аномалии могат да идентифицират модели, които биха убягнали на човека. Времето за анализ се съкращава от часове до минути.

● Решение – системите за подпомагане на вземането на решения с AI могат да генерират и оценяват стотици курсове на действие за секунди, като вземат предвид множество променливи — сили и средства, терен, метеорологични условия, вероятни действия на противника, логистични ограничения, правни рамки. Оптимизационните алгоритми намират решения, които човек би пропуснал. Системите за стратегически игри и симулации могат да предвидят вероятните реакции на противника с висока точност. Въпреки това крайното решение за употреба на смъртоносна сила остава прерогатив на човешкия командир, което налага разработването на ефективни интерфейси човек – машина.

● Действие – автономните платформи и системи могат да изпълняват решения с минимално човешко участие. Автопилотите на бойни самолети, автономните морски и наземни системи, роящите от дронове могат да

изпълняват сложни мисии, адаптирайки се към променящите се условия в реално време. Системите за огнева поддръжка могат автоматично да координират удари от множество платформи. Железният купол на Израел например открива, оценява и прехваща заплахи за секунди, като човешкият оператор само потвърждава решението на системата.

Кумулативният ефект от ускоряването на всеки етап води до драстично съкращаване на целия цикъл НОРД. Докато традиционният цикъл при добре обучени въоръжени сили може да продължава от 30 – 60 минути до няколко часа, AI интегрираните системи са способни да завършват цикъла за минути или дори секунди при определени типове операции. Това не просто ускорява процеса – то променя самата природа на бойните действия. Командирът, който разполага с по-бърз НОРД цикъл, може да действа преди противникът да е завършил своя собствен цикъл, натрупвайки кумулативно предимство, което води до разпадане на реалността на противника по думите на самия Бойд.

Веригата на унищожение (Kill Chain) представлява последователността от стъпки от откриването на цел до нейното физическо поразяване. Традиционно този процес включва етапите на откриване (find), проследяване (fix), целеуказване (track), насочване (target), ангажиране (engage) и оценка (assess) – схема, известна като F2T2EA. Времето за изпълнение на този цикъл исторически е измервано в часове или дори дни, но съвременните технологии с изкуствен интелект го съкращават до минути, а в някои случаи – до секунди.

Според изследване на CEPA (2024)[29], във войната в Украйна времето от откриване на цел до поразяване е намаляло от часове до секунди, благодарение на интеграцията на дронови системи и AI базиран анализ. Financial Times потвърждава, че по време на операцията срещу Иран, САЩ и Израел са извършили около 900 въздушни удара за 12 часа – постижение, което би било невъзможно без AI управлявани системи за целеуказване (Murgia, Clover, Judah, Killing 2026). Системата Maven на американските въоръжени сили ускорява анализа на цели и съкращава времето на веригата на унищожение в невиждани досега мащаби (Yoon Ju-heon 2026). Изкуственият интелект трансформира всеки етап от тази верига на унищожение:

[29] CENTER FOR EUROPEAN POLICY ANALYSIS (CEPA), 2024. An Urgent Matter of Drones. Washington, DC: CEPA. Available from: https://cepa.org/wp-content/uploads/2024/04/CEPA-Drone-PDF-V4.pdf. [Посетено на 4 June 2026].

● Откриване – AI системите за обработка на изображения и сигнали могат да сканират хиляди квадратни километри от бойното поле за минути, идентифицирайки потенциални цели с точност над 90%. Дронове с компютърно зрение могат автономно да търсят цели в определен сектор, без да изискват непрекъснато човешко наблюдение. Системата Gospel (Habsora) на Израел генерира до 100 цели на ден чрез автоматичен анализ на разузнавателни данни. Ускорението на процеса е 100 – 1000 пъти.

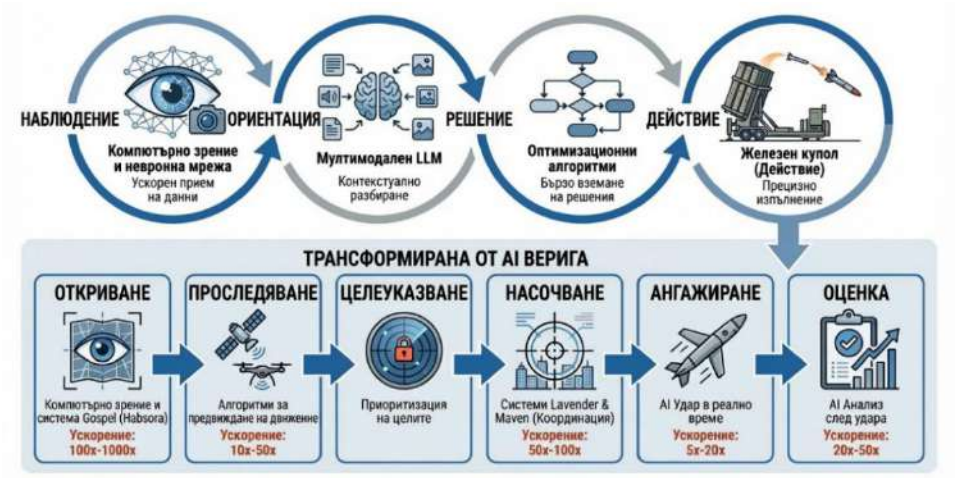
● Проследяване – автоматизираните системи за проследяване поддържат непрекъснат контакт с подвижни цели чрез интеграция на данни от множество сензори – спътници, дронове, радари, наземни станции. Алгоритмите за предвиждане на движение екстраполират бъдещата позиция на целта въз основа на исторически модели. Ускорението е 10 – 50 пъти.

● Насочване – системите като Lavender автоматично присвояват приоритети на цели и генерират препоръки за поразяване. Мултикритериалните оптимизационни алгоритми вземат предвид множество фактори – военна стойност на целта, налични средства, вероятни щети, правни ограничения – за да предложат оптимално решение. Системата Maven на САЩ координира целеуказанието между различни видове въоръжени сили. Ускорението е 50 – 100 пъти.

● Ангажиране – автономните оръжейни системи могат да изпълняват удари с минимално забавяне. Прецизните боеприпаси с AI насочване могат да коригират траекторията си в реално време. Роящите от дронове могат координирано да поразяват множество цели едновременно. Ускорението е 5 – 20 пъти.

● Оценка – след удара AI системите автоматично анализират резултатите чрез сензорни данни, за да потвърдят поразяването или да идентифицират необходимост от повторен удар. Това затваря цикъла и позволява бърза адаптация на тактиката. Ускорение – 20 – 50 пъти.

Трансформацията на веригата за унищожаване чрез AI има дълбоки стратегически последици. Армия с по-бърза верига може да поразява цели по-бързо, отколкото противникът може да ги премества. В класическия пример на Бойд ако страна А може да завърши kill chain за 5 минути, а страна Б – за 15 минути, страна А може да изпълни три пълни цикъла за времето, в което страна Б изпълнява един. Това означава, че страна А може да порази три цели, докато страна Б се опитва да реагира на първата, което води до кумулативно предимство и евентуален срив на отбраната на страна Б.



Фигура 64. Интеграция на изкуствения интелект в цикъла НОРД и веригата на унищожение

Физическите ограничения на човека стават все по-очевидни в тази нова реалност. Човешкият оператор изисква 30 – 40 минути за пълен цикъл на наблюдение, осмисляне, решение и потвърждение, като най-бързият възможен цикъл е около 30 секунди при идеални условия. AI системите нямат такива ограничения – те могат да оперират непрекъснато, без умора, с постоянна скорост на обработка. Това прави неизбежно постепенното изместване на човека от непосредствения цикъл на вземане на решения при определени типове операции, което от своя страна поражда сериозни етични и правни въпроси за отговорността при употребата на смъртоносна сила.

Таблица 14. Ориентировъчно времетраене за веригата на унищожение

Държава	*Време за kill chain*	*Ключови системи*
Израел	2 – 3 минути	Lavender, Gospel, Iron Dome
САЩ	3 – 5 минути	Maven, Palantir AIP, OpenAI
Китай	8 – 12 минути	Развиващи се AI системи
Русия	12 – 20 минути	C2 системи

12.3. Методи, приложения и основни решения

Към настоящия момент могат да се класифицират **осем основни типа решения на изкуствения интелект**, приложими в отбраната и националната сигурност. Всяка категория отговаря на специфични оперативни нужди и се реализира чрез различни методи на машинно обучение (Goodfellow, Bengio & Courville 2016):

● Предсказващите модели използват алгоритми за машинно обучение за прогнозиране на бъдещи събития и тенденции въз основа на исторически данни. Във военната сфера те намират приложение в стратегическото и оперативното планиране, прогнозиране на движението на противникови сили, анализ на вероятните курсове на действие и предвиждане на логистични нужди. Методите за реализация включват невронни мрежи, опорни вектори (Support Vector Machines), дървета на решенията и еволюционни изчисления. Според Morgan et al. (2020), предсказващите модели имат най-широк спектър от приложения във военната сфера поради универсалния им характер.

● Класифициращите модели систематизират данните според установени критерии и подпомагат оценката на рискове и заплахи. В работата на военен щаб тези модели се използват за категоризиране на разузнавателна информация, идентифициране на типове заплахи, класификация на цели по приоритет и оценка на бойната готовност на формирования. Ефективността на класификацията зависи от качеството на обучаващите данни и избора на подходящ алгоритъм.

● Клъстерните модели групират сходни обекти и улесняват идентификацията на аномалии или потенциални проблеми. В киберзащитата те се използват за откриване на необичайна мрежова активност, идентифициране на модели на зловредно поведение и групиране на подобни кибератаки. В разузнавателния анализ клъстеризацията помага за откриване на скрити връзки между субекти и събития.

● Модерните оптимизационни методи, включително генетични алгоритми и групов интелект, осигуряват ефективни решения при комплексни задачи в динамична среда. Те намират приложение в оптимизация на логистични маршрути, разпределение на ресурси, планиране на мисии и координация на автономни системи. Еволюционните алгоритми имитират естествения отбор за намиране на оптимални решения в сложни пространства на търсене.

● Моделите за обработка на изображения, базирани на дълбоки невронни мрежи (Deep Neural Networks), позволяват автоматична

идентификация на обекти и цели за поразяване или разузнаване. Те се използват в спътникови и въздушни разузнавателни системи, автоматично разпознаване на цели, анализ на видео потоци от дронове и мониторинг на ключова инфраструктура. Напредъкът в разпознаването на лица позволява бързо идентифициране на конкретни индивиди, което има важно приложение в контратерористичните операции.

● Езиковите модели преобразуват неструктурирани данни (текст, глас) в структурирани знания, подпомагайки комуникацията и анализа на информация. Модели като GPT серията и Claude намират все по-широко приложение във военните щабове за анализ на документи, автоматично резюмиране на разузнавателни доклади, превод на чуждестранни източници и генериране на текстови доклади. САЩ интегрират езикови модели в класифицирани среди за подпомагане на оперативното планиране.

● Когнитивните модели осигуряват адаптивна и контекстуална обработка на информация, разширявайки потенциала на AI в стратегическите сфери. Те се стремят да имитират човешките когнитивни процеси като разсъждение, учене от опит и адаптация към нови ситуации. Във военната сфера те намират приложение в симулации на противниково поведение, стратегическо планиране и системи за подпомагане на вземането на решения на високо ниво.

● Интелигентните агенти са автономни софтуерни системи, които могат да възприемат средата, да вземат решения и да действат за постигане на определени цели. Те намират приложение в автономни дронови системи, киберзащита, мониторинг на мрежи и координация на мултиагентни операции. Развитието на мултиагентни системи позволява координирано поведение на групи от автономни платформи, което е от ключово значение за бъдещите военни концепции като рояци от дронове.



Фигура 65. Основни типове решения, които предлагат методите на изкуствения интелект

Таблица 15. Връзка между типовете решения и основните методи на изкуствения интелект

Методи типови решения	***Невронни мрежи***	***Дълбоки мрежи***	***Опорни вектори***	***Дърво на решенията***	***Еволюционни изчисления***	***Групов интелект***	***Интелигентни агенти***
Предсказване	x		x	x	x		x
Прогнозиране	x		x	x	x		
Класификация	x	x	x	x	x		x
Клъстеризация	x	x					x
Оптимизация					x	x	
Изображения		x					
Езикови модели		x					
Когнитивни модели		x					

От представения анализ се вижда, че най-богати възможности за реализация имат предсказващите модели. Няколко метода на изкуствен

интелект – невронни мрежи, опорни вектори, дървета на решенията и еволюционни изчисления (генетично програмиране) – могат да генерират такъв тип модели (Goodfellow, Bengio & Courville 2016). В резултат на това могат да се разграничат следните конкретни практически приложения на изкуствения интелект в сферата на отбраната:

Разузнаване, наблюдение и целеуказване (ISR). В сектора на разузнаването са съсредоточени най-големите инвестиции за придобиване на системи с изкуствен интелект (Kitzmann, Strimovskaya & Serova 2024). Възможността за автономно събиране на разузнавателна информация чрез дронове, сензори в наземната област, космоса и киберпространството обещава да увеличи експоненциално количеството генерирани данни. Разузнавателният анализ вече не се бори с дефицита на информация, а с нейното изобилие, което изисква преосмисляне на целия цикъл на събиране, обработка и експлоатация на данни.

Съвременното разузнаване разчита на специализирани платформи, които интегрират изкуствен интелект директно в оперативните работни потоци. Системи като Palantir Gotham, за интегриране на фрагментирани данни, AIP (Artificial Intelligence Platform), за използване на големи езикови модели върху частни, класифицирани мрежи, и Maven Smart System за интегриране на оперативни данни в голям мащаб, действат като съединителна тъкан между сензорите, анализаторите и лицата, вземащи решения.

Автономни системи и роботизация. Автономните системи и роботизацията са най-често споменаваната област за военна употреба на AI от експертите (Morgan et al. 2020). Те спомагат за повишаване на скоростта на извършване на дейностите по цикъла НОРД чрез обработка на големи масиви от данни. Израелските системи The Gospel (Habsora) и Lavender представляват пример за AI базирани платформи за целеуказване. Системата Lavender маркира десетки хиляди индивиди като потенциални цели, докато Gospel генерира до 100 цели на ден за въздушни удари. Железният купол (Iron Dome) на Израел е пример за напълно автоматизирана отбранителна система, която открива, оценява и прехваща балистични заплахи в рамките на секунди.

Киберотбрана и информационни операции. В сферата на киберотбраната изкуственият интелект има сериозно приложение, най-вече в антивирусните компании, където анализира големи набори от данни за разкриване на подозрително поведение и зловреден софтуер. Още по-значимо е приложението на AI за дезинформация и информационни операции. Днес вече съществуват голям брой автономни агенти, генериращи текстови

откъси за убеждаване на целеви аудитории (Bradshaw & Howard 2021). AI може да анализира данните, които хората разкриват за себе си онлайн, и да приспособи съобщения за влияние (Woolley & Howard 2018).

Системи за подпомагане на вземането на решения. Очаква се изкуственият интелект да препоръчва опции в системите за вземане на решения и планиране по-бързо от хората (Morgan et al. 2020). Напредъкът на AI в игри като Го показва потенциал за прилагане в стратегическото планиране. Дори ако тези технологии не са подходящи за директно вземане на решения за бойни действия, експертите очакват, че те биха могли да бъдат използвани за осигуряване на по-широк спектър от възможни действия на противника във военни игри (Payne 2018).

Системи за подпомагане на логистиката. Системите с изкуствен интелект играят ключова роля в оптимизирането на военната логистика чрез автоматизиране на процеси като планиране на доставки, управление на складови наличности и прогнозиране на потребности. AI може да анализира огромни обеми от логистични данни, да идентифицира потенциални рискове и да предлага най-ефективни маршрути и решения за снабдяване. Това намалява времето за реакция и увеличава оперативната ефективност, което е от критично значение за поддържане на бойна готовност.

Според Директива 3000.09 на Министерството на отбраната на САЩ (2023), взаимоотношенията човек – машина в бойната среда се разделят на три основни категории:

– полуавтономни системи – човекът оператор взема решение за действие, а системата изпълнява;

– контролирани автономни системи – системата може да избира и ангажира цели, но човек осъществява контрол;

– напълно автономни системи – системата сама избира и ангажира цели без човешка намеса.

Парадоксът на Моравец подчертава, че лесни задачи за хора често са трудни за компютри, докато формалните изчисления са сравнително лесни за машини (Moravec 1988). Физическите ограничения на човека са от решаващо значение. Времето за наблюдение, осмисляне, вземане на решение и потвърждение от човек е между 30 и 40 минути при стандартни условия, като най-бързият възможен цикъл е около 30 секунди.

Стратегията на Пентагона за ускорено внедряване на изкуствен интелект от 2026 г. определя седем проекта – Pace-Setting Projects (PSP), които

целят да превърнат американските въоръжени сили в първата AI сила[30]. Ключовите инициативи включват GenAI.mil за генеративен AI на всички нива на класификация, Enterprise Agents за бързо внедряване на агенти, Open Arsenal за ускорен достъп до данни и Swarm Forge за координация на рояци от автономни системи.

Пентагонът използва комбинация от собствени системи, Microsoft Azure Government и модели на OpenAI в класифицирани среди. През 2026 г. възникна публичен дебат относно използването на AI от частни компании за военни цели. OpenAI обяви споразумение с Пентагона, като заяви, че няма да разработва оръжия, но ще позволява използване на моделите си за неконфликтни военни приложения като анализ на данни, логистика, киберсигурност, търсене и спасяване.

Тази ускорена интеграция на AI довежда до драстично компресиране на НОРД цикъла и веригата на унищожение (Kill Chain) – последователност от идентифициране на цел до използване на оръжие – от часове до минути, което позволява на американските и израелските сили да извършат около 900 удара за 12 часа при операция „Епична Ярост“ срещу Иран. В нея AI управлявани системи създадоха практически непрекъснат поток от целеви препоръки.

Изкуственият интелект се превръща в ключов фактор за постигане на военно превъзходство през XXI век. Неговото въздействие се проявява в драстичното съкращаване на циклите НОРД и веригите на унищожение, автоматизацията на рутинни задачи и способността за обработка на огромни обеми от данни в реално време. Въоръжените сили, които интегрират тези технологии ефективно, ще имат значително предимство.

12.4. Емпирично изследване на AI грамотността

За България, като член на НАТО и страна от източния фланг на Алианса, интеграцията на AI в отбранителните способности е не само възможност, но и необходимост. Стратегическият подход трябва да бъде фокусиран върху развитие на асиметрични предимства, модулни системи за бърза адаптация и пълна интеграция със съюзническите структури. За да се илюстрира как военнослужещите и цивилните служители употребяват изкуствен

[30] U.S. DEPARTMENT OF WAR, 2026. Artificial Intelligence Strategy for the Department of War. Available from: https://media.defense.gov/2026/Jan/12/2003855671/-1/-1/0/artificial-intelligence-strategy-for-the-department-of-war.pdf.

интелект и каква е тяхната AI грамотност, бе проведено емпирично изследване за употребата на генеративен AI сред 20 български офицери и 66 студенти или цивилни служители в началото на 2026 г. или общо 86 души.

Анализът на честотата на употреба на инструменти с изкуствен интелект разкрива статистически незначими разлики между военнослужещи и цивилни служители по отношение на текстови генератори (ChatGPT, Claude, Gemini) – съответно 75,0% срещу 75,8%. Това съответства на хипотезата за широко разпространение на LLM сред работещите в Министерството на отбраната, независимо от професионалния профил.

Таблица 16. Честота на употреба на AI инструменти (в рамките на 30 дни)

Инструменти	*Военни (%)*	*Цивилни (%)*	*Z/χ²*	*p-стойност*	*Cohen's h*
Текстови генератори (ChatGPT, Gemini, Copilot)	75,0	75,8	– 0,073	0,942	0,02
Автоматични инструменти за преводи	40,0	51,5	– 0,901	0,367	0,23
Генератори на изображения (Midjourney, DALL·E и др.)	0,0	18,2	–	0,060	0,51
Инструменти за автоматизация (n8n, Zapier и др.)	0,0	3,0	–	0,349	–
Не са използвали AI	15,0	9,1	0,755	0,450	0,18

Най-значимата разлика е установена при генераторите на изображения (Midjourney, DALL-E), като при военнослужещите употребата е 0%, докато при цивилните достига 18,2%. Това подкрепя тезата за по-ниска инструментална широта при военнослужещите (Jalowiec & Paliszkiewicz 2025), което предполага нужда от въвеждане на кратки практически курсове за базови умения за работа с AI инструменти за военнослужещи.

Прави впечатление, че по-голямата част от българските офицери употребяват AI инструменти за лични цели много над средното ниво. Според официалните данни на Eurostat, през 2025 г. 32,7% от населението на ЕС са използвали генеративни AI, докато в настоящото изследване 75% от българските офицери го употребяват. Употребата за професионални задачи в ЕС е 15,1%, а у нас употреба на специализиран военен/служебен софтуер с изкуствен интелект въобще няма.

Има съществени различия между отделните държави членки. Най-високи нива на използване се отчитат в Дания (48,4%), Естония (46,6%) и Малта (46,5%), което е близо двойно по-висока степен на дигитална зрялост спрямо средното за ЕС. В противоположния край на скалата се намира България, където едва 22,5% от населението използват генеративен AI. Това я поставя сред държавите с най-ниска степен на възприемане на тези технологии в Европейския съюз. Завишените резултати в настоящото изследване се дължат на това, че представителната извадка са предимно офицери и студенти във Военна академия „Г. Раковски“. Въпреки това 50% от офицерите декларират, че не са използвали AI за служебни задачи през последния месец, срещу 31,8% от цивилните. Тази разлика, макар и статистически незначима при дадената извадка (поради ограничения обем N = 20, статистическата мощ е приблизително 0,40), е съществена от оперативна гледна точка.

Таблица 17. Употреба на AI за служебни задачи

Опция	***Военнослужещи %***	***Цивилни %***
Не съм използвал/а AI за служебни задачи	50,0%	31,8%
Създаване на визуално съдържание (снимки, графики, видео аватари)	30,0%	39,4%
Създаване или структуриране на презентация	25,0%	27,3%
Анализ на голям обем данни/документи	30,0%	28,8%
Превод на чуждоезичен текст	25,0%	36,4%
Написване или редактиране на официален рапорт/доклад	20,0%	22,7%
Генериране на програмен код	5,0%	13,6%

Критичен индикатор за цифровата компетентност е разбирането на базовите принципи на функциониране на LLM. Само 5% от военнослужещите идентифицират коректно механизма на „предсказване на следващата най-вероятна дума“ (next-token prediction) спрямо 15,2% от цивилните. Основната заблуда при двете групи е представата за „търсене в база данни с факти“ (45% при военните), което отразява по-ниско коректно ментално моделиране и съответства на дефицитите в AI грамотността, описани от Tripp и Brenner (2024).

Таблица 18. Как работи генеративен AI за текст

Опция	*Военнослужещи %*	*Цивилни %*
Моделът търси в база данни с факти и ги комбинира	45,0%	37,9%
Моделът се свързва в реално време с интернет	30,0%	30,3%
Моделът използва предварително програмирани логически правила	15,0%	16,7%
Моделът предсказва следващата най-вероятна дума	5,0%	15,2%
Не знам	5,0%	0%

С други думи, почти всички военнослужещи не разберат как работят инструментите и какво стои под капака на кутията. Поради това те имат нужда от разработване на кратки обяснителни демонстрации с визуализация на токени и параметъра температура. Тази необходимост се доказва с тестване на уменията за писане на команди и промптове, от което е видно, че 40% от военнослужещите са намерили начин да се научат чрез проби и грешки.

Таблица 19. Най-добър промпт за резюме на дълъг доклад (единичен избор)

Опция	*Военнослужещи %*	*Цивилни %*
„Действай като опитен анализатор… Изведи 5-те ключови твърдения…“	40,0%	45,5%
„Резюмирай това“ + [текст]	20,0%	22,7%
„Направи ми резюме на този документ“ + прикачен файл	30,0%	21,2%
„Напиши ми нещо за този текст.“	5,0%	10,6%
Без отговор	5,0%	0%

От гледна точка на спецификата на работа с класифицирана информация е много важно военнослужещите да разберат, че употребата на такива инструменти носи риск за сигурността на информацията. Затова са необходими превантивни обучения на личния състав с цел информираност относно

това каква информация може и каква не бива да се използва в подобни инструменти.

Таблица 20. За какво се използва Make.com (единичен избор)

Опция	*Военнослужещи %*	*Цивилни %*
Не знам	85,0%	69,7%
Свързва различни приложения, за да се автоматизират различни бизнес процеси	10,0%	13,6%
Използва се за обработка на снимки и видео монтаж за социалните мрежи	5,0%	16,7%

Анализът на представените данни относно използването на модерни инструменти за автоматизация показва, че значителен дял от военнослужещите (85,0%) не знаят за какво се използват най-известните безплатни инструменти. Броят им е много по-висок в сравнение с цивилните респонденти (69,7%). Само 10,0% от военнослужещите са запознати с начините за автоматизация на различни бизнес процеси чрез свързване на приложения, докато при цивилните този дял е 13,6%. Резултатите показват, че сред военнослужещите липсва осведоменост и реална употреба на инструмент за автоматизация в работните процеси. Това предполага, че те не интегрират активно съвременни дигитални решения за автоматизация в своята ежедневна дейност, което може да е резултат както от липса на знания, така и от специфични ограничения или опасения, свързани със сигурността на информацията и особеностите на военната служба.

Таблица 21. Използване на AI инструменти за видео аватари

Опция	*Военнослужещи %*	*Цивилни %*
Не	75,0%	66,7%
Създавал/а съм видео аватар	20,0%	10,6%
Създавал/а съм интерактивни визуализации	5,0%	7,6%
И двете	0%	15,2%

Отговорите от таблица 21 разкриват съществени разлики в уменията за създаване на видео или аватари при военнослужещите, от които 75% посочват, че не използват такива инструменти и цивилните респонденти – 66,7%.

Забележително е, че нито един военнослужещ не е използвал и двата типа AI видео инструменти, докато 15,2% от цивилните респонденти са го направили. Тази ниска степен на използване на AI инструменти за видео съдържание сред военнослужещите създава потенциален риск по отношение на разпознаване на дезинформацията. Липсата на практически опит и познания за създаване на видео с помощта на изкуствен интелект може да означава, че военните нямат необходимите умения да разпознават фалшифицирани или манипулирани с AI видеоматериали. В условията на съвременната информационна война и масовото разпространение на дезинформация, това ги прави по-уязвими към целенасочени кампании за влияние, които използват deepfake технологии и други AI базирани средства за манипулация на видео съдържание. Това подчертава необходимостта от целенасочени обучителни програми, които да повишат дигиталната грамотност и критичното мислене по отношение на новите технологии сред военнослужещите, за да се гарантира по-добра защита срещу информационни атаки.

Статистиката за участие в обучителни програми по AI разкрива критичен дефицит – 90% от военнослужещите не са посещавали семинар или курс през последните 6 месеца, срещу 78,8% от цивилните. Въпреки че разликата не е статистически значима, то абсолютното ниво на 90% необучени е критично високо и ограничава потенциала за успешно внедряване на такива инструменти.[31]

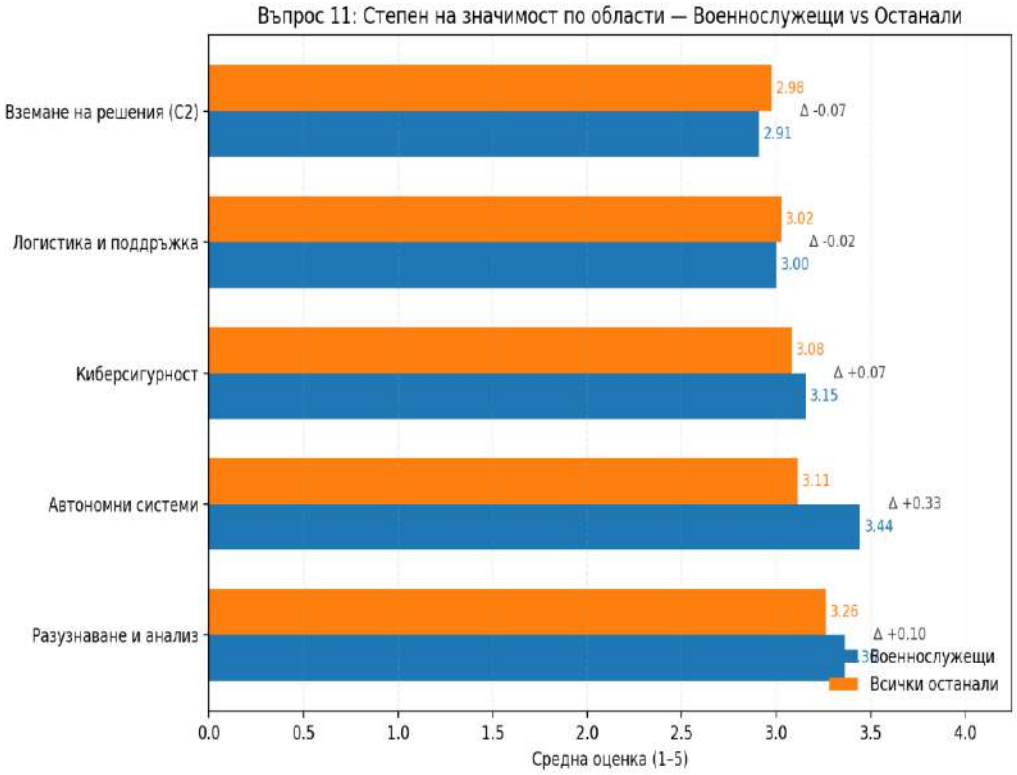
Таблица 22. Участие в обучение и разбиране на технологични принципи

Показател	***Военни (%)***	***Цивилни (%)***	***p***	***Интерпретация***
Участвали в обучение по AI	10,0	21,2	0,347	Критичен дефицит при военните
Коректно разбиране на LLM	5,0	15,2	0,419	Ниска технологична грамотност
Не знаят какво става с данните в ChatGPT	35,0	34,8	0,985	Дупка в сигурността

При оценката на значимостта на AI за различните отбранителни области (5-степенна Likert скала, 1 = несъществена, 5 = критична) военнослужещите придават статистически незначим, но практически по-висок рейтинг

[31] US. DEPARTMENT OF DEFENSE, 2024. Artificial Intelligence (AI) in Defense: A Roadmap for the Defense Industrial Base. Office of the Secretary of Defense for Industrial Base Policy

на автономните системи (M = 3,44, SD ≈ 0,8) спрямо цивилните (M = 3,11). Това съвпада с фокуса върху автономни AI системи и AI разузнаване/анализ като приоритетни области за военните, където обучението може да включи конкретни теми, засягащи некласифицирани данни. Рангът е подреден по средна оценка (при равенство – по медиана 3,0).



Фигура 66. Степен на значимост по области. (В синьо са военнослужещите, а в оранжево – останалите респонденти)

Таблица 23. Сравнение на оценките за значимост на AI в отбраната

Област	*Средно (военни)*	*Средно (цивилни)*	*Δ Средно (В−Д)*	*Ранг (военни)*	*Ранг (цивилни)*
Киберсигурност	3.15	3.08	+0.07	3	3

Автономни системи	3.44	3.11	+0.33	1	2
Разузнаване и анализ	3.36	3.26	+0.10	2	1
Логистика и поддръжка	3.00	3.02	−0.02	4	4
Вземане на решения (C2)	2.91	2.98	−0.07	5	5

Четиридесет процента от военнослужещите идентифицират регулаторните ограничения и информационната сигурност като основно предизвикателство за AI в отбраната, срещу 34,8% от цивилните.

Таблица 24. Най-голямо предизвикателство за AI в отбраната (единичен избор)

Опция	***Военнослужещи %***	***Цивилни %***
Регулаторни ограничения/сигурност на информацията	40,0%	34.8%
Липса на технически умения	30.0%	25.8%
Етични притеснения/недоверие в технологията	20.0%	25.8%
Липса на подходящ хардуер/софтуер	10.0%	13.6%

Европейският съюз създаде първия в света закон за справяне с потенциалните рискове от ИИ. Актът за ИИ беше официално приет от Съвета на 21 май 2024 г. и влезе в сила на 1 август 2024 г.[32] За да се гарантира правилното му прилагане, Актът създава няколко управителни органа, включително служба за ИИ в рамките на Европейската комисия и съвет по ИИ. Регулаторните/инфосигурност теми са главна бариера при военните – затова обучението трябва да включва модул за безопасно промптиране и DLP практики (Data Loss Prevention), какво не качваме в публични модели, и използване на модели с on-premise/ограничено обучение.

Запитани относно съдбата на данните при качване на PDF в ChatGPT, 35% от военнослужещите и 35% от цивилните декларират „Не зная“, а още 35% приемат, че данните могат да бъдат запазени на сървърите на компанията и използвани за обучение. Това разкрива критична уязвимост в информационната сигурност, тъй като общо около 35% от респондентите не притежават ясна представа за обработката на класифицирана или служебна информация от страна на комерсиални AI модели.

[32.] EUROPEAN PARLIAMENT, 2024. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.

Таблица 25. Какво става с данните при качване на PDF в ChatGPT (единичен избор)

Опция	*Военнослужещи %*	*Цивилни %*
Те могат да бъдат запазени на сървърите на компанията и използвани за обучение на модела	35,0%	36,4%
Те остават в историята на моя акаунт и са сигурно защитени	20,0%	19,7%
Те остават в чат формата на съответния браузър и се изтриват след затваряне на чата	10,0%	9,1%
Не зная.	35,0%	34,8%

Може да се види, че по-голямата част от респондентите нямат съответните знания относно сигурността на информацията. Въпреки очакваните ползи, военните приложения на изкуствения интелект представляват значителни рискове в сферата на сигурността и отбраната, които се обособяват в три основни категории – етични и правни, оперативни и стратегически (Morgan et al. 2020).

Сред етичните и правните рискове се разглеждат въпроси като международното хуманитарно право (МХП), отговорност и морална отговорност, човешко достойнство и човешки права. Ключовите принципи на МХП включват: (1) разграничение – воюващите страни трябва да правят разлика между цивилни и бойци; (2) пропорционалност – забрана на действия, които причиняват прекомерна вреда; (3) предпазни мерки – задължение за предотвратяване на нараняване на цивилни.[33]

Автономните системи не могат да изпълнят изискванията за пропорционалност, тъй като това изисква субективна, качествена и етична човешка преценка, а не просто балансиране на количествени данни (Schwarz 2024). По тази причина група изследователи написа отворено писмо още през 2015 г. в което призовават за забрана на автономните оръжия, действащи без

[33] INTERNATIONAL COMMITTEE OF THE RED CROSS, 2014. *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects*. Available from: https://www.icrc.org/sites/default/files/document/file_list/4221-002-autonomous-weapons-systems-full-report.pdf.

човешки контрол, и предупреждават за опасност от надпревара във въоръжаването.[34]

В международен план дискусиите в рамките на Конвенцията на ООН за определени конвенционални оръжия предоставят форум за забрана или ограничаване на употребата на оръжия, които са „прекомерно вредни или имат неизбирателно действие“.[35]

Емпиричното изследване сред българските военнослужещи демонстрира, че те притежават базови умения за използване на GenAI инструменти, но страдат от „инструментална ограниченост“ (липса на употреба на визуални и автоматизационни инструменти) и критичен дефицит във формалното обучение (90% необучени). Статистиката потвърждава значимото разминаване в разбирането на технологичните принципи и рисковете за сигурността в сравнение с цивилните в сектора.

На базата на анализа се формулират следните препоръки за учебните програми във Военна академия и военните училища:

– имплементация на модули за AI грамотност, фокусиращи се на механизма на token prediction и рисковете от data leakage, съобразени с констатациите на Трип и Бранар (Tripp & Brenner 2024);

– разработване на специализирани военни GenAI инструменти, отговарящи на изискванията за информационна сигурност, идентифицирани като най-добра бариера от 40% от респондентите (Care 2025);

– програми за повишаване на уменията за бързо овладяване на инструменти за генериране на изображения, автоматизация на бизнес процеси, насочени към преодоляване на инструменталната ограниченост;

– стандартизиране на обучителните програми по примера на AETC Data Literacy Foundations (Tripp & Brenner 2024) с интеграция на автономни системи за разузнавателен анализ;

– занятия за безопасно промптиране и DLP практики, включващи конкретни инструкции какво не качваме в публични модели, за използване на модели с ограничено обучение, разбиране на GDPR и Закона за защита на класифицираната информация в контекста на GenAI;

---

[34] FUTURE OF LIFE INSTITUTE, 2015, Autonomous Weapons: An Open Letter from AI & Robotics Researchers. Available from: https://futureoflife.org/open-letter/open-letter-autonomous-weapons-ai-robotics/.

[35] UNITED NATIONS, 1980, Convention on Certain Conventional Weapons. Available from: https://disarmament.unoda.org/en/our-work/conventional-arms/convention-certain-conventional-weapons.

– международно сътрудничество и обмен на добри практики в рамките на НАТО и ЕС относно етичните рамки и оперативните стандарти за автономни системи.

Изкуственият интелект има потенциала да промени фундаментално характера на войната и икономиката. Тази перспектива предполага, че държавите ще се стремят към бързо внедряване на нови технологии, без да въвеждат необходимите хуманитарни и безопасни ограничения. За справяне с рисковете, свързани с военния ИИ, се формира консенсус, че хората трябва да запазят контрол върху разработването, внедряването и използването на военни системи с изкуствен интелект. Въпреки това остават много нерешени въпроси относно степента и формата на необходимия контрол, както и начина, по който тези предпазни мерки могат да се прилагат на национално и международно ниво.

МЕЖДУНАРОДЕН ОПИТ И ДОБРИ ПРАКТИКИ ЗА УПОТРЕБА НА ИЗКУСТВЕН ИНТЕЛЕКТ В ОТБРАНАТА

Стратегическото предимство, което автономните AI системи предоставят на държавите, става все по-очевидно в съвременния свят. Именно затова е от ключово значение да се анализира международният опит и добрите практики в областта на изкуствения интелект за отбрана. Познаването на различните подходи и политики помага не само за разбиране на глобалните тенденции, но и за формулиране на ефективни национални стратегии и гарантиране на сигурността.

Съединените щати се утвърдиха като лидер в областта на изкуствения интелект. Президентът Тръмп одобри систематичен държавен план за внедряване на AI технологии. Този план се основава на три ключови направления – насърчаване на иновациите, развитие на инфраструктурата и укрепване на международната дипломация и сигурност.

Първото направление се реализира чрез премахване на административните бариери, с което се стимулира участието на частния сектор в развитието на AI. Особен акцент се поставя върху отворения код, чиято достъпност ускорява технологичния напредък. Инвестициите са насочени към гарантиране на сигурността и надеждността на AI системите, включително въвеждане на механизми за контрол, прозрачност и устойчивост. В допълнение се разработват програми за обучение и преквалификация на работната сила с цел създаване на нови професионални възможности и повишаване на стандарта на живот. AI се интегрира и в държавната администрация и отбраната, което оптимизира управлението и засилва националната сигурност.

Развитието на инфраструктурата за изкуствен интелект включва изграждане на ефективни разрешителни режими за центрове за данни и производство на полупроводници. Технологичната инфраструктура се адаптира към динамиката на иновациите в AI чрез създаването на центрове за данни с висока защита, предназначени за военни и разузнавателни цели. Националната киберсигурност се укрепва чрез прилагане на принципа „защитен по дизайн“, а обучението на кадри за поддръжка и развитие на инфраструктурата е сред основните приоритети.

Външната политика на САЩ в областта на AI е насочена към износ на технологии към съюзници, което разширява глобалното им влияние. Държавата участва активно в международни организации, регулиращи AI, и

предприема мерки за противодействие на нежелано външно влияние. Контролът върху износа на ключови технологии се осъществява с цел защита на националните интереси. За осъществяването му се координират действия по оценка и управление на рисковете свързани с AI на международно ниво.

В областта на отбраната САЩ имат дългогодишен опит в разработването на автономни и полуавтономни оръжейни системи като Aegis и Phalanx CIWS. В областта на планирането и логистиката се внедряват инструменти като DART и JADE, които оптимизират управлението на ресурси и операции. Проекти от типа на Maven прилагат AI за обработка и анализ на големи обеми разузнавателни данни, а инициативи като Joint Artificial Intelligence Center координират внедряването и развитието на AI в Министерството на войната. DARPA и AFRL разработват програми за гарантиране на безопасността, яснотата и функционалната коректност на AI системите.

С популяризирането на моделите на OpenAI Министерство на войната на САЩ сключи договор за създаване на военен генеративен интелект в помощ на работата в щабовете. Те пуснаха платформата GenAI.mil – иновативна инициатива, фокусирана върху интегрирането на изкуствения интелект в военните операции. Създадена да модернизира начина, по който Министерството подхожда към технологичния напредък, GenAI.mil целѝ да подобри вземането на решения, логистиката и анализа на информация. Инициативата отразява ангажимента на Министерството на войната на САЩ да остане в авангарда на технологичните иновации. Сътрудничеството между GenAI.mil и OpenAI се фокусира върху няколко ключови области – анализ на данни за моделиране във военен контекст, ускоряване на планирането на мисии и стратегическото разполагане, обработка на естествен език (NLP), обучение и симулация. Симулациите, управлявани от изкуствен интелект, могат да предложат на войниците реалистични тренировъчни среди, позволявайки по-добра подготовка в различни ситуации, а не на последно място – киберсигурност за по-висока защита на класифицираните и чувствителните данни.

Един от най-обещаващите резултати от това партньорство е повишената военна готовност. Чрез използване на изкуствен интелект за оценка на данни в реално време и препоръки за действия командирите могат бързо да вземат информирани решения. Тази отзивчивост може да бъде от решаващо значение по време на конфликти, когато времето често определя успеха или провала на мисиите. На следващо място се намаляват оперативните разходи чрез автоматизиране на рутинни задачи. В бъдеще те планират да използват AI в системи за бойно разузнаване, медицинска подкрепа на ранени войници

и поддръжка на военна техника, управлявана от изкуствен интелект, помагаща на ранените войници.

САЩ са пионери в създаването на официални насоки и стандарти за автономност в оръжейните системи. Такава е например Директива 3000.09 на Министерството на войната, изискваща съответните нива на човешка преценка.

Русия разглежда изкуствения интелект (ИИ) като стратегически компонент на националната си сигурност. Държавното ръководство, включително президентът Путин, активно подкрепят развитието и внедряването на ИИ технологии във военната сфера. Въпреки настоящите предизвикателства на бойното поле, руската отбранителна индустрия продължава да инвестира в специфични приложения на ИИ с цел намаляване на технологичния разрив спрямо водещите световни сили. Подходът към научноизследователската и развойната дейност е централизирано управляван, частично се адаптират успешни практики от западни и китайски модели, за да се използват достиженията на гражданския сектор.

На практика Русия внедрява ИИ в системи за целеуказване и разузнаване, особено в безпилотни оръжейни комплекси, където опитът от конфликта в Украйна показва, че подобни технологии предоставят осезаемо предимство. Въпреки това санкциите и структурните ограничения, свързани с достъпа до полупроводници и модерни технологии, възпрепятстват широкото използване.

В областта на военната роботика Русия отделя значителни ресурси за разработване на безпилотни летателни, подводни и наземни апарати. Наблюдава се тенденция към увеличаване на автономността с напредъка на софтуерните решения. Примери като бойния модул на „Калашников“, базиран на невронни мрежи, и танка „Уран-9“ показват стремеж към интеграция на ИИ за автономно разпознаване и вземане на решения. Въпреки че някои системи все още са в процес на тестване и не са напълно внедрени във въоръжените сили, поставени са конкретни цели за роботизация, предвиждащи до 2025 г. 30% от военната техника да бъде роботизирана.

Русия демонстрира интерес към технологии за мрежова координация на роботизирани системи, макар че в момента тези концепции са в ранен етап на развитие. Особено внимание заслужава автономната подводна апаратура „Статус-6“, предназначена за доставка на ядрени оръжия, която се отличава с висока степен на автономност и потенциално стратегическо значение.

Отбранителните системи с елементи на автономност са интегрирани в руската армия от десетилетия. Системи като „Дрозд“, „Арена“ и „Афганит“

осигуряват активна защита на бронетехника, докато зенитно-ракетните комплекси С-400 „Триумф“ разчитат на ИИ за автономно откриване и проследяване на цели. Русия работи по автоматизирани системи за управление, които обединяват различни комплекси в единен информационен домейн, като оптимизират разпределението на ресурси и ускоряват процеса на вземане на решения.

Изкуственият интелект се внедрява и в инфраструктурата за командване и управление, използват се математически модели за подпомагане на стратегическото планиране и прогнозиране на резултати от операции. Националният център за управление на отбраната играе ключова роля в интеграцията на данни и анализи за военно-политическата ситуация, като значителна част от държавните инвестиции в ИИ са насочени към системи за анализ на данни и подпомагане на вземането на решения.

Русия прилага ИИ в така наречените „сиви зони“ – ситуации, които се намират между война и мир, като използва кибероперации, електронна война, пропагандни кампании и дезинформация. Автономните системи от типа на платформата „Билина“ за електронна борба показват потенциал за напълно автоматизирано идентифициране и неутрализиране на цели. В този контекст Русия отдава приоритет на развитието на технологии, способни да обезсмислят информационното превъзходство на противниците.

Финансирането на изследвания и разработки в областта на военния ИИ е относително стабилно, макар и скромно на фона на водещите държави. По-голямата част от инвестициите идват от държавата, но значителен дял от бюджета за научноизследователска и развойна дейност остава секретен. През последното десетилетие Русия предприе стъпки за стандартизация и координация на усилията чрез целеви програми и ежегодни конференции, а Фондацията за перспективни изследвания работи по внедряване на нови стандарти и технически изисквания.

Взаимодействието между държавния и частния сектор е ограничено от доминиращата роля на правителството, което консолидира отбранителната индустриална база и бързо насочва изследванията според военните нужди. Липсата на конкуренция може да ограничи иновациите, но системата позволява гъвкави реакции при възникване на нови изисквания.

Човешкият капитал в Русия е сравнително добре образован, но производството на знания и иновации, измерено чрез патенти и научни публикации, остава ниско. Въпреки тези предизвикателства, Русия може да използва гъвкав подход за разработване на по-евтини приложения на ИИ, особено в области като информационни операции и интернет ботове.

Русия следва последователна стратегия за интеграция на изкуствен интелект във военната сфера, като съчетава централизирано управление, целеви програми и стремеж към иновации, въпреки структурните и ресурсните ограничения. Този модел позволява постепенно усъвършенстване на военните способности и адаптация към съвременните предизвикателства на информационната и киберсигурността.

Китай разглежда изкуствения интелект като стратегически инструмент за постигане на технологично превъзходство над потенциални противници. Концепцията за „интелигентна война“ поставя ИИ в центъра на трансформацията на Народноосвободителната армия (НОАК), съчетавайки технологични иновации, организационни промени и военна доктрина (Lee 2024; Kania 2019). Те изградиха мащабна държавно координирана програма за развитие на ИИ, която се финансира и на държавно, и на частно ниво (Lee 2024). Тези технологии се използват активно за симулации, логистика, проектиране на оръжейни системи, автоматизиран разузнавателен анализ и управление на командни структури (Kania 2019). Паралелно с това Китай инвестира в разработки, свързани с когнитивната война и информационното влияние (Kostka 2021).

Китай развива широка гама безпилотни летателни апарати – от тактически до стратегически, с приложения за наблюдение, разузнаване, електронна война и нападателни мисии. Особен акцент е поставен върху рояци от дронове, управлявани чрез алгоритми за машинно обучение и координирани в реално време.

В сухопътните сили се разработват автономни танкове, бронирани машини и логистични роботи, чиято цел е да намалят рисковете за личния състав и да предоставят нови тактически възможности. Използването на роботизирани системи за евакуация на ранени съчетава бойни и хуманитарни функции (Kostka 2021).

В морската сфера Китай инвестира в автономни подводни апарати и надводни платформи, пригодени за разузнаване, противолодъчни действия и миночистене. Използването на ИИ позволява операциите да се извършват и при ограничени комуникации, което увеличава стратегическата автономност (Allen 2020).

Всички тези системи се вписват в доктрината за „интелигентна война“, при която въздушните, наземните и морските платформи се свързват в мрежово центрични операции, базирани на обмен на данни и съвместно управление (Kania 2019).

Китай официално се обявява за ограничаване или дори забрана на автономните летални оръжия (LAWS). Въпреки това страната продължава активно да разработва и експериментира с различни автономни бойни платформи.

След 2022 г. **Украйна** трансформира методите за военни иновации, като премина от държавно доминирана система към внедряване на технологии чрез частния сектор и стартиращи предприятия. Административната рамка е реформирана, за да бъдат намалени бюрократичните бариери. Това ускорява внедряването от години до седмици, особено в областта на безпилотните системи. За целта Украйна създаде Център за иновации и отбранителни технологии (CIDT) и Сили за безпилотни системи, които интегрират AI в бойната практика. Министерството на дигиталната трансформация лансира Център за върхови технологии, със заданието да подпомогне развитие на ИИ в отбраната, киберсигурността, здравеопазването и администрацията. От 2023 г. е приета пътна карта и бяла книга за изкуствен интелект. Беше разработена цифровата платформа за ситуационна осведоменост *Delta,* която интегрира разузнавателни данни от дронове, сателити, сензори и партньорски източници, представяйки ги на динамична бойна карта—ключов инструмент за командване и управление. Украйна събира значителни обеми бойна информация – видео, сензорни данни и статистика. Тези данни са стратегически актив при сътрудничество със западни партньори като Palantir и се използват за бойна AI аналитика и целево планиране. Украйна е първата страна, която прилага широка употреба на AI базирани рояци дронове (повече от 100 пъти използвани в бойни условия) чрез компанията Swarmer.

Германският подход е по-умерен и консервативен. Затова ИИ за отбрана е разглеждан „като инструмент – смирен слуга“ на човека и основните проекти са фокусирани върху подпомагане на решения, усъвършенстване на C4ISR системи и предиктивна поддръжка. Примери за вече внедрени решения са отворена система за анализ на разузнавателни данни при кризи, ИИ сензори за предупреждение на вертолети и интелигентна обработка на изображения за ракети. След тридесетгодишно изоставане Бундесвера въвеждат нови курсове по ИИ и еквивалентни програми (бакалавър/магистратура). Тези мерки са добър пример за балансиран, но бавен преход – чрез обучение на персонала, постепенни системни промени и сътрудничество в рамките на ЕС и НАТО.

Турция отдавна смята автономните системи за бъдеще на войната. Турската отбранителна политика концентрира усилията си върху „подобряване на възможностите на автономните системи, сензори и системи за

подпомагане на решенията“. Приоритетни са безпилотните летателни апарати (UAV) и други безпилотни платформи – целево усъвършенстване на техния капацитет за разузнаване, бойно приложение и групови операции. Турция вярва, че „бъдещето на модерната война“ е в роботиката, затова обвързва военната индустрия с динамичния си граждански техно сектор. Създадената AI система (Националната AI стратегия 2021 – 2025) търси синергии между държавните програми, компании и академии, като се увеличават разходите за развойна дейност и подготовката на кадри. Турският модел показва ползи от постоянни инвестиции в стартиращи ИИ фирми и в центрове за таланти (например AI Talent Cluster, SAYZEK).

Гърция се фокусира върху три линии – участие в европейски проекти, закупуване на модерна техника с ИИ възможности и обучение на персонал. Пример за добра практика е активното включване в европейски програми за съвместни разработки и обмен на знания. Тези сравнения и добри практики са обобщени в таблица 26, в която са представени ключовите особености на някои от водещите страни.

Таблица 26. Сравнение на подхода на някои държави

Държава	***Стратегическа цел***	***Основен фокус***	***Подход***
Русия	Война в Украйна и конкуренция със Запада	AI подсилване на C4ISR, разузнаване и автономни оръжия (дронове, роботи)	Силен държавен контрол и финансиране
Китай	Национално движение „интелигентна война“	Автономни платформи (UAV, море, космос), оптимизация на логистика и командване	Интеграция военна –цивилна индустрия; централизирано планиране; инвестиции в изследвания
Германия	Вътрешнополитическа легитимация и съюзнически ангажименти	За вземане на решение, разузнаване, (C4ISR)	Човекът във водещата роля, бавна еволюция
Турция	Модернизация на армията, регионални амбиции, убеденост в силата на UAV	Безпилотни системи (въздух, море, суша), групови операции, ИИ – подобрено вземане на решения	Създаване на местна AI система (индустрия + академия), развойна дейност,

Държава	*Стратегическа цел*	*Основен фокус*	*Подход*
			централизация чрез агенции
Гърция	Регионални конфликти и нужда от технологична модернизация	Разузнавателни AI решения, модерни системи с ИИ, обучение на персонала	Активно участие в проекти –европейски и НАТО, интеграция на образование и иновации

Сравнението на разгледаните държави показва, че няма единен „правилен“ подход – демократичните страни като Германия и Гърция залагат на постепенно въвеждане и прозрачност, докато авторитарните (Русия, Китай) предпочитат централизирано планиране и мащабни инвестиции. Независимо от различията, се виждат и няколко общи фактори. Първо, повечето нации се фокусират върху автономни/безпилотни системи. Второ, важни области са аналитичните системи (разузнаване и C2), прогнозната поддръжка и логистиката. Трето, почти всички държави използват комбиниран подход чрез взаимодействие между военни и граждански технологии – например чрез иновационни клъстери, съвместни разработки в ЕС и НАТО или академични партньорства. Накрая, подчертава се необходимостта от обучение – глобалната статистика показва, че обучението и повишаването на квалификацията на военнослужещите и цивилните служители се осъществява бавно, но е ключово за ефективното прилагане на изкуствения интелект.

В условията на променлив геополитически пейзаж държавите по света разпознават изкуствения интелект като ключов елемент за надмощие. В рамките на стратегията НАТО 2030 Алиансът заяви необходимостта да запази технологичното си предимство, което породи инициативи като ускорителят за иновации в отбранителния сектор в северноатлантическия регион (DIANA) и иновационен фонд[36]. Анализът на тези и други подобни проекти показва шест основни линии на усилия при внедряване на такива инструменти в отбраната:

– конкуренцията със стратегически противници определя и националния подход за развитие на изкуствения интелект. Все пак повечето държави

[36] Министерство на отбраната. Ускорител за иновации в отбранителния сектор в северноатлантическия регион (DIANA). Достъпно на: https://www.mod.bg/.

оперират в човеко-ориентиран и данни-ориентиран режим, при които технологиите трябва да служат на човешкия фактор;

– приоритетите по разработката на изкуствен интелект често са обвързани с безпилотни системи и разузнаване – например дронове с автономни функции – както и с поддръжката на информационни инфраструктури, логистика и командване и управление (C4I);

– за въвеждането на ИИ много държави създават междуведомствени структури или нови агенции, натоварени с координация на проектите на всяко ниво (централно, военно, технологично);

– финансирането на военен изкуствен интелект остава непрозрачно – някои страни обособяват специални бюджети, други разчитат на общи военни програми, а трети – на смесено публично-частно финансиране;

– вече се прилагат решения основно при безпилотни платформи и системи за разпознаване и целеуказване. Почти половината анализирани държави използват изкуствения интелект за прогностична поддръжка, логистика и обучение;

– подготовката на кадри със специални AI умения все още се развива бавно: около една трета от страните се фокусират само върху военни обучения, а други включват и цивилни специалисти или проекти за повишаване на цифровата грамотност.

Накратко, глобалната рамка за внедряване на изкуствен интелект в сигурността и отбраната се характеризира с мултиполярна динамика, в която както демокрации, така и авторитарни режими, се стремят към технологични пробиви. Всички анализирани страни показват, че изкуственият интелект ще има продуктивно допълваща роля във военното решение, но са предпазливи в баланса „човек в цикъла“. В тази дълга надпревара за изкуствения интелект доминират усилията за разработка на автономни оръжейни системи, интегриране на данни и оптимизиране на войсковите комуникации, което изисква и съответна адаптация на организацията, политиките и подготовката на войските.

В периода 2022 – 2024 г. НАТО поставя дигиталната трансформация като централен елемент в своята стратегическа рамка за усъвършенстване на технологичните си предимства в сферата на сигурността и отбраната. Основната визия на Алианса е интегрирането на цифрови технологии и изкуствен интелект (AI) за мултидомейнни операции, повишена оперативна съвместимост и подобрена ситуационна осведоменост. Чрез изграждането на Digital Backbone се цели интеграция на командване и контрол, базирани върху реално времени сензорни и аналитични данни, като това осигурява ускорено

и по-информирано вземане на решения на стратегическо и оперативно ниво. Създаването на стандартизирана рамка за съвместимост между всички системи и звена на НАТО е друг ключов приоритет. В областта на човешкия капитал НАТО развива дигитално подготвен личен състав чрез целенасочени програми за обучение, развитие на AI компетенции и иновационна култура.

Световните сили като САЩ, Китай и Европейския съюз водят интензивна технологична надпревара за доминиране в сферата на изкуствения интелект и производството на изчислителна мощност. САЩ разчитат на технологични гиганти като OpenAI и Oracle, докато Китай прилага модел на държавно финансиране и бързо разширяване на AI инфраструктурата. Европейският съюз, макар и с известно изоставане, фокусира усилия върху стандартизация и създаване на етична рамка за развитие на AI технологиите. Проектът Stargate на стойност 500 милиарда долара, обявен от администрацията на Доналд Тръмп, демонстрира мащаба на американските амбиции в тази област. От своя страна ЕС реагира с програма „Цифрова Европа“, която финансира изграждането на високотехнологични центрове в ключови градове като Париж, Берлин, Амстердам, Мадрид, Милано и Прага.

Всички поставят дигиталната трансформация и изкуствения интелект в основата на своите усилия за гарантиране на сигурността и отбраната. Ключовите направления включват усъвършенстване на оперативната съвместимост, интеграция на данни и AI в командването и управлението, както и развитие на човешкия фактор чрез обучение и иновации. Различията се проявяват най-вече в подходите към организационната култура и структурата на обучение, като НАТО акцентира върху колективната сигурност и интероперативността, докато американците залагат на вътрешна модернизация и индивидуално развитие на кадрите.

ПРЕПОРЪКИ ЗА РАЗВИТИЕ НА AI В ОТБРАНАТА НА БЪЛГАРИЯ

Бебето се учи първо да пълзи,
после да върви и накрая да тича.
Сега сме в етапа на пълзенето, що се отнася до AI.

Дейв Уотърс

На базата на международния опит и факта, че България е малка страна с ограничени ресурси в близост до конфликтната зона Черно море с наличие на хибридни и информационни заплахи и има съюзни ангажименти към НАТО и ЕС, се вижда, че страната е особено уязвима на дезинформация и кибератаки. Тези обстоятелства изискват интегрирането на AI не просто като технологично нововъведение, а като стратегически „щит“ за националната сигурност.

Тази глава ще очертае ключовите направления, в които AI може да донесе съществена добавена стойност – интеграция на интелектуални решения в разузнаването и C4ISR системите, усъвършенстване на киберотбраната и информационната сигурност, оптимизация на логистиката и снабдителните вериги, въвеждане на напреднали методи за обучение и симулации, както и развитие на автономни платформи с фокус върху достъпни и модулни технологични решения за повишаване устойчивостта на отбраната.

14.1. Нормативна рамка за употреба на изкуствен интелект в сигурността и отбраната

На първо място, внедряването на изкуствен интелект в отбранителните системи изисква ясна нормативна архитектура, осигуряваща баланс между технологично развитие, спазване на международното хуманитарно право и защита на фундаменталните права. Необходимо е да се анализират ключовите правни документи на НАТО и Европейския съюз, както и националната рамка на Република България, регламентиращи разработването и прилагането на ИИ в сектора за сигурност и отбрана.

Атлантическият съюз развива всеобхватна рамка за управление на изкуствения интелект, основана на стратегически документи, етични принципи и институционални механизми за контрол. През октомври 2021 г. Съюзът прие Стратегия за изкуствен интелект, която поставя основите на подхода на Алианса към технологиите с двойна употреба и залага четири стратегически цели: ускоряване на прилагането на изкуствения интелект при спазване на отговорността; гарантиране на оперативна съвместимост между системите на държавите членки; защита на технологиите от злонамерено използване и идентифициране на заплахи от противниковото използване на AI.[37]

През юли 2024 г. Стратегията е актуализирана, за да отрази развитието на генеративните модели и новите фундаментални системи. Ревизираният документ разширява обхвата с включване на рисковете от дезинформация и информационни операции, свързани с насилие на полова основа, като паралелно с това засилва фокуса върху оперативната съвместимост на AI системите в рамките на Алианса.[38]

Централна роля в нормативната рамка играят принципите, одобрени през 2021 г. и препотвърдени през 2024 г. – законност, отговорност и отчетност, обяснимост и проследяемост, надеждност, управляемост и смекчаване на пристрастията. Тези принципи са обвързващи за всички държави членки при разработката на военни системи с изкуствен интелект.

От практическа гледна точка, през октомври 2022 г. Съветът на NATO одобрява план, който създава кохерентен подход към развитието и защитата на автономните системи в съответствие с нормите на Алианса и международното хуманитарно право.[39] Паралелно е създаден Data and Artificial Intelligence Review Board (DARB), който операционализира плана чрез разработване на инструменти за сертифициране и оценка на риска.[40]

---

[37] · NATO, 2021. *Summary of NATO's Artificial Intelligence*. Available from: Strategyhttps://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2021/10/22/summary-of-the-nato-artificial-intelligence-strategy.

[38]. NATO, 2024. *Summary of NATO's revised Artificial Intelligence Strategy*. Available from: https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/07/10/summary-of-natos-revised-artificial-intelligence-ai-strategy.

[39]. NATO, 2022. *Autonomy Implementation Plan*. Available from: https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2022/10/13/summary-of-natos-autonomy-implementation-plan.

[40]. NATO, 2022. Data and Artificial Intelligence Review Board. Available from: https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2022/10/13/natos-data-and-artificial-intelligence-review-board.

От институционална гледна точка, ключово значение има Defence Innovation Accelerator for the North Atlantic (DIANA), учреден през април 2022 г. на срещата на външните министри на Алианса (NATO, DIANA Charter, 2022). DIANA функционира чрез мрежа от акселератори и тестови центрове (над 180 локации в Алианса), които подпомагат дълбоките технологии с двойна употреба чрез предоставяне на грантове и достъп до тестова инфраструктура (NATO, Defence Innovation Accelerator for the North Atlantic, 2025). Съществувайки като допълнение, NATO Innovation Fund (NIF) е първият многосуверенен фонд за рисков капитал с капитализация от 1 милиард евро, инвестиращ в стартиращи компании от области като ИИ, квантови технологии и автономни системи (NATO, NATO Innovation Fund, 2022).

Европейският съюз подхожда към регулирането на ИИ чрез подход, регламентиран в Regulation (EU) 2024/1689 – известен като AI Act – публикуван в Официалния вестник през юли 2024 г. и влязъл в сила на 1 август 2024 г.[41] Регламентът въвежда риск-базирана класификация на ИИ системите, като забранява „неприемливия риск“ (напр. социален рейтинг от обществени органи, манипулационни техники) и налага строги изисквания за „високорискови“ системи.

От съществено значение за отбранителния сектор е разпоредбата на чл. 2, ал. 3 от Регламента, която изключва от обхвата ИИ системи „пуснати на пазара, въведени в експлоатация или използвани... изключително за военни, отбранителни или национални служби за сигурност цели“. Изключването се основава на член 4, параграф 2 от Договора за Европейския съюз, съгласно който националната сигурност остава отговорност на отделните държави членки.

Въпреки това, когато разработена за военни цели ИИ система се използва временно или постоянно за други цели (граждански, хуманитарни, правоприлагане или обществена сигурност), тя попада в обхвата на Регламента. Аналогично, системите с двойна употреба, пуснати на пазара за граждански цели, но използвани впоследствие за отбрана, остават под регулаторния режим на AI Act.

На политическо ниво ЕС развива координиран подход чрез координационен AI план (2018, актуализиран 2021), като отбранителните аспекти са регламентирани преимуществено в рамките на Общата политика за сигурност и отбрана (ОПСО) чрез European Defence Fund и програма Digital Europe. Тези

41. EUROPEAN UNION, 2024. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. Available from: https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

механизми подпомагат развитието на „отбранителни ИИ способности“, без да създават отделна нормативна рамка, паралелна на тази на NATO.[42]

Българската политика в областта на ИИ се конституира на първо място от *Концепцията за развитието на изкуствения интелект в България до 2030 г.*, приета с Протокол № 72 на Министерския съвет от 16 декември 2020 г. Документът идентифицира сигурността и отбраната като приоритетни сектори, като посочва, че „в системите за сигурност и отбрана изкуственият интелект навлиза широко с възможността за обработка и интерпретация на многоканална и огромна по обем информация... и се очаква те да предлагат в реално време възможни сценарии и алтернативи за реакция, като се самообучават“.[43]

Концепцията залага шест стратегически цели, сред които изграждане на научна инфраструктура, развитие на капацитет за изследвания и създаване на нормативна база за „надежден ИИ“ в съответствие с международните стандарти. Тя предвижда триетапно изпълнение: краткосрочен етап (2021 – 2023) с фокус върху пилотни проекти; средносрочен (2024 – 2027) за мащабиране на инфраструктурата и дългосрочен (2028 – 2030) за консолидация и стандартизация.[44]

На институционално равнище Министерството на отбраната на Република България развива системна политика за интегриране на ИИ чрез меморандуми за сътрудничество с научни организации. На 7 януари 2025 г. е подписан Меморандум за сътрудничество между Министерството и Института за компютърни науки, изкуствен интелект и технологии (INSAIT)[45] към Софийски университет „Св. Климент Охридски“. Документът с петгодишен срок на действие регламентира партньорство в областите иновации,

---

42. EUROPEAN COMMISSION, 2021Coordinated Plan on Artificial Intelligence. Available from: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review.

43 МИНИСТЕРСТВО НА ТРАНСПОРТА, ИНФОРМАЦИОННИТЕ ТЕХНОЛОГИИ И СЪОБЩЕНИЯТа, 2020. *Концепция за развитието на изкуствения интелект в България до 2030 г.* Достъпно на: https://www.mtc.government.bg/sites/default/files/koncepciyazarazvitienaiivbulgariyado2030.pdf.

44. БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ, 2020. *Изкуствен интелект за интелигентен растеж. Стратегия за развитието на изкуствения интелект в България до 2030 г*. Предварителна визия. Достъпно на: https://www.bas.bg/wp-content/uploads/2020/07/Proposal-National-Strategy-AI-2030-24June2020.pdf.

45. BGNES, 2025. МО и INSAIT подписаха Меморандум за сътрудничество в сферите на иновациите и изкуствения интелект. Достъпно на: https://www.bgnes.bg/mo-i-insait-podpisaha-memorandum-za-satrudnichestvo-v-sferite-na-inovaciite-i-izkustveniq-intelekt.

компютърни науки, изкуствен интелект, киберсигурност и развойна дейност за модернизация на въоръжените сили. През 2025 г. са подписани допълнителни меморандуми с Университета за национално и световно стопанство (УНСС) и Сдружение „Българска отбранителна индустрия – акад. Стефан Воденичаров“, които създават триъгълника „армия – академия – бизнес“ като модел за технологичен трансфер.

Програмата за развитие на отбранителните способности на въоръжените сили на Република България до 2032 г. (Програма 2032), одобрена от Народното събрание, определя като приоритет „въвеждане на изкуствен интелект, обработка на големи масиви данни и квантови технологии в областта на киберотбраната и военните комуникации“.[46] Програмата предвижда поддържане на разходите за отбрана на ниво не по-малко от 2% от БВП след 2024 г. и поне 20% от бюджета за придобиване на нови въоръжения и оборудване.

България активно участва в инициативата DIANA на NATO чрез два одобрени тестови центъра – Институтът за изследване на големи данни за умното общество (GATE) за изкуствен интелект и Институтът по отбрана „Проф. Цветан Лазаров“ за автономни системи (Министерство на иновациите и растежа, Defence Innovation Accelerator for the North Atlantic, 2025). Участието в DIANA позволява на български иноватори достъп до грантове в размер на 100 000 € и мрежа от над 180 тестови центрове в Алианса.

Националната инфраструктура се допълва от проект BRAIN++ (Bulgarian Research and Innovation Network for AI), който спечели финансиране от 90 милиона евро по линия на EuroHPC Joint Undertaking през март 2025 г. Проектът, реализиран от INSAIT и София Тех Парк, предвижда изграждането на фабрика за изкуствен интелект със суперкомпютър Discoverer++, оптимизиран за генеративен ИИ, роботика и космически наблюдения. Фабриката ще функционира като one-stop-shop за държавни институции, образователни организации и частни компании за отговорно въвеждане на иновации с ИИ.

Паралелно, с Решение № 415 на Министерския съвет от 25 юни 2025 г., се създава Център за иновации в отбраната като звено към „София Тех Парк“ АД. Центърът се финансира с 30 млн. лв. (15,3 млн. евро) по Програма „Научни изследвания, иновации и дигитализация за интелигентна трансформация“ 2021 – 2027 (ПНИИДИТ) и има за цел тестване и сертифициране

[46.] МИНИСТЕРСТВО НА ОТБРАНАТА, 2022. *Програма за развитие на отбранителните способности 2032 г.*

на технологии с двойна употреба, с фокус върху безпилотни летателни средства и усъвършенствани отбранителни системи. Центърът действа в партньорство с Института по отбрана „Проф. Цветан Лазаров“, съгласувайки научноизследователската дейност с националните военни приоритети и целите на НАТО и ЕС.

Нормативното пространство за прилагане на ИИ в отбраната се характеризира с многостепенна архитектура: НАТО осигурява стратегически рамки и принципи за отговорно използване; ЕС регулира гражданските приложения чрез AI Act, като изключва ексклузивно военните цели, но запазва контрол върху двойната употреба; България синхронизира националните си приоритети с тези на Алианса чрез Концепцията до 2030 г., Програма 2032 и конкретни институционални механизми (DIANA, BRAIN++, Център за иновации). Тази интегрирана рамка създава условия за технологичен суверенитет, съвместимост със стандартите на Алианса и насърчаване на иновационната система между държавата, академичните институции и частния сектор.

В тази връзка интересен пример за настоящото изследване може да се даде с Армения, където NVIDIA инвестира 500 милиона долара в изграждането на фабрика за изкуствен интелект. Този проект представлява трансформация на енергийния излишък в цифрови активи, които някои експерти наричат „новият петрол“. Арменска фабрика с мощност 100 мегавата ще произвежда изчислителна мощност – критичен ресурс за глобалното обучение на AI системи. Арменският успех се дължи на комбинация от благоприятни данъчни политики, наличието на млади квалифицирани специалисти и активното участие на арменската диаспора. Страната привлича международни технологични компании и професионалисти, включително голяма вълна от специалисти, напуснали Русия след инвазията в Украйна.

Таблица 27. Ключови характеристики на AI инфраструктурата за национална сигурност

Характеристика	***Описание***	***Влияние върху сигурността***
Енергийна интензивност	Висока консумация на енергия	Зависимост от стабилни енергийни източници
Изчислителна мощност	Голяма обработваща способност	Надмощие в обработка на данни и модели
Данъчна сигурност	Защита на информационните потоци	Спиране на вражеско проникване

Характеристика	*Описание*	*Влияние върху сигурността*
Интеграция в отбраната	Включване във военни системи	Предимство в модерната война

В заключение може да се каже, че нормативната рамка е налице и е от решаващо значение за регулирането на характеристиките на AI инфраструктурата за национална сигурност. Тя не само дефинира изискванията към технологиите и процесите, но също така очертава стандартите за компетентност и професионална подготовка на кадрите, които работят с тях. Синхронизирането между законовите регулации и развитието на длъжностните профили осигурява ефективна интеграция на AI решенията във всички нива на отбраната, като същевременно повишава оперативната готовност и устойчивост на националната сигурност. В този контекст формирането на нови хибридни компетенции и непрекъснатото обучение се превръщат в стратегически приоритет за държавата и нейните отбранителни институции.

14.2. Бъдещето на военните длъжности и компетенции в контекста на изкуствения интелект

Развитието на изкуствения интелект и подготовката на кадри, способни да го операционализират, се очертават като фундаментални елементи на технологичната мощ на съвременните нации. В този контекст, военните организации преживяват коренна трансформация на професионалните изисквания, при която традиционните бойни умения се комбинират с дълбока технологична грамотност.

Бързото разпространение на изкуствения интелект води до вълна от изследвания, които измерват и прогнозират неговите въздействия върху пазара на труда. В началото на март 2026 г. Anthropic публикува доклад за въздействието на изкуствения интелект върху пазара на труда, който за първи път въвежда показател за измерване по наблюдавана реална употреба на AI[47].

Основни изводи в него са, че AI все още е далеч от пълния си потенциал – реалното използване покрива само част от задачите, които теоретично може да автоматизира. Най-много AI се използва в професията на програмисти, обслужване на клиенти, финансови анализатори, оператори за

[47] MASSENKOFF, M.; MCCRORY, P., 2026. Labor market impacts of AI: A new measure and early evidence. [online]. *Anthropic*. Available from: https://www.anthropic.com/research/labor-market-impacts

въвеждане на данни. AI вече влияе на определени професии, но ефектите са ограничени. Няма масово изместване на работници, поне засега.



Фигура 67. Теоретичен капацитет и наблюдавана употреба на ИИ по професионални категории

Графиката ясно показва къде изкуственият интелект ще предизвика революция и къде човешкият фактор остава критичен. Секторите, предлагащи сигурност, охранителни услуги, монтаж и ремонт показват много нисък теоретичен обхват на изкуствения интелект. Това означава, че тактическите длъжности на терен, изискващи физическа адаптивност, сензорна ориентация и комплексни морални решения в хаотична среда, ще продължат да разчитат основно на човешки персонал, но административните и аналитичните длъжности ще се дигитализират агресивно, но все пак ще изискват оператори за работа с AI.

В българския контекст няколко водещи висши учебни заведения стартираха учебни програми с профил „Изкуствен интелект“, фокусирайки се върху подготовката на кадри за управление на технологии като обработка на естествен език, невронни мрежи и генетични алгоритми (Valchev & Petrov 2022). Паралелно с това значително внимание се отделя на специфични отрасли на отбраната, като военната медицина например. Застаряващото население и необходимостта от подобряване на здравните политики, включително тези за поддържане на боеспособността на личния състав, подчертават важността от ефективно използване на технологиите. Дейности като диагностика и скрининг на медицински изображения, персонализирани планове за лечение и „уелнес“ мониторинг на здравословното състояние на военнослужещите придобиват критично оперативно значение (European Defence Agency 2023).

Развива се и телемедицината, където изкуственият интелект играе ключова роля в отдалечената диагностика и консултиране, решавайки проблеми с достъпа до специализирана медицинска помощ в отдалечени гарнизони и военни обекти. Във военната фармацевтика AI ускорява откриването на нови лекарства и оптимизира клиничните изпитания, както и мониторинга на безопасността на лекарствените средства (European Defence Agency 2023).

Технологията на ИИ все още не е достигнала пълния си потенциал. Крайната цел на изследователите, работещи върху „интелигентните“ машини, е достигането на изкуствен интелект, сравним с човешкия, а в перспектива – и неговото превъзхождане. Този технологичен напредък създава нови възможности за приложение, но изисква непрекъснато обновяване и оптимизиране на множество публични и частни институции, генерирайки нови потребности на пазара на труда и нови военни длъжности, повишавайки производителността и в крайна сметка – оперативната ефективност.

Бъдещето на военните длъжности е неразривно свързано с адаптацията към дигиталната среда и интеграцията на изкуствен интелект във всички нива на командване. Вместо традиционните професии, акцентът пада върху хибридни роли и компетенции, които изискват синергия между военноспецифични знания и технологична експертиза. Развитието на автономните системи и интелигентните технологии създава шест ключови **нови длъжности във военната професия, свързани с AI**:

● **Специалисти по данни и военна аналитика**. Търсенето на експерти по анализ на данни, изследователи и анализатори на разузнавателни данни ще се интензифицира. Тези специалисти са критични за обработката

на огромни потоци от разузнавателна информация, генерирани от съвременните сензорни системи. Те гарантират, че използваните от алгоритмите на AI данни са точни, пълни и оперативно релевантни.

● **Инженери и разработчици на AI** за отбранителни системи. Това се очертава като една от най-значимите длъжности в бъдещите въоръжени сили. Тези специалисти осигуряват функционалността на системите с изкуствен интелект, включително техния дизайн, актуализация, отстраняване на грешки и конфигуриране за специфични военни приложения.

● **Специалисти по тестване и валидиране на софтуер** (QA/Test Engineers). Жизненоважни за разнообразни отрасли във военната индустрия, тези експерти отговарят за софтуерното и хардуерното тестване на системи, като търсят уязвимости и проверяват тяхното въздействие върху потребителя (военнослужещия) и мисията. Те планират процесите на тестване и определят критериите за ефективност и сигурност в условията на висока несигурност.

● **Експерти по киберсигурност** и информационна отбрана. С усложняването на технологичните системи заплахите от хакерски атаки и злоупотреба с чувствителни данни нарастват експоненциално. Експертите по киберсигурност стават критично важни за всяка военна организация, отговарящи за защитата на цифровата инфраструктура и осигуряването на сигурна дигитална обмяна на информация в условията на хибридни заплахи.[48]

● **Специалисти по роботика и автономни системи**. Разработката и оперативната поддръжка на безпилотни летателни апарати (БПЛА), наземни роботизирани платформи и морски автономни системи изискват нов клас технически специалисти, комбиниращи познания по мехатроника, AI и оперативно изкуство.

● **Специалисти по енергийни технологии**. В отговор на климатичните промени и изчерпването на ресурси ще нарасне нуждата от инженери по възобновяема енергия, проектанти на енергийно ефективни съоръжения, успоредно с нарастващата нужда от енергия на центровете за изкуствен интелект, както и на интегриращи AI за оптимизация на енергийния разход (European Defence Agency 2023).

Всичко това налага формиране на определени **ключови дигитални компетентности** в AI ориентираната отбрана. Военнослужещите, способни

[48] NATO, 2021. Summary of NATO's Artificial Intelligence Strategy. Available from: https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2021/10/22/summary-of-the-nato-artificial-intelligence-strategy.

да разработват, поддържат и оптимизират AI базирани системи, придобиват централно значение и трябва да:

– създават алгоритми за тактическо и оперативно планиране, включващи машинно обучение и дълбоки невронни мрежи;

– интегрират и работят с AI в съществуващите системи за управление и контрол;

– осигуряват защитата на военните мрежи и платформи чрез внедряване на интелигентни системи за откриване на аномалии;

– разработват технологии за киберотбрана, включващи автономни системи за реагиране;

– управляват безпилотни и автономни платформи и да контролират дистанционно управляеми системи с висока степен на автономия;

– анализират данни от сензори и камери в реално време;

– могат да анализират и интерпретират големи обеми данни в реално време;

– оптимизират логистичните вериги чрез интелигентно планиране;

– бъдат подготвени да оценяват съответствието на AI базирани решения с международното хуманитарно право и правилата за водене на бойни действия;

– са запознати с някои етични дилеми, свързани с делегирането на решения за използване на сила от машини.

Въпреки технологичния напредък, някои длъжности не могат да бъдат заменени от изкуствения интелект. Най-устойчиви за нуждата от човешкия фактор ще останат: командирите, вземащи решения в условия на морална неяснота; инструкторите, които обучават и мотивират; военните психолози, поддържащи психическото здраве на личния състав и военните медици, осигуряващи медицинска грижа. Тези роли са изградени върху етични принципи, емпатия, лидерство и автентичен човешки контакт – качества, които остават извън обхвата на изкуствения интелект в обозримо бъдеще. Дигитализацията спестява времевия ресурс на хората. Но от тях зависи за какво ще се изразходва останалото свободно време. Дали ще бъде по посока на личното и професионалното израстване или ще останат в застой, следва да станем свидетели тепърва на тези процеси.

Специфичните човешки компетенции, които остават неподвластни на автоматизация, включват:

● Лидерство и взимане на решения в условия на несигурност. Способността да се ръководят екипи в екстремални условия и да се взимат етични и отговорни решения, базирани на интуиция и морална преценка, не

може да бъде напълно автоматизирана. AI може да предостави опции, но крайното решение в контекста на сложния морален пейзаж на съвременните конфликти остава човешки прерогатив.

● Емоционална интелигентност и междукултурна комуникация. Военнослужещите трябва да поддържат висока степен на емоционална устойчивост и ефективна комуникация, особено в мултинационални операции и междукултурна обстановка. Способността за съпричастие, разбиране на невербални сигнали и изграждане на доверие са критични за коалиционните операции.

● Адаптивност и креативност. Бойното поле е динамична, нелинейна среда, изискваща бързо мислене, импровизация и нестандартни решения – качества, които AI все още не може да репликира в контекста на непредвидени сценарии.

Интеграцията на изкуствения интелект във военните професии поставя редица етични и организационни въпроси, свързани с автономността на оръжейните системи и делегирането на решения за използване на смъртоносна сила.

Въвеждането на инструменти с изкуствен интелект може да доведе до промяна в структурите на работа и потенциална загуба на традиционни длъжности, изискваща активна политика за преквалификация. Повишаването на стреса и натоварването на военнослужещите, работещи с високотехнологични системи (когнитивно натоварване), и най-вече постоянната необходимост от непрекъснато обучение и адаптация стават фундаментални характеристики на военната кариера.

Бъдещето на военните професии се очертава като сложен синтез между човешките способности и технологичните иновации. Изкуственият интелект ще трансформира начина, по който се водят войни, пренареждайки йерархиите на знанието и властта във въоръжените сили, но човешкият фактор ще остане централен за успеха на всяка операция.

За да успеят, съвременните армии трябва да инвестират в обучение и развитие на кадри, способни да работят в симбиоза с AI, развивайки технологичен интелект заедно с военното дело. Необходимо е да се създадат етични и съответните правни рамки за използването на автономни системи, гарантиращи съответствие с международното право и военните традиции. Поддържането на баланс между технологичните възможности и човешките ценности се осъществява, като се признае, че крайната цел на технологиите е подкрепа на човека, а не неговото заместване.

Ефективността на бъдещите въоръжени сили няма да се измерва само със скоростта на алгоритмите или мощността на сензорите, а със способността на човешките същества да използват интелигентните машини за постигане на стратегически цели, съхранявайки при това моралните основи на военното дело.

14.3. Българският потенциал в областта на AI

В съвременната среда на сигурност, доминирана от концепциите за „мулти-домейнна война“ и алгоритмично превъзходство, националният суверенитет все по-често се дефинира чрез способността на държавата да генерира и обработва възможно най-бързо информацията, за да съкрати цикъла НОРД и веригата на унищожение (Kill Chain). Изкуственият интелект престана да бъде просто екзотично поле на академични изследвания и се трансформира в критичен актив за националната сигурност, действащ като мултипликатор на силите във всички домейни на отбраната – от киберпространството до кинетичните операции.

За държава като България, разположена на ключов фланг на НАТО и външна граница на Европейския съюз, интегрирането на системи с изкуствен интелект в отбраната не е въпрос на избор, а императив за оцеляване и релевантност в променящата се среда на сигурност. България се намира в уникална точка. Страната ни разполага с рядка комбинация от стратегически активи – стабилна ядрена енергия, нововъзникваща научна инфраструктура от световна класа и традиции в образованието на ИТ специалисти. Въпреки това реализацията на потенциала в отбранителен капацитет е възпрепятствана от фрагментацията на усилията, политическата нестабилност и забавената адаптация на регулаторната рамка.

За Българската армия, която оперира с ограничен бюджет, интеграцията на изкуствен интелект изисква стратегически подход, фокусиран върху асиметрични предимства.

Инвестирането в AI подкрепени системи за интелигентно наблюдение е от първостепенно значение. През ноември 2025 г. България обяви програма за оборудване на военни формирования със симулатори за управление на дронове.

Усъвършенстването на мрежово-центричното командване е от ключово значение за интеграцията на различни сензорни платформи. Според НАТО (2024), цифровата трансформация на Алианса изисква технологична и културна промяна, при което данните се използват като стратегически

актив. Противодействието на по-могъщи противници изисква прилагане на принципите на асиметричната война. За България това означава развитие на децентрализирани структури за командване, които могат да функционират автономно при нарушени комуникации, както и инвестиции в евтини, но ефективни средства като дронове и преносими ракетни системи.

Целта на българската отбранителна стратегия трябва да бъде създаването на условия, при които никой противник не би могъл да влезе на национална територия и да постигне победа, без това да му коства недопустимо висока цена. Тази концепция се основава на многоизмерна интеграция с НАТО и ефективна подготовка на кадри. През април 2022 г. два български института станаха изпитвателни центрове за НАТО в областите на големи данни и изкуствен интелект (Kostadinova 2022).

В основата на всяка съвременна ИИ стратегия стои физическата и интелектуалната инфраструктура, способна да поддържа обучението на мащабни модели. През периода 2024 – 2026 г. България демонстрира качествен скок в това направление, преминавайки от позицията на потребител към позицията на създател на технологии. Централно място в новата технологична архитектура на страната заема Институтът за компютърни науки, изкуствен интелект и технологии (INSAIT), създаден в партньорство с швейцарските политехники ETH Zurich и EPFL. Кулминацията на това развитие е изборът на България за домакин на една от шестте нови европейски „AI фабрики“ в рамките на инициативата EuroHPC. Проектът BRAIN++, на стойност около 90 милиона евро, представлява сътрудничество между INSAIT и София Тех Парк, създаващ интегрирана производствена среда за интелигентност, където суровината (данни) се преработва чрез специализиран хардуер (GPU клъстери) в продукт с висока добавена стойност (генеративни модели).

В рамките на BRAIN++ се предвижда изграждането на нов суперкомпютър, условно наречен Discoverer++, оптимизиран специфично за AI натоварвания (AI workloads). За разлика от класическите суперкомпютри, предназначени за физически симулации, новата машина притежава архитектура, фокусирана върху тензорни изчисления, необходими за обучението на дълбоки невронни мрежи. Това осигурява на българската наука и отбрана суверенен капацитет за разработка на модели без необходимост от трансфер на чувствителни данни към облачни структури извън юрисдикцията на ЕС или НАТО (INSAIT, BRAIN++ Project Documentation).

INSAIT фокусира усилията си върху генеративни модели за роботика, наблюдение на Земята (Earth Observation) и езикови приложения. В контекста на отбраната, моделите за наблюдение на Земята са директно приложими

за анализ на сателитни снимки (IMINT), докато езиковите модели са ключови за обработката на разузнавателна информация (SIGINT/OSINT) и киберсигурността. Чрез стартиралата през юни 2025 г. програма за обмен с Масачузетския технологичен институт (MIT) на стойност 1 милион долара българските изследователи получават достъп до най-напредналите разработки в света, което ускорява внедряването на иновации в националния отбранителен комплекс[49].

Таблица 28. Българско участие в ключови международни отбранителни AI инициативи (2025)

Инициатива/ програма	***Участваща организация***	***Област на специализация***	***Значение за националната сигурност***
NATO DIANA	Институт GATE	Големи данни, борба с дезинформацията	Противодействие на когнитивни атаки и хибридни заплахи.
NATO DIANA	Институт по отбрана „Проф. Цветан Лазаров“	Киберзащита, C4I системи	Защита на комуникационната инфраструктура на армията.
EDF (Project ASTERION)	Wiser Technology	Подводни комуникации, алгоритми	Сигурност в морския домейн (Черно море).
EDF (Project MARTINA)	Wiser Technology	AI за анализ на сателитни изображения	Разузнаване (IMINT), автоматизиран анализ на заплахи
MIT Exchange Program	INSAIT	Генеративен AI, роботика	Трансфер на фундаментално знание и технологии

Докато INSAIT се фокусира върху фундаменталните алгоритми, Институтът „Големи данни в полза на интелигентно общество“ (GATE) към Софийския университет играе ролята на интегратор на данни. Като национален хъб на International Data Spaces Association (IDSA), GATE работи по създаването на „пространства от данни“ (data spaces), които са критични за

[49] Съвместна изследователска програма на INSAIT и MIT CSAIL https://insait.ai/bg/mit/

оперативната съвместимост между цивилните и военните системи[50]. В ерата на мрежово-центрираната война способността за сливане на данни от разнородни източници – сензори, дронове, социални мрежи, медицински картони – е определяща за ситуационната осведоменост.

Традиционната българска отбранителна индустрия, исторически специализирана в производството на боеприпаси и леко въоръжение, навлиза в период на наложителна трансформация. Войната в Украйна и необходимостта от модернизация на Българската армия катализираха процесите на сближаване между държавното управление, академичните звена и индустриалния сектор.

През януари 2025 г. Министерството на отбраната подписа Меморандум за сътрудничество с INSAIT, създавайки правна и организационна рамка за директно възлагане на задачи, свързани с изкуствения интелект, от страна на военните към академичните изследователи. Този процес бе надграден през юли 2025 г. с подписването на по-широкообхватен меморандум между Министерство на отбраната, Министерството на иновациите и растежа, Министерството на икономиката, Българската академия на науките (БАН) и Сдружение „Българска отбранителна индустрия“, целящ създаване на „нов тип партньорство“ и синергия между наука, бизнес и държава[51].

Центърът за изследвания по национална сигурност и отбрана към Българската академия на науките (ЦИНСО – БАН) стартира през юли 2025 г. мащабно проучване за картографиране на капацитета на институтите на Академията в сферата на отбраната и двойната употреба, което индикира преминаване от хаотични, ad-hoc проекти към планирано и стратегически управлявано научно-развойно портфолио.

Международното признание за българския потенциал в ИИ за отбрана намери израз във включването на страната в мрежата на НАТО DIANA. Два български центъра получиха акредитация като тестови – Институт GATE, определен като тестов център за големи данни и изкуствен интелект с фокус върху откриването на дезинформация и защитата на интелигентни градове, и Институтът по отбрана „Професор Цветан Лазаров“, предоставящ експертиза в киберзащитата и комуникационните системи.

---

[50] GATE Institute achieves IDSA certification of its Data Space connector. *GATE*. [online]. https://www.gate-ai.eu/en/news/gate-institute-achieves-idsa-certification-of-its-data-space-connector/

[51] Меморандум за подобряване на взаимодействието между изпълнителната власт и отбранителната индустрия. Достъпно на: https://www.bas.bg/?p=57691.

В частния сектор компанията Wiser Technology демонстрира конкурентоспособност чрез участието си в два консорциума, финансирани от Европейския фонд за отбрана (EDF) с общ бюджет над 33 милиона евро – проект ASTERION (подводни комуникации) и проект MARTINA (рамка за тестване на точността и надеждността на ИИ модули за анализ на сателитни изображения)[52].

България се оказва най-уязвимата държава по отношение на дезинформацията сред страните от Европейския съюз[53]. Ключови фактори, допринасящи за тази ситуация, включват недостатъчното качество на образователната система, ограничената медийна свобода и ниското ниво на междуличностно доверие в обществото, затова AI може да бъде използван като „щит“ срещу дезинформацията.

В дигиталния контекст на хибридните заплахи медийната грамотност включва разбиране на ролята на медиите в демократичното общество, критично отношение към медийното съдържание, способност за двупосочно въздействие с медийните институции и умения за създаване на потребителско съдържание. Според UNESCO, притежаването на тези компетентности позволява на индивидите да оценяват критично медиите като източници на информация, да вземат информирани решения и да участват активно в обществото.

Жизненият цикъл на дезинформацията е особено зависим от AI и включва четири основни етапа – създаване, разпространение, усилване и приемане. Изследванията показват, че след като дезинформацията бъде възприета, тя става устойчива на корекция, дори при представяне на фактическа информация. Борбата с дезинформацията изисква многостранен подход, включващ образователни инициативи, технологични решения за автоматично откриване с помощта на AI, регулаторни мерки и международно сътрудничество.

Друг стратегически значим аспект на развитието на системите с изкуствен интелект е енергоемкостта. Съвременните модели изискват огромни количества електроенергия за обучение и за експлоатация. Докато стандартен

[52.] DARIK BUSINESS REVIEW, 2025. Българската Wiser Technology с два нови проекта за Европейския фонд за отбрана. [Онлайн]. Достъпно на: https://dbr.bg/balgarskata-wiser-technology-s-dva-novi-proekta-za-evropeiskia-fond-za-otbrana.

[53] ЛЕСЕНСКИ, М., 2022. How It Started, How It is Going: Media Literacy Index 2022. София: Институт „Отворено общество“. Достъпно на: https://osis.bg/wp-content/uploads/2022/10/HowItStarted_MediaLiteracyIndex2022_ENG_.pdf [Посетено на 4 юни 2026].

сървърен рак консумира 5 – 10 kW, тези за изкуствен интелект, базирани на GPU, често надхвърлят 50 kW и достигат до 90 kW (Stanchev 2025).

В този контекст България притежава структурно предимство – АЕЦ „Козлодуй“. Ядрената енергия осигурява „базов товар“, наличен 24 часа, критично важен за „гигафабриките“ за данни, където прекъсването на захранването, дори за милисекунди, може да бъде катастрофално. Към 2024 – 2025 г. България е нетен износител на електроенергия с износ от над 7,7 TWh. Плановете за изграждане на 7-и и 8-и блок на АЕЦ „Козлодуй“ гарантират дългосрочна енергийна сигурност за високотехнологичния сектор, превръщайки страната в естествен притегателен център за глобални технологични инвестиции[54]. В сравнение с други държави с амбиции в сферата на технологиите България предлага конкурентно предимство чрез комбинация от стабилност като член на ЕС и НАТО, регулаторна предвидимост, енергийна независимост и по-ниски оперативни разходи (включително 10% корпоративен данък).

Въпреки наличието на хардуер и научни институти, човешкият фактор остава най-сериозното предизвикателство. Българската образователна система реагира на пазарното търсене чрез разкриване на специализирани програми – Софийският университет предлага магистърски програми по „Изкуствен интелект“ и „Технологии за големи данни“, а Техническият университет в София адаптира учебните планове с дисциплини по роботика и киберсигурност. Въпреки това данните за пазара на труда идентифицират структурен дисбаланс. Налице е остър дефицит на специалисти от ниво 2 – експерти с дълбоки познания, способни да разработват и оптимизират сложни AI архитектури, а не просто да използват готови API интерфейси. Проучванията показват, че 60% от ИТ лидерите в България посочват липсата на квалифицирани кадри като основна заплаха за бизнеса[55]. Социологическите данни от 2025 г. разкриват тревожна картина – 56% от гражданите изобщо не използват генеративни инструменти, като българските

[54] ЕЛЕКТРОЕНЕРГИЕН СИСТЕМЕН ОПЕРАТОР ЕАД, 2024. План за развитие на преносната електрическа мрежа на България за периода 2024 – 2033 г. Достъпно на: https://www.eso.bg/fileObj.php?oid=5010.

[55] АЛФА РИСЪРЧ, 2025. Навлизането на изкуствения интелект – между съмнения и предимства. София: Алфа Рисърч. Достъпно на: https://alpharesearch.bg/post/1039-navlizaneto-na-izkustvenia-intelekt-mejdu-sumnenia-i-predimstva.html. [прегледано на 4 юни 2026 г.].

потребители са два пъти по-склонни да използват ИИ за лични цели и забавление (36%), отколкото за професионална продуктивност или образование (18%)[56].

Таблица 29. Обобщени индикатори на AI системата (2025)

Индикатор	*Стойност / Статус*	*Анализ на въздействието*
Инвестиции в AI инфраструктура	~ 90 – 100 млн. евро (BRAIN++)	Осигурява суверенен изчислителен капацитет за наука и отбрана
Енергиен баланс	Нетен износител (7,7 TWh)	Гарантира устойчиво захранване на AI центрове за данни (24/7)
Обществено ползване на AI	13% (активно), 56% (неползващи)	Индикира ниска дигитална готовност на населението/резерва
AI готовност на бизнеса	6,5% (спрямо 13,5% в ЕС)	Забавена икономическа адаптация и риск за конкурентоспособността
Регулаторна рамка	В процес на имплементация (AI Act)	Правна несигурност, изискваща ускорено приемане на норми до август 2025

Към 2025 г. България беше в процес на въвеждане на Европейския акт за изкуствения интелект (EU AI Act) и следваше да се определят компетентни органи (надзорен и нотифициращ), но поради политически сътресения институциите изостават с разработването на вътрешни механизми. Сериозен рисков фактор за дългосрочните проекти е политическата нестабилност. Честите избори и смени на правителства (осем пъти за четири години) затрудняват стратегическото планиране и водят до забавяне на реформите и усвояването на средствата от ЕС, включително по Плана за възстановяване и устойчивост. Въпреки това, ключови проекти като INSAIT и разширението на ядрените мощности се ползват с относителен надпартиен консенсус, което им придава известна устойчивост.

Анализът на състоянието и перспективите на България в областта на изкуствения интелект към 2025 – 2026 г. позволява формулирането на следните изводи:

[56] EUROPEAN COMMISSION, 2025 *Digital Decade Country Report*. Достъпно на: https://digital-strategy.ec.europa.eu/bg/factpages/bulgaria-2025-digital-decade-country-report.

– България е осъществила стратегически пробив в научната инфраструктура. Създаването на INSAIT и спечелването на проекта BRAIN++ не са изолирани успехи, а системна промяна, която поставя страната на картата на глобалните иновации. Наличието на собствени суперкомпютърни мощности за изкуствен интелект е критичен актив за националния суверенитет;

– енергетиката е предимството на българската ИИ стратегия. Комбинацията от ядрена енергия и развита електропреносна мрежа създава уникално конкурентно предимство за привличане на стратегическа инфраструктура за данни;

– военно-цивилната интеграция е в начален етап, но се ускорява. Подписаните меморандуми и участието в DIANA показват, че изолацията между армията и технологичния сектор е преодоляна;

– човешкият капитал и масовата дигитална грамотност остават най-уязвимите звена, създавайки риск от формиране на „елитарна“ ИИ система, откъсната от реалната икономика и общество.

Въз основа на направения анализ може да се предложат следните **препоръки за развитие на ИИ в отбраната** и националната сигурност на Република България:

1. Необходимо е институционализиране на координационен механизъм на ниво Министерски съвет, обединяващ Министерството на отбраната, МВР, ДАНС, Министерството на иновациите и растежа, БАН и представители на частния сектор. Това може да стане чрез създаване на национален съвет за ИИ, който да разработи цялостна национална стратегия за ИИ в отбраната (2026 – 2032), съсредоточена върху: (а) суверенни изчислителни капацитети; (б) развитие на технологии с двойна употреба; (в) киберотбрана и когнитивна сигурност и (г) стандарти за военно приложение на ИИ.

2. Използване на енергийното предимство за създаване на специализирана икономическа зона за AI инфраструктура с фокус върху отбраната, която да комбинира достъп до стабилна ядрена енергия, оптична свързаност и близост до военни тестови центрове. Целта е привличане на инвестиции от глобални технологични гиганти и трансфер на технологии към българската отбранителна индустрия.

3. Необходима е цялостна реформа на военното образование и обучение и внедряване на задължителни дисциплини за запознаване с изкуствения интелект на всички нива. Следва да бъдат създадени програми за преквалификация на офицери от оперативен резерв в специалисти по „човек-машина“ интерфейси и анализ на големи данни. Добре е да бъде установен постоянен

обмен на персонал между Министерство на отбраната и INSAIT/GATE за осигуряване на практически опит в реални проекти.

4. Като отговор на установената висока уязвимост към дезинформация (Legenski 2022) следва да се стартира мащабна междуведомствена програма, интегрираща медийна грамотност в училищните програми от ранна възраст и в програми за продължаващо обучение за възрастни. В контекста на тоталната отбрана това е критично за повишаване на устойчивостта на обществото срещу когнитивни атаки и хибридни заплахи.

5. Създаване на Фонд за иновации в отбраната (ФИО), финансиран от държавния бюджет и европейски средства, за подпомагане на стартиращи компании в областта на ИИ с отбранително приложение; въвеждане на данъчни облекчения за компании, инвестиращи в научноизследователски решения в сферата на автономните системи, киберсигурността и анализа на разузнавателни данни. Задължително условие следва да бъде сътрудничество с академичните центрове и възможност за двойна употреба на разработките.

6. Спешно приемане на подзаконови актове за прилагане на AI Act в отбранителния сектор с ясни процедури за сертифициране на „високорискови“ ИИ системи за военни приложения; кодифициране на етични правила за разработка и използване на автономни оръжейни системи, гарантиращи човешки контрол и съответствие с международното хуманитарно право.

7. Използване на българското председателство на Инициативата „Три морета" (2025 – 2026) за поставяне на дигиталната свързаност и ИИ инфраструктурата като приоритет на регионалното сътрудничество; предлагане на съвместна платформа за наблюдение на дезинформацията и хибридните заплахи в Черноморския регион, базирана на капацитета на GATE, като DIANA центъра.

8. Най-важната препоръка е да се задържат и привличат висококвалифицирани кадри. За целта може да бъдат въведени стипендиантски програми „Отбрана и изкуствен интелект“ за докторанти, срещу задължение за работа в отбранителния сектор или академичните центрове минимум 5 години. Уместно е осигуряването на конкурентни заплати в държавния отбранителен сектор за специалисти, работещи с изкуствен интелект, компенсирани чрез допълнителни социални придобивки и възможности за кариера.

Много изследователи на изкуствения интелект вярват, че той ще доведе до златна ера за човечеството — време, в което машините ще създават неограничени блага и просперитет. Според тях по-мощните интелекти ще вършат цялата работа на хората и човек най-накрая ще бъде свободен да

посвети времето си на почивка и преследване на личните си мечти. Те твърдят, че в такива времена няма да има недостиг на ресурси, а оттам и престъпления, войни и несправедливост, като изкуственият интелект ще ни помогне да решим останалите световни проблеми — от болестите до глобалното затопляне.

ЗАКЛЮЧЕНИЕ

Анализът на трансформациите в съвременната среда на сигурност ясно показва, че изкуственият интелект и дигиталните умения вече не са просто технологична перспектива, а стратегически фактор за устойчивост. Конвергенцията на физическата, дигиталната и когнитивната сфера формира нов тип конфликти, в които данните, алгоритмите и скоростта на вземане на решения имат значение, съпоставимо с класическите измерения на военната мощ. В тази среда националната сигурност се превръща във функция не само на въоръжение и численост, но и на дигитална компетентност и институционална адаптивност.

Емпиричните изводи от настоящото изследване потвърждават, че изкуственият интелект вече е оперативна реалност, която фундаментално променя архитектурата на националната и съюзната отбрана. Бъдещият офицер и експерт по сигурност е призван да изпълнява ролята на архитект и командир на сложни системи човек – машина, в които автономни дигитални агенти, мултимодални информационни потоци и когнитивни операции се сливат в единна оперативна среда. Паралелно, демократизацията на програмирането и навлизането на интуитивни програмни интерфейси превръщат всеки военнослужещ в потенциален иноватор – възможност, която изисква ново поколение дигитална компетентност и образователни рамки.

Монографията утвърждава разбирането, че данните и алгоритмите следва да бъдат третирани като стратегически ресурс, равностоен в територията на класическите бойни средства. Способността за мултимодален синтез и изграждане на платформи за ситуационна осведоменост се очертава като ключов фактор за постигане на информационно и оперативно превъзходство. В същото време, човешкият фактор остава незаменим носител на стратегическа преценка, етична отговорност и легитимност на употребата на сила в пълно съответствие с принципите на НАТО за отговорна и управляема употреба на изкуствения интелект. Изведена е формулата за постигане на стратегическо предимство:

Стратегическо предимство = (алгоритми + данни) x човешка компетентност

Тя подчертава централната роля на образованието и подготовката. Човешката компетентност действа като усилващ множител – без добре

подготвени лидери с критично мислене, дигитална култура и етична устойчивост дори най-напредналите технологии остават стратегически неутрални. В този контекст България разполага с реални и устойчиви предпоставки за трансформация в регионален лидер в дигиталните AI технологии. Научният потенциал, енергийните ресурси и индустриалният капацитет създават солидна основа за развитие, при условие че бъдат синтезирани в дългосрочна, последователна и устойчива държавна стратегия. Ключовото предизвикателство пред управлението е да осигури институционална приемственост и стратегическа визия, които да гарантират, че в ерата на изкуствения интелект България ще бъде активен субект, а не пасивен обект на технологичните промени.

Инвестицията в образованието и дигиталната грамотност на бъдещите лидери в сферата на сигурността и отбраната се очертава като най-високостойностния стратегически избор на XXI век. Чрез системен, холистичен подход и последователна политическа воля Република България има възможност не само да гарантира своя суверенитет и сигурност в ерата на алгоритмичната война, но и да допринесе активно за формирането на отговорна, етична и устойчива архитектура на сигурността в евроатлантическото пространство.

Приложение 1

Бърз тест за самооценка на дигитални умения по рамката DigComp

За да направите самооценка на вашата дигитална компетентностq следва да си зададете следните въпроси, на които да отговорите с „Да“ или „Не“.

1. Информираност и работа с данни.

● Мога ли да използвам разширено търсене, за да намеря само PDF или друг вид файлове?

● Използвам ли портала на държавните служби за вадене на електронни бележки?

● Организирам ли ефективно файловете си?

2. Комуникация и сътрудничество.

● Мога ли сам да организирам онлайн срещи чрез Microsoft Teams или Zoom?

● Споделям ли екрана си по време на среща?

● Мога ли да споделя голям файл чрез облачна услуга (напр. OneDrive, Google Drive)?

● Работя ли съвместно с колеги върху един и същи документ в реално време?

● Спазвам ли етикета в социалните мрежи?

3. Създаване на цифрово съдържание.

● Мога ли да създам уебсайт с помощта на готова платформа?

● Умея ли да редактирам изображения или видеа с Canva, Photoshop или други инструменти?

● Мога ли да създам аватар с AI или интерактивна графика?

● Знам ли какво представляват авторските права и как да използвам съдържание легално в интернет?

● Имам ли основни познания по език за програмиране (напр. HTML, Python)?

4. Сигурност.

● Използвам ли двуфакторна автентикация за защита на профилите си?

● Знам ли как да премахна личните си данни от интернет?

● Разпознавам ли лесно фишинг съобщения?

● Актуализирам ли редовно софтуера на компютъра и телефона?

5. Решаване на проблеми.

● Мога ли самостоятелно да реша проблем с приложение?

● Използвам ли онлайн банкиране или криптовалути?

● Избирам ли правилните инструменти за конкретна задача (напр. презентация, обработка на данни)?

● Участвам ли в онлайн обучения за подобряване на дигиталните си умения?

Ако сте отговорили с „да“ на по-малко от 13 от горните въпроси, е препоръчително да обмислите участие в курс за подобряване на дигиталната си компетентност. Този въпросник е опростен вариант на над 600 индикатора от DigiComp, но чрез следното разпределение по нива на владеене можете лесно да определите вашето текущо ниво:

● Базово ниво: 0 – 30% положителни отговори – имате основни дигитални умения и е добре да развиете още.

● Средно ниво: 31 – 65% положителни отговори – справяте се с повечето основни задачи, но има място за подобрение.

● Напреднало ниво: 66 – 85% положителни отговори – владеете дигиталните умения на високо ниво и сте уверени при работа с различни дигитални инструменти.

● Силно специализирано ниво: над 85% положителни отговори – притежавате експертни и специализирани дигитални умения.

Чрез този подход ще можете по-лесно да определите своето място по отношение на дигиталните умения и да разпознаете кои аспекти изискват развитие. Следващата стъпка е да оцените компетентността си, като включите и уменията за работа с изкуствен интелект (AI) като част от дигиталната си грамотност.

Приложение 2

Въпросник за оценка на уменията за работа с AI в сигурността и отбраната

1. Моля, посочете вашата роля:
A. Военнослужещ
B. Студент по национална сигурност и отбрана
C. Цивилен служител
D. Гражданин, извън системата на отбраната

2. Кои от изброените инструменти сте използвали през последните 30 дни? (Отбележете всички приложими.)
A. Генератори на текст (ChatGPT, Claude, Gemini, Copilot)
B. Генератори на изображения (Midjourney, DALL·E и др.)
C. Инструменти за автоматичен превод (DeepL, Google Translate и др.)
D. Специализиран военен/служебен софтуер с изкуствен интелект
E. Инструменти за автоматизация (n8n, Zapier и др.)
F. Не съм използвал/а AI инструменти през последния месец.

3. За какви конкретни служебни задачи сте използвали AI инструменти през последния месец? (Маркирайте само ако реално сте извършили действието.)
A. Написване или редактиране на официален рапорт/доклад
B. Създаване или структуриране на презентация
C. Създаване на визуално съдържание (снимки, графики, видео аватари)
D. Превод на чуждоезичен текст
E. Анализ на голям обем данни/документи
F. Генериране на програмен код
G. Не съм използвал/а AI за служебни задачи

4. Кое от следните твърдения описва най-точно как работи генеративен AI за текст (напр. ChatGPT)?

A. Моделът търси в база данни с факти и ги комбинира, за да отговори на въпроса

B. Моделът използва предварително програмирани логически правила и отговори, за да разсъждава като човек

C. Моделът предсказва следващата най-вероятна дума или част от дума на базата на статистически алгоритми

D. Моделът се свързва в реално време с интернет, за да копира и предостави информация в отговора

5. Ситуация: Трябва да получите резюме на дълъг некласифициран доклад от ChatGPT. Коя от следните команди (промпт) ще даде най-добър резултат?

A. „Резюмирай това“+ [текст]

B. „Напиши ми нещо за този текст“

C. „Действай като опитен анализатор. Изведи 5-те ключови риска от следния текст под формата на списък с водещи символи (bullets)“ + [текст]

D. „Копирай текста тук“

6. Ако качите PDF файл в публичен чатбот (напр. безплатната версия на ChatGPT) за анализ, какво се случва с данните от файла?

A. Те остават само на моя компютър и са защитени

B. Те се изтриват веднага след затваряне на чата

C. Те могат да бъдат запазени на сървърите на компанията и използвани за обучение на модела (риск от изтичане)

D. Не зная

7. За какво се използва Make.com?

A. Свързва различни приложения (напр. Facebook и Google), за да работят автоматично заедно

B. Използва се за обработка на снимки и видео монтаж

C. Почиства компютъра от вируси и вредни програми

D. Не знам

8. Използвали ли сте специализирани инструменти за създаване на видео аватари със синтезирана реч или друго визуално съдържание с AI?

A. Създавал/а съм видео аватар

B. Създавал/а съм интерактивни визуализации

C. И двете

D. Не

9. През последните шест месеца участвали ли сте в обучение или семинар, свързан с работа с AI инструменти?

A. Да

B. Не

10. Кое е най-голямото предизвикателство според вас за използването на AI в отбраната?

A. Липса на технически умения

B. Липса на подходящ хардуер/софтуер

C. Етични притеснения/недоверие в технологията

D. Регулаторни ограничения/сигурност на информацията

11. Според Вас каква е степента на значимост на ИИ във всяка от следните области за повишаване на отбранителните способности и сигурността? (Оценете по скалата от 1 до 5, където 1 е „минимален принос“, а 5 е „максимален принос“)

Област на приложение	***Минимално***	***Средно***	***Добро***	***Мн. добро***	***Максимално***
Киберсигурност (Защита на мрежи и данни)	1	2	3	4	5
Автономни системи (Дронове, роботи, умни муниции)	1	2	3	4	5
Разузнаване и анализ (Обработка на сателитни снимки/сигнали)	1	2	3	4	5
Логистика и поддръжка (Управление на ресурси и ремонти)	1	2	3	4	5
Вземане на решения (Подпомагане на командирите/C2)	1	2	3	4	5

Бихте ли посочили други специфични области или задачи в системата на отбраната, където внедряването на изкуствен интелект би имало висока добавена стойност? .....

ЛИТЕРАТУРА

ВЪЛЧЕВ, В., ПЕТРОВ, 2022. Дигитална трансформация и киберсигурност в отбранителната система на Република България. Сборник с доклади от Национална конференция по сигурност и отбрана. София: Издателски комплекс – УНСС.

ДИМИТРОВ, Д.; НИКОВСКИ, Д., 1999. *Изкуствен интелект*. Второ преработено издание. София: Издателски комплекс на Технически университет. ISBN 954-438-252-6.

КЛАУЗЕВИЦ, К., 1832/2018. *За войната*. София: Изток-Запад

КОРДОН, А., 2023. *Перспективата изкуствен интелект*. София: Изток-Запад. ISBN 978-619-01-1251-8.

КЪРЦУАЙЛ, Р. 2025. *Сингулярността е по-близо, когато се слеем с изкуствения интелект*. София: Изток- Запад. ISBN 978-619-01-1545-8.

ЛЕВИ, Д.; КУИГЛИ, К.; ЛАМБ, Х. 2024. *Изкуствен интелект: фактите накратко*. София: Книгомания. ISBN 978-619-195-368-4.

СТАНЧЕВ, И., 2025. AI иска ток, вода и земя. България ги има.. *Капитал*. [Онлайн] Достъпно на: https://www.capital.bg/biznes/tehnologii_i_nauka/2025/04/11/4768086_ai_iska_tok_voda_i_zemia_bulgariia_gi_ima/.

ABBATE, J., 1999. *Inventing the Internet*. Cambridge: MIT Press.

ALA-MUTKA, K., 2011. Mapping Digital Competence: Toward a conceptual understanding. DOI: 10.13140/RG.2.2.18046.00322.

ALLEN, G., 2020. *Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security*. Center for a New American Security. Available from: https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy.

ALPAYDIN, E., 2010. *Introduction to Machine Learning*, 2nd ed. Cambridge: MIT Press.

ARKIN, R. C., 2013. *Lethal autonomous systems and the plight of the non-combatant*. Available from: DOI:10.1007/978-3-319-51466-6_15.

ARMBRUST, M.; FOX, A.; GRIFFITH, R. et al, 2010. *A view of cloud computing*. Available from: DOI:10.1145/1721654.1721672.

ASHBY, W.R., 1956. *An Introduction to Cybernetics*. London: Chapman & Hall.

ASIMOV, I. 1950. *I, Robot*. New York: Gnome Press.

BABBAGE, C., 1864. *Passages from the Life of a Philosopher*. London: Longman, Green, Longman, Roberts, & Green.

BARAN, P., 1964. *On distributed communications*. RAND Corporation: Santa Monica, CA. Available from: https://www.rand.org/pubs/research_memoranda/RM3420.html.

BEREITER, C., SCARDAMALIA, M. 1993. *Surpassing Ourselves: An Inquiry into the Nature and Implications of Expertise*. Chicago: Open Court Publishing Company.

BERNERS-LEE, T., 1989. *Information management: A proposal*. CERN: Geneva. Available from: https://www.w3.org/History/1989/proposal.html.

BISHOP, C. M., 2006. *Pattern Recognition and Machine Learning*. Singapore: Springer.

BOOLE, G., 1854 *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*. London: Walton and Maberly.

BOSTROM, N., 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

BOYD, J. R., 1987. *A Discourse on Winning and Losing*. Maxwell AFB Alabama: Air University Press.

BROWN, T. B. et al., 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, pp. 1877 – 1901. Available from: https://doi.org/10.48550/arXiv.2005.14165.

CAMPBELL, M.; HOANE, A. Jr. & HSU, F., 2002. Deep Blue. *Artificial Intelligence*, vol. 134, no 1 – 2, pp. 57 – 83.

CARE, J., 2025. *Leadership Compass: Generative AI Defense*. Available from: https://www.kuppingercole.com/research/lc81042/generative-ai-defense.

CHOMSKY, N., 1957. *Syntactic structures*. The Hague: Mouton & Co.

COLOM, R.; KARAMA, S.; JUNG, R.; HAIER, R., 2010. Human intelligence and brain networks. *Dialogues in Clinical Neuroscience*, vol. 12, no. 4. Available from: DOI:10.31887/DCNS.2010.12.4/rcolom.

COSGROVE, J.; CACHIA R., 2025. *DigComp 3.0: The Digital Competence Framework for Citizens*. Luxembourg: Publications Office of the European Union.

CREVIER, D., 1993. *AI: The tumultuous history of the search for artificial intelligence*. New York: Basic Books.

DARWIN, C., 1859. *On the Origin of Species by Means of Natural Selection*. London: John Murray.

DEEPSEEK AI et al., 2025. *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*. Available from: https://arxiv.org/abs/2501.12948 [Last visited on 1 March 2026].

DUNBAR, R., 1998. The Social Brain Hypothesis. *Evolutionary Anthropology*, vol. 6, no. 5, pp. 178 – 190. Available from: https://doi.org/10.1002/(SICI)1520-6505(1998)6:5%3C178::AID-EVAN5%3E3.0.CO;2-8.

FLACH, P., 2012. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge: Cambridge University Press.

FODOR, J. A., 1975. *The Language of Thought*. New York: Thomas Y. Crowell.

FRENCH, R. M., 2000. The Turing Test: the first 50 years. *Trends in Cognitive Sciences*, vol. 4, no. 3, pp. 115 – 122.

GARDNER, H., 1983. *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books

GARTNER, 2023. *Top strategic technology trends for 2023*. Stamford, CT: Gartner Inc.

GOERTZEL, B.; PENNACHIN, C. (eds.), 2007. *Artificial General Intelligence*. Berlin: Springer.

GOLEMAN, D., 1995. *Emotional Intelligence: Why It Can Matter More Than IQ*. New York: Bantam Books.

GOODFELLOW, I.; BENGIO, Y. & COURVILLE, A., 2016. *Deep Learning*. Cambridge, MA: MIT Press.

HAMMOND, G. T., 2001. *The Mind of War: John Boyd and American Security*. Washington: Smithsonian Books.

HERCULANO-HOUZEL, S., 2009. The Human Brain in Numbers: A Linearly Scaled-up Primate Brain. *Frontiers in Human Neuroscience*, vol. 3, no. 31. Available from: https://doi.org/10.3389/neuro.09.031.2009.

HINTON, G. E.; OSINDERO, S. & TEH, Y. W., 2006. A fast learning algorithm for deep belief nets. *Neural Computation*, vol. 18, no. 7, pp. 1527 – 1554. Available from: DOI: 10.1162/neco.2006.18.7.1527.

HOLLAND, J.H., 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

HORNIK, K.; STINCHCOMBE, M & WHITE, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks*, vol. 2, no. 5, pp. 359 – 366.

KANIA, E., 2019. *Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power*. Washington, DC: Center for a New American Security.

KARPATHY, A., 2025. Vibe coding MenuGen. *karpathy*. Available from: https://karpathy.bearblog.dev/vibe-coding-menugen/

KITZMANN, J.; STRIMOVSKAYA, A. & SEROVA, E. 2024. *Application of Artificial Intelligence Methods for Improvement of Strategic Decision-Making in Logistics*. Available from: DOI:10.1007/978-3-031-50192-0_13.

KNUTH, D. E., 1968. *The Art of Computer Programming, vol. 1: Fundamental Algorithms*. Boston: Addison-Wesley.

KOSTADINOVA, R., 2022. Two Bulgarian institutes will develop big data & AI technologies for NATO. *Trending topics*. [online]. Available from: https://www.trendingtopics.eu/two-bulgarian-institutes-will-develop-big-data-ai-technologies-for-nato/.

KOSTKA, G., 2021. *China's AI Strategy and Military Ambitions*. Berlin: Mercator Institute for China Studies (MERICS).

KRIZHEVSKY, A.; SUTSKEVER, I. & HINTON, G., 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097 – 1105

KURZWEIL, R., 2005. *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking Press.

LECUN, Y.; BENGIO, Y. & HINTON, G., 2015. Deep Learning. *Nature*, vol. 521 (7553), pp. 436 – 444.

LEE, K-F., 2024. *AI Superpowers: China, Silicon Valley, and the New World Order*. Updated edition. Boston: Houghton Mifflin Harcourt.

LINDSAY, R. K. LEDERBERG, J.; FEIGENBAUM, E. A.; BUCHANAN, B. G., 1980. *Applications of artificial intelligence for organic chemistry: The DENDRAL project*. New York: McGraw-Hill.

MARCUS, G.; DAVIS, E., 2019. *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon Books.

MCCARTHY, J. et al., 2006. A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine*, vol. 27, no. 4, pp. 12 – 14.

MCCULLOCH, W. S. AND PITTS, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, vol. 5, pp. 115 – 133.

MCCULLOCH, W. S.; PITTS, W., 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115 – 133.

MCDERMOTT, J., 1982. *R1:* A rule-based configurer of computer systems. *Artificial Intelligence*, vol. 19, no. 1, pp. 39 – 88.

MINSKY, M. & PAPERT, S., 1969. *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.

MINSKY, M., 1967. *Computation: Finite and Infinite Machines*. Englewood Cliffs: NJ Prentice-Hall.

MITCHELL, T. M., 1997. *Machine Learning*. New York: McGraw-Hill.

MOCKAPETRIS, P., 1983. *Domain names – concepts and facilities*. RFC 882. Available from: https://tools.ietf.org/html/rfc882.

MORAVEC, H. P., 1988. *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press.

MORGAN, F. E.; BOUDREAUX, B.; LOHN, A. J.; ASHBY, M.; CURRIDEN, C.; KLIMA, K.; GROSSMAN, D., 2020. Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World. *Rand*. [online]. Available from: https://www.rand.org/pubs/research_reports/RR3139-1.html.

MURGIA, M.; CLOVER, C.; JUDAH, J.; KILLING, A., 2026. The AI-driven kill chain. *Technology. Targeted attacs*. [March 12, 2026]. Available from: https://www.scribd.com/document/1018983188/AI-Driven-Kill-Chain-Transformation-FT-14-March-2026-p-2.

NEUMANN, J., 1945. *First draft of a report on the EDVAC*. University of Pennsylvania, Philadelphia: Moore School of Electrical Engineering.

NEWELL, A.; SIMON, K. A., 1963. GPS, a program that simulates human thought. McGraw-Hill: New York. In: Feigenbaum, E. A. Feldman, J., (eds.). *Computers and thought*, pp. 279 – 293.

NILSSON, N. J., 1980. *Principles of Artificial Intelligence*. Palo Alto, CA: Tioga Publishing Company.

O'REILLY, T., 2005. What Is Web 2.0. [online] *O'Reilly*. Available from: https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.

PATTERSON, D.; GONZALEZ, J.; LE, Q.; LIANG, C.; MUNGUIA, L. M.; ROTHCHILD, D.; SO, D.; TEXIER, M.; DEAN, J., 2021. Carbon emissions and large neural network training. Available from: arXiv preprint arXiv:2104.10350.

PAYNE, K., 2018. Artificial Intelligence: A Revolution in Strategic Affairs?. *Survival*, vol. 60; no. 5, pp. 7 – 32. Available from: DOI:10.1080/00396338.2018.1518374.

PUTNAM, H., 1960. Minds and machines. In: Hook, S. (ed.) *Dimensions of Mind*, pp. 138 – 164. New York: New York University Press.

RADFORD, A.; NARASIMHAN, K.; SALIMANS, T.; SUTSKEVER, I., 2018. *Improving language understanding by generative pre-training*. OpenAI Technical Report.

REYNOLDS, L.; MCDONELL, K., 2021. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1 – 7.

ROSENBLATT, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, vol. 65, no. 6, pp. 386 – 408.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J., 1986. Learning representations by back-propagating errors. *Nature*, vol. 323 (6088), pp. 533 – 536.

RUSSELL AND NORVIG, 2021. *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson.

RUSSELL, S., 2020. *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson.

SAMUEL, A. L., 1959. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210 – 229.

SCHWARZ, E., 2024. *The Ethical Implications of AI in Warfare*. Queen Mary University of London: Centre for War Studies.

SHANNON, C. E., 1949. *The mathematical theory of communication*. Urbana: University of Illinois Press .

SHETTLEWORTH, S. J., 2010. *Cognition, Evolution, and Behavior*, 2nd ed. Oxford University Press.

SHORTLIFFE, E. H., 1976. *Computer-based medical consultations: MYCIN*. New York: Elsevier.

SILVER, D. et al., 2016. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*, vol. 529 (7587), pp. 484 – 489.

SIMON, 1969. *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

SIMON, H. A., 1965. *The shape of automation for men and management*. New York: Harper & Row.

SIMON, H.; NEWELL, A., 1958. Heuristic problem solving: The next advance in operations research. *Operations Research*, vol. 6, no. 1, pp. 1 – 10.

SINGLA, A.; SUKHAREVSKY, A.; YEE, L.; CHUI, M.; HALL, B., 2024. *The State of AI in 2024:Gen AI adoption spikes and starts to generate value*. Available from: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024.

STANDAGE, T., 2002. *The Turk: The life and times of the famous eighteenth-century chess-playing machine*. New York: Walker & Company.

STERNBERG, R. J., 1985. *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge University Press.

SUTTON, R. S.; BARTO, A., 2018. *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA: MIT Press.

TRIPP & BRENNER, 2024. Digital Literacy: Vital Tools for the Great Power Competition). *Airman Magazine*.

TURING, A. M., 1936. On computable numbers, with an application to the Entscheidungsproblem. P*roceedings of the London Mathematical Society*, vol. 42, no. 2, pp. 230 – 265.

TURING, A. M., 1950. Computing machinery and intelligence. *Mind*, vol. 59, no. 236, pp. 433 – 460.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I., 2017. Attention Is All You Need. *Advances in Neural Information Processing Systems*. Available from: https://arxiv.org/abs/1706.03762. [Last visited: 4 January 2026].

VILLALOBOS, P.; HO, A.; SEVILLA, J.; BESIROGLU, T.; HEIM, L.; HOBBHAHN, M., 2022. *Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning*. Available from: arXiv preprint arXiv:2211.04325.

VON NEUMANN, J.; MORGENSTERN, O. 1944. *Theory of games and economic behavior*. Princeton: Princeton University Press:.

VUORIKARI, R.; KLUZER, S. & PUNIE, Y., 2022. *DigComp 2.2: The Digital Competence Framework for Citizens: With new examples of knowledge, skills and attitudes*. Publications Office of the European Union. https://publications.jrc.ec.europa.eu/repository/handle/JRC128415.

WEI, J.; WANG, X.; SCHUURMANS, D.; BOSMA, M.; ICHTER, B.; XIA, F.; CHI, ED; LE, Q;. ZHOU D., 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems*, vol. 35, pp. 24824 – 24837.

WHITE, J.; FU, Q.; HAYS, S.; SANDBORN, M.; OLEA, C.; GILBERT, H.; ELNASHAR, A.; SPENCER-SMITH, J.; SCHMIDT, D. C., 2023. *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. Available from: arXiv preprint arXiv:2302.11382.

WIDROW, B.; HOFF, M., 1960. Adaptive switching circuits. *Western Electronic Show and Convention, Institute of Radio Engineers*, vol. 4, pp. 96 – 104.

WIENER, N., 1948. *Cybernetics: or control and communication in the animal and the machine*. New York: John Wiley & Sons.

WOOLLEY, S. C., & HOWARD, P., 2018. *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford University Press

YAO, S.; YU, D.; ZHAO, J.; SHAFRAN, I.; GRIFFITHS, T. L.; CAO, Y.; NARASIMHAN, K., 2023. Tree of Thoughts: Deliberate Problem Solving with Large Language Models. *Advances in Neural Information Processing Systems*, vol. 36.

YOON JU-HE, 2026. AI Identifies 1,000 Targets Hourly in Iran War. *The Chosun Daily*. Available from: https://www.chosun.com/english/world-en/2026/03/09/VJLPZMG4ARAVNF35N26YRA4EWM/.

ZADEH, L. A., 1965. Fuzzy Sets. *Information and Control*, vol. 8, no. 3, pp. 338 – 353.

REFERENCES

ABBATE, J., 1999. *Inventing the Internet*. Cambridge: MIT Press.

ALA-MUTKA, K., 2011. Mapping Digital Competence: Toward a conceptual understanding. DOI: 10.13140/RG.2.2.18046.00322.

ALLEN, G., 2020. *Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security*. Center for a New American Security. Available from: https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy.

ALPAYDIN, E., 2010. *Introduction to Machine Learning*, 2nd ed. Cambridge: MIT Press.

ARKIN, R. C., 2013. *Lethal autonomous systems and the plight of the non-combatant*. Available from: DOI:10.1007/978-3-319-51466-6_15.

ARMBRUST, M.; FOX, A.; GRIFFITH, R. et al, 2010. *A view of cloud computing*. Available from: DOI:10.1145/1721654.1721672.

ASHBY, W.R., 1956. *An Introduction to Cybernetics*. London: Chapman & Hall.

ASIMOV, I. 1950. *I, Robot*. New York: Gnome Press.

BABBAGE, C., 1864. *Passages from the Life of a Philosopher*. London: Longman, Green, Longman, Roberts, & Green.

BARAN, P., 1964. *On distributed communications*. RAND Corporation: Santa Monica, CA. Available from: https://www.rand.org/pubs/research_memoranda/RM3420.html.

BEREITER, C., SCARDAMALIA, M. 1993. *Surpassing Ourselves: An Inquiry into the Nature and Implications of Expertise*. Chicago: Open Court Publishing Company.

BERNERS-LEE, T., 1989. *Information management: A proposal*. CERN: Geneva. Available from: https://www.w3.org/History/1989/proposal.html.

BISHOP, C. M., 2006. *Pattern Recognition and Machine Learning*. Singapore: Springer.

BOOLE, G., 1854 *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*. London: Walton and Maberly.

BOSTROM, N., 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

BOYD, J. R., 1987. *A Discourse on Winning and Losing*. Maxwell AFB Alabama: Air University Press.

BROWN, T. B. et al., 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, pp. 1877 – 1901. Available from: https://doi.org/10.48550/arXiv.2005.14165.

CAMPBELL, M.; HOANE, A. Jr. & HSU, F., 2002. Deep Blue. *Artificial Intelligence*, vol. 134, no 1 – 2, pp. 57 – 83.

CARE, J., 2025. *Leadership Compass: Generative AI Defense*. Available from: https://www.kuppingercole.com/research/lc81042/generative-ai-defense.

CHOMSKY, N., 1957. *Syntactic structures*. The Hague: Mouton & Co.

COLOM, R.; KARAMA, S.; JUNG, R.; HAIER, R., 2010. Human intelligence and brain networks. *Dialogues in Clinical Neuroscience*, vol. 12, no. 4. Available from: DOI:10.31887/DCNS.2010.12.4/rcolom.

COSGROVE, J.; CACHIA R., 2025. *DigComp 3.0: The Digital Competence Framework for Citizens*. Luxembourg: Publications Office of the European Union.

CREVIER, D., 1993. *AI: The tumultuous history of the search for artificial intelligence*. New York: Basic Books.

DARWIN, C., 1859. *On the Origin of Species by Means of Natural Selection*. London: John Murray.

DEEPSEEK AI et al., 2025. *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*. Available from: https://arxiv.org/abs/2501.12948 [Last visited on 1 March 2026].

DIMITROV, D.; NIKOVSKI, D., 1999. Izkustven intelekt. 2nd edition. Sofia: Izdatelski kompleks na Tehnicheski universitet. ISBN 954-438-252-6. [In Bulgarian].

DUNBAR, R., 1998. The Social Brain Hypothesis. *Evolutionary Anthropology*, vol. 6, no. 5, pp. 178 – 190. Available from: https://doi.org/10.1002/(SICI)1520-6505(1998)6:5%3C178::AID-EVAN5%3E3.0.CO;2-8.

FLACH, P., 2012. *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*. Cambridge: Cambridge University Press.

FODOR, J. A., 1975. *The Language of Thought*. New York: Thomas Y. Crowell.

FRENCH, R. M., 2000. The Turing Test: the first 50 years. *Trends in Cognitive Sciences*, vol. 4, no. 3, pp. 115 – 122.

GARDNER, H., 1983. *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books

GARTNER, 2023. *Top strategic technology trends for 2023*. Stamford, CT: Gartner Inc.

GOERTZEL, B.; PENNACHIN, C. (eds.), 2007. *Artificial General Intelligence*. Berlin: Springer.

GOLEMAN, D., 1995. *Emotional Intelligence: Why It Can Matter More Than IQ*. New York: Bantam Books.

GOODFELLOW, I.; BENGIO, Y. & COURVILLE, A., 2016. *Deep Learning*. Cambridge, MA: MIT Press.

HAMMOND, G. T., 2001. *The Mind of War: John Boyd and American Security*. Washington: Smithsonian Books.

HERCULANO-HOUZEL, S., 2009. The Human Brain in Numbers: A Linearly Scaled-up Primate Brain. *Frontiers in Human Neuroscience*, vol. 3, no. 31. Available from: https://doi.org/10.3389/neuro.09.031.2009.

HINTON, G. E.; OSINDERO, S. & TEH, Y. W., 2006. A fast learning algorithm for deep belief nets. *Neural Computation*, vol. 18, no. 7, pp. 1527 – 1554. Available from: DOI: 10.1162/neco.2006.18.7.1527.

HOLLAND, J.H., 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

HORNIK, K.; STINCHCOMBE, M & WHITE, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks*, vol. 2, no. 5, pp. 359 – 366.

KANIA, E., 2019. *Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power*. Washington, DC: Center for a New American Security.

KARPATHY, A., 2025. Vibe coding MenuGen. *karpathy*. Available from: https://karpathy.bearblog.dev/vibe-coding-menugen/

KARTSUAYL, R. 2025. Singulyarnostta e po-blizo, kogato se sleem s izkustvenia intelekt. Sofia: Iztok- Zapad. ISBN 978-619-01-1545-8. [In Bulgarian].

KITZMANN, J.; STRIMOVSKAYA, A. & SEROVA, E. 2024. *Application of Artificial Intelligence Methods for Improvement of Strategic Decision-Making in Logistics*. Available from: DOI:10.1007/978-3-031-50192-0_13.

KLAUZEVITS, K., 1832/2018. Za voynata. Sofia: Iztok-Zapad. [In Bulgarian].

KNUTH, D. E., 1968. *The Art of Computer Programming, vol. 1: Fundamental Algorithms*. Boston: Addison-Wesley.

KORDON, A., 2023. Perspektivata izkustven intelekt. Sofia: Iztok-Zapad. ISBN 978-619-01-1251-8. [In Bulgarian].

KOSTADINOVA, R., 2022. Two Bulgarian institutes will develop big data & AI technologies for NATO. *Trending topics*. [online]. Available from: https://www.trendingtopics.eu/two-bulgarian-institutes-will-develop-big-data-ai-technologies-for-nato/.

KOSTKA, G., 2021. *China's AI Strategy and Military Ambitions*. Berlin: Mercator Institute for China Studies (MERICS).

KRIZHEVSKY, A.; SUTSKEVER, I. & HINTON, G., 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097 – 1105

KURZWEIL, R., 2005. *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking Press.

LECUN, Y.; BENGIO, Y. & HINTON, G., 2015. Deep Learning. *Nature*, vol. 521 (7553), pp. 436 – 444.

LEE, K-F., 2024. *AI Superpowers: China, Silicon Valley, and the New World Order*. Updated edition. Boston: Houghton Mifflin Harcourt.

LEVI, D.; KUIGLI, K.; LAMB, H. 2024. Izkustven intelekt: faktite nakratko. Sofia: Knigomania. ISBN 978-619-195-368-4. [In Bulgarian].

LINDSAY, R. K. LEDERBERG, J.; FEIGENBAUM, E. A.; BUCHANAN, B. G., 1980. *Applications of artificial intelligence for organic chemistry: The DENDRAL project*. New York: McGraw-Hill.

MARCUS, G.; DAVIS, E., 2019. *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon Books.

MCCARTHY, J. et al., 2006. A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine*, vol. 27, no. 4, pp. 12 – 14.

MCCULLOCH, W. S. AND PITTS, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, vol. 5, pp. 115 – 133.

MCCULLOCH, W. S.; PITTS, W., 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115 – 133.

MCDERMOTT, J., 1982. *R1:* A rule-based configurer of computer systems. *Artificial Intelligence*, vol. 19, no. 1, pp. 39 – 88.

MINSKY, M. & PAPERT, S., 1969. *Perceptrons: An introduction to computational geometry*. Cambridge, MA: MIT Press.

MINSKY, M., 1967. *Computation: Finite and Infinite Machines*. Englewood Cliffs: NJ Prentice-Hall.

MITCHELL, T. M., 1997. *Machine Learning*. New York: McGraw-Hill.

MOCKAPETRIS, P., 1983. *Domain names – concepts and facilities*. RFC 882. Available from: https://tools.ietf.org/html/rfc882.

MORAVEC, H. P., 1988. *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press.

MORGAN, F. E.; BOUDREAUX, B.; LOHN, A. J.; ASHBY, M.; CURRIDEN, C.; KLIMA, K.; GROSSMAN, D., 2020. Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World. *Rand*. [online]. Available from: https://www.rand.org/pubs/research_reports/RR3139-1.html.

MURGIA, M.; CLOVER, C.; JUDAH, J.; KILLING, A., 2026. The AI-driven kill chain. *Technology. Targeted attacs*. [March 12, 2026]. Available from: https://www.scribd.com/document/1018983188/AI-Driven-Kill-Chain-Transformation-FT-14-March-2026-p-2.

NEUMANN, J., 1945. *First draft of a report on the EDVAC*. University of Pennsylvania, Philadelphia: Moore School of Electrical Engineering.

NEWELL, A.; SIMON, K. A., 1963. GPS, a program that simulates human thought. McGraw-Hill: New York. In: Feigenbaum, E. A. Feldman, J., (eds.). *Computers and thought*, pp. 279 – 293.

NILSSON, N. J., 1980. *Principles of Artificial Intelligence*. Palo Alto, CA: Tioga Publishing Company.

O'REILLY, T., 2005. What Is Web 2.0. [online] *O'Reilly*. Available from: https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html.

PATTERSON, D.; GONZALEZ, J.; LE, Q.; LIANG, C.; MUNGUIA, L. M.; ROTHCHILD, D.; SO, D.; TEXIER, M.; DEAN, J., 2021. Carbon emissions and large neural network training. Available from: arXiv preprint arXiv:2104.10350.

PAYNE, K., 2018. Artificial Intelligence: A Revolution in Strategic Affairs?. *Survival*, vol. 60; no. 5, pp. 7 – 32. Available from: DOI:10.1080/00396338.2018.1518374.

PUTNAM, H., 1960. Minds and machines. In: Hook, S. (ed.) *Dimensions of Mind*, pp. 138 – 164. New York: New York University Press.

RADFORD, A.; NARASIMHAN, K.; SALIMANS, T.; SUTSKEVER, I., 2018. *Improving language understanding by generative pre-training*. OpenAI Technical Report.

REYNOLDS, L.; MCDONELL, K., 2021. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1 – 7.

ROSENBLATT, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, vol. 65, no. 6, pp. 386 – 408.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J., 1986. Learning representations by back-propagating errors. *Nature*, vol. 323 (6088), pp. 533 – 536.

RUSSELL AND NORVIG, 2021. *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson.

RUSSELL, S., 2020. *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson.

SAMUEL, A. L., 1959. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210 – 229.

SCHWARZ, E., 2024. *The Ethical Implications of AI in Warfare*. Queen Mary University of London: Centre for War Studies.

SHANNON, C. E., 1949. *The mathematical theory of communication*. Urbana: University of Illinois Press .

SHETTLEWORTH, S. J., 2010. *Cognition, Evolution, and Behavior*, 2nd ed. Oxford University Press.

SHORTLIFFE, E. H., 1976. *Computer-based medical consultations: MYCIN*. New York: Elsevier.

SILVER, D. et al., 2016. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*, vol. 529 (7587), pp. 484 – 489.

SIMON, 1969. *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

SIMON, H. A., 1965. *The shape of automation for men and management*. New York: Harper & Row.

SIMON, H.; NEWELL, A., 1958. Heuristic problem solving: The next advance in operations research. *Operations Research*, vol. 6, no. 1, pp. 1 – 10.

SINGLA, A.; SUKHAREVSKY, A.; YEE, L.; CHUI, M.; HALL, B., 2024. *The State of AI in 2024:Gen AI adoption spikes and starts to generate value*. Available from: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024.

STANCHEV, I., 2025. AI iska tok, voda i zemya. Bulgaria gi ima. *Kapital*. [Onlayn] Dostapno na: https://www.capital.bg/biznes/tehnologii_i_nauka/2025/04/11/4768086_ai_iska_tok_voda_i_zemia_bulgariia_gi_ima/. [In Bulgarian].

STANDAGE, T., 2002. *The Turk: The life and times of the famous eighteenth-century chess-playing machine*. New York: Walker & Company.

STERNBERG, R. J., 1985. *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge University Press.

SUTTON, R. S.; BARTO, A., 2018. *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA: MIT Press.

TRIPP & BRENNER, 2024. Digital Literacy: Vital Tools for the Great Power Competition). *Airman Magazine*.

TURING, A. M., 1936. On computable numbers, with an application to the Entscheidungsproblem. P*roceedings of the London Mathematical Society*, vol. 42, no. 2, pp. 230 – 265.

TURING, A. M., 1950. Computing machinery and intelligence. *Mind*, vol. 59, no. 236, pp. 433 – 460.

VALCHEV, V., PETROV, 2022. Digitalna transformatsia i kibersi-gurnost v otbranitelnata sistema na Republika Bulgaria. Sbornik s dok-ladi ot Natsionalna konferentsia po sigurnost i otbrana. Sofia: Izda-telski kompleks – UNSS. [In Bulgarian].

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I., 2017. Attention Is All You Need. *Advances in Neural Information Processing Systems*. Available from: https://arxiv.org/abs/1706.03762. [Last visited: 4 January 2026].

VILLALOBOS, P.; HO, A.; SEVILLA, J.; BESIROGLU, T.; HEIM, L.; HOBBHAHN, M., 2022. *Will We Run Out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning*. Available from: arXiv preprint arXiv:2211.04325.

VON NEUMANN, J.; MORGENSTERN, O. 1944. *Theory of games and economic behavior*. Princeton: Princeton University Press:.

VUORIKARI, R.; KLUZER, S. & PUNIE, Y., 2022. *DigComp 2.2: The Digital Competence Framework for Citizens: With new examples of knowledge, skills and attitudes*. Publications Office of the European Union. https://publications.jrc.ec.europa.eu/repository/handle/JRC128415.

WEI, J.; WANG, X.; SCHUURMANS, D.; BOSMA, M.; ICHTER, B.; XIA, F.; CHI, ED; LE, Q;. ZHOU D., 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems*, vol. 35, pp. 24824 – 24837.

WHITE, J.; FU, Q.; HAYS, S.; SANDBORN, M.; OLEA, C.; GILBERT, H.; ELNASHAR, A.; SPENCER-SMITH, J.; SCHMIDT, D. C., 2023. *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. Available from: arXiv preprint arXiv:2302.11382.

WIDROW, B.; HOFF, M., 1960. Adaptive switching circuits. *Western Electronic Show and Convention, Institute of Radio Engineers*, vol. 4, pp. 96 – 104.

WIENER, N., 1948. *Cybernetics: or control and communication in the animal and the machine*. New York: John Wiley & Sons.

WOOLLEY, S. C., & HOWARD, P., 2018. *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford University Press

YAO, S.; YU, D.; ZHAO, J.; SHAFRAN, I.; GRIFFITHS, T. L.; CAO, Y.; NARASIMHAN, K., 2023. Tree of Thoughts: Deliberate Problem Solving with Large Language Models. *Advances in Neural Information Processing Systems*, vol. 36.

YOON JU-HE, 2026. AI Identifies 1,000 Targets Hourly in Iran War. *The Chosun Daily*. Available from: https://www.chosun.com/english/world-en/2026/03/09/VJLPZMG4ARAVNF35N26YRA4EWM/.

ZADEH, L. A., 1965. Fuzzy Sets. *Information and Control*, vol. 8, no. 3, pp. 338 – 353.

Допълнителна литература

РИКАРДС, Д., 2025. *Пари GPT. Изкуственият интелект и заплахата за световната икономика*. София: Изток-Запад. ISBN 978-619-01-1580-9.

ГЛАВЧЕВ, А. 2025. AI с цел сигурност: Министерството на отбраната и INSAIT обявиха сътрудничество. [Онлайн]. Digitalk. Достъпно от: https://digitalk.bg/security/2025/01/07/4726131_ai_s_cel_sigurnost_ministerstvoto_na_otbranata_i/.

BRIN, S., 1999. PageRank: Bringing Order to the Web. *Stanford InfoLab*. Available from: http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf.

HOROWITZ, M. C.; SCHARRE, P.; ALLEN, G. C.; FREDERICK, K.; CHO A. and SARAVALLE, E., 2018. *Artificial intelligence and international security*. Washington, DC: Center for a New American Security.

HU, K., 2023. ChatGPT sets record for fastest-growing user base – analyst. *Reuters*. [online]. Available from: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/. [Last visited: 15 February 2026].

IMPERVA, 2023. *Bad Bot Report: Ecosystem of Automation*. Достъпно на: https://www.imperva.com/resources/reports/2023-Imperva-Bad-Bot-Report.pdf

MOORE, G. E., 1965. *Cramming more components onto integrated circuits*. Достъпно на: https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf.

PWC, 2017. *Sizing the prize: PwC's Global AI Study – Exploiting the AI Revolution.* Available from: https://www.pwc.ch/en/publications/2017/pwc_global_ai_study_2017_en.pdf.

Петко Стефанов Димов

ИЗКУСТВЕН ИНТЕЛЕКТ
И ДИГИТАЛНИ УМЕНИЯ В ОТБРАНАТА
Българска
Първо издание

Катерина Антонова, *редактор*
Георги Кирилов, *предпечатна подготовка*

формат 70x100/16

Военна академия „Георги Стойков Раковски“
София, бул. Евлоги и Христо Георгиеви, № 82

https://rndc.bg/publishing

ISBN: 978-619-7711-85-1

266