Софийски университет „Св. Климент Охридски“
Факултет по математика и информатика

РАЗРАБОТКА

по

Отговорен изкуствен интелект: етични, правни, социални и икономически аспекти

спец. Технологии за големи данни,
I курс, зимен семестър, академ. година 2024/2025

Предизвикателства и ограничения при системи с изкуствен интелект с отворен код

Изготвил:
Стефан Димитров Велев,
Ф. Н.: 0MI3400521

Преподавател:
Александра Цветкова

11.01.2025
гр. София

СЪДЪРЖАНИЕ

1. Увод

В последните години **изкуственият интелект (ИИ)** набира все по-голяма популярност. Науката, която изследва начините за това как да накараме компютрите да правят неща, които за хората изглеждат разумни, навлиза постепенно във всяка една сфера на живота.

В съчетание с многобройните възможности, възникващи от използването на ИИ, **отвореният код** идва с потенциал за тласък на повече иновации както в публичния, така и в частния сектор. Предимствата включват способността за увеличаване на прозрачността, улесняване на одита на ИИ и по този начин повишаване на доверието на гражданите, като същевременно се стимулират икономическите дейности и специфичното за дадена област познание. Сред основните ограничения в сферата се включват правни предизвикателства, технически предизвикателства, ограничения по отношение на управление на данните, управление на риска, обществени и етични предизвикателства.

Целта на настоящата разработка е да обхване въпросите, които системите с ИИ с отворен код поставят. Това ще бъде направено през призмата на различните категории предизвикателства, които са налице в областта – правни, технически, обществени, етични и др. Ще бъдат разгледани възможностите, но и недостатъците, които могат да бъдат в резултат от използването на такива системи. В изложението ще бъдат представени примери, които да онагледят теорията по темата.

2. Понятие за софтуер с отворен код

Софтуерът с отворен код е компютърен софтуер, чийто изходен код е достъпен под лиценз (или договореност като публично достояние), който позволява на потребителите да изучават, променят и подобряват софтуера и да го разпространяват отново в модифицирана или немодифицирана форма. Често се разработва публично, в сътрудничество. Предоставянето на достъп до изходния код не е само по себе си достатъчно условие, за да може един софтуер да се счита за такъв с отворен код.

Инициативата за отворен код (***Open Source Initiative***) е световна организация с нестопанска цел, основана през 1998 г. и действаща като водещ авторитет при разработката и разпространението на софтуер с отворен код. Според **дефиницията за отворен код** (***Open Source Definition – ver. 1.9***), която тя предлага, софтуерът трябва да отговаря на **10 критерия**, които е необходимо да бъдат налични, за да може да се нарече с отворен код:

1. Свободно разпространение: *Лицензът не трябва да ограничава никоя страна да продава или раздава софтуера като компонент от обобщено разпространение на софтуер, съдържащо програми от няколко различни източника. Лицензът не трябва да изисква възнаграждение или друга такса за такава продажба.*

2. Достъпност и интегритет на изходния код: *Програмата трябва да включва изходен код и да позволява разпространение в изходен код, както и в компилиран вид.*

3. Допускане на модификации: *Лицензът трябва да позволява модификации и производни произведения, както и тяхното разпространение при същите условия като лиценза на оригиналния софтуер.*

4. Интегритет на изходния код на автора: *Лицензът може да изисква оригиналният софтуер да се разпространява непокътнат, но само ако модификациите могат да се разпространяват като отделни файлове (patch files) без ограничения.*

5. Без дискриминация срещу отделни хора или групи

6. Без дискриминация по области и сфери на приложение: *Лицензът не трябва да ограничава никого от използването на програмата в каквато и да е област на дейност.*

7. Разпространение на лицензa: *Правата, свързани с програмата и нейния изходен код, трябва да са приложими за всички следващи разпространения, без да е необходимо прилагането на други лицензи от тези страни.*

8. Лицензът не трябва да бъде специфичен за даден продукт: *Правата, свързани с разпространението на програмата, не трябва да зависят от това дали тя е част от определена софтуерна дистрибуция.*

9. Лицензът не трябва да ограничава друг софтуер: *Лицензът не трябва да поставя ограничения на друг софтуер, който се разпространява заедно с лицензирания софтуер (например да бъде с отворен код).*

10. Лицензът трябва да бъде технологично неутрален: *Никоя разпоредба на лицензa не може да се основава на конкретна технология или стил интерфейс.*

Софтуерът с отворен код се предлага под специфичен тип лицензи за авторски права. В съчетание с многобройните възможности, произтичащи от използването на ИИ, напредъкът може да бъде безпрецедентен в обхват и дълбочина, в редица различни сектори, за автоматизиране на процеси, изграждане на прозрачност и иновации в услугите. **Подходът с отворен код** идва с голям потенциал за иновационен капацитет и усвояване, както в публичния, така и в частния сектор, благодарение на способността на лица и организации свободно да използват повторно софтуера под **лицензи с отворен код**.

Отвореният изходен код обикновено се съхранява в **хранилище** и се споделя публично. Всеки може да получи достъп до **хранилището**, за да използва кода независимо или да допринесе за подобрения в дизайна и функционалността на цялостния проект.

Софтуерът с отворен код обикновено идва с лиценз за разпространение. Този лиценз включва условия, които определят как разработчиците могат да използват, изучават, модифицират и най-важното да разпространяват софтуера. Сред най-популярните лицензи за отворен код са:

- *MIT License*
- *GNU General Public License (GPL) 2.0*
- *Apache License 2.0*
- *GNU General Public License (GPL) 3.0*
- *BSD License 2.0*

Когато изходният код се промени, софтуерът с отворен код трябва да предоставя информация за това какво е променено, както и включените методи. В зависимост от **лицензионните условия** софтуерът, получен в резултат на тези модификации, може или не може да се изисква да бъде предоставен безплатно.

3. Системи с изкуствен интелект с отворен код

3.1 Дефиниция

Следвайки същите идеи зад понятието „софтуер с отворен код“, **Инициативата за отворен код** (***Open Source Initiative***) определя **системите с ИИ с отворен код** като системи с ИИ, които са свободно достъпни за:

- **Използване** на системата за всякакви цели и без да е необходимо да се изисква разрешение;
- **Изучаване** как системите работят и **разбиране** как са получени резултатите от тях;
- **Модифициране** на системите за всякаква цел, включително за промяна на изхода им;
- **Споделяне** на системите с други лица, които да ги използват с или без модификации за всякаква цел.

Тези свободи се прилагат както за напълно функционална система, така и за отделни елементи на система. **Предварително условие** за упражняване на тези свободи е наличието на достъп до предпочитана форма за извършване на модификации на системата. За системите с наличен модул за **машинно самообучение**, това включва следните елементи:

- **Информация за данните:** Достатъчно подробна информация за данните, използвани за обучение на системата, така че човек с подходящите умения да може да изгради по същество еквивалентна система. Информацията за данните се предоставя при условия, одобрени от OSI. По-специално, това трябва да включва:
  - ***пълно описание на всички данни***, използвани за обучение, включително (ако се използват) на неподлежащи на споделяне данни, разкриване на произхода на данните, техния обхват и характеристики, как данните са получени и избрани, процедури за етикетиране и методологии за обработка и филтриране на данни;
  - ***списък*** на всички ***публично достъпни данни*** за обучение и откъде могат да бъдат получени;
  - ***списък*** на всички ***данни за обучение***, които могат да бъдат получени от ***трети страни*** и къде да бъдат получени, включително срещу заплащане.
- **Кодът:** Пълният изходен код, използван за обучение и работа на системата. Кодът представлява пълната спецификация на това как данните са били обработени и филтрирани и как е извършено обучението. Кодът трябва да бъде предоставен под лицензи, одобрени от OSI:
  - *например, ако се използва, това трябва да включва код, използван за обработка и филтриране на данни, код, използван за обучение, включително използвани аргументи и настройки, валидиране и тестване, поддържащи библиотеки като токенизатори и код за търсене на хиперпараметри, код за изводи и архитектура на модела.*
- **Параметрите:** Параметрите на модела, като ***тегла*** или други ***конфигурационни настройки***. Параметрите се предоставят при условия, одобрени от OSI:
  - *например, това може да включва контролни точки от ключови междинни етапи на обучение, както и крайното състояние на оптимизатора.*

Лицензионните или други условия, прилагани към тези елементи и към всяка комбинация от тях, може да съдържат условия, които изискват всяка модифицирана версия да бъде пусната при същите условия като оригинала.

Популярни категории проекти за ИИ с отворен код включват големите езикови модели (LLMs), инструменти за машинен превод, чатботове и др. За да могат разработчиците

на софтуер да произвеждат ресурси за ИИ с отворен код, те трябва да се доверят на различните други софтуерни компоненти с отворен код, които да използват при разработването му. Спекулира се, че системите с ИИ с отворен код имат потенциално повишен риск в сравнение с тези със затворен код, тъй като е хипотетично възможно от тях да бъдат премахнати ключови елементи като протоколите за безопасност и др. По подобен начин пък от друга страна се спекулира, че системите с ИИ със затворен код водят до повишен риск в сравнение с тези с отворен код поради проблеми със зависимостите от създателя, поверителността, непрозрачните алгоритми, корпоративния контрол и ограничената наличност, което същевременно може да доведе до потенциално забавяне в полезните иновации.

Историята на ИИ с отворен код е преплетена както с развитието на технологиите с ИИ, така и с растежа на движението за **софтуер с отворен код** (***open-source software movement***). ИИ с отворен код се развива значително през последните няколко десетилетия с приноса на различни академични институции, изследователски лаборатории, технологични компании и независими разработчици. Ще проследим много накратко това развитие от самото начало до наши дни.

3.2 Ранно развитие на системите с изкуствен интелект и софтуера с отворен код

Концепцията за ИИ датира от средата на 20-ти век, когато учени като Алън Тюринг (1912–1954) и Джон Маккарти (1927–2011) полагат основите на съвременните теории и алгоритми за ИИ. Ранните изследвания се фокусират върху разработването на системи за **символен ИИ с логически разсъждения** и базирани на правила **експертни системи**. През този период идеята за софтуер с отворен код започва да се оформя, като пионер в тази област е Ричард Столман (1953), който се застъпва за свободния софтуер като средство за насърчаване на сътрудничеството и иновациите в програмирането. Тези свои идеи той аргументира подробно в неговата статия *„Защо софтуерът трябва да бъде безплатен“*.

Фондацията за свободен софтуер (***Free Software Foundation***), основана през 1985 г. от Р. Столман, е една от първите големи организации, която популяризира идеята за софтуер, който може да бъде свободно използван, модифициран и разпространяван. Идеите от това движение в крайна сметка повлияват на развитието на ИИ с отворен код, тъй като повече разработчици започват да виждат потенциалните ползи от отвореното сътрудничество при създаването на софтуер, включително модели с ИИ и алгоритми за тях.

3.3 Поява на системи с изкуствен интелект с отворен код

През 90-те години **софтуерът с отворен код** започва да набира все по-голяма популярност, тъй като Интернет свързаността улеснява сътрудничеството отвъд географските граници. Възходът на **машинното самообучение** и **статистическите методи** също водят до разработването на все по-практични инструменти за ИИ. Въпреки това, едва в началото на новия век **ИИ с отворен код** започва да набира популярност с излизането на основополагащи библиотеки и рамки, които са достъпни за всеки да може да използва и при желание да допринесе за тяхното развитие.

Една от най-ранните рамки за ИИ с отворен код е ***Scikit-learn***, пусната през 2007 г. ***Scikit-learn*** бързо се превръща в една от най-широко използваните библиотеки за машинно самообучение поради своята лекота на използване и стабилна функционалност, предоставяща имплементации на множество общи алгоритми като класификация, регресия и групиране. По същото време други библиотеки за машинно обучение с отворен код като ***OpenCV***, ***Torch*** и

Theano са разработени от технологични компании и изследователски лаборатории, допълнително подкрепяйки растежа на системите с ИИ с отворен код.

3.4 Възход на моделите и платформите за изкуствен интелект с отворен код

Второто десетилетие на новия век отбелязва значителна промяна в развитието на ИИ, водена от появата на **дълбоко обучение** (***deep learning***) и **невронни мрежи** (***neural networks***). Рамките за дълбоко обучение с отворен код като ***TensorFlow*** (разработена от *Google Brain*) и ***PyTorch*** (разработена от *AI Research Lab* на *Facebook*) революционизират картата на ИИ, като правят сложните модели за дълбоко обучение по-достъпни. Тези рамки позволяват на изследователите и разработчиците да изграждат и обучават сложни невронни мрежи за специфични задачи като **разпознаване на изображения**, **обработка на естествен език** (NLP) и **автономно шофиране**.

През това време **генеративни модели** с ИИ като ***BERT*** на *Google* (2018) за обработка на естествен език и серията ***GPT*** на *OpenAI* (2018) за генериране на текст също стават широко достъпни в началото под формата на отворен код. Първоначалните модели целят да демонстрират потенциала на ИИ да промени индустриите чрез подобряване на разбирането и генерирането на човешките езици, предизвиквайки допълнителен интерес към разработването на ИИ с отворен код.

3.5 Ключови етапи при моделите с изкуствен интелект с отворен код в наши дни

През последните години се наблюдава **непрекъснат растеж** и **съзряване** на ИИ с отворен код. Компаниите и изследователските организации започват да пускат широкомащабни предварително обучени модели сред обществеността, което довежда до бум както в търговските, така и в академичните приложения на ИИ. По-специално, ***Hugging Face***, компания, фокусирана върху обработката на естествен език, се превръща в център за разработване и разпространение на най-съвременни модели с ИИ с отворен код, включително версии на трансформатори като ***GPT-2***, ***GPT-3*** и ***BERT***.

С обявяването на ***GPT-2***, *OpenAI* първоначално планира да запази изходния код на своите модели поверителен, цитирайки опасения относно „злонамерени“ приложения. След като обаче се сблъсква с обществена реакция, компанията публикува изходния код за *GPT-2* в *GitHub* три месеца след пускането му. *OpenAI* не е публикувала публично изходния код или предварително обучените тегла за моделите ***GPT-3*** или ***GPT-4***, въпреки че техните функционалности могат да бъдат интегрирани от разработчиците чрез API на ***OpenAI***.

Възходът на **големите езикови модели** (LLM) и в цялост на генеративния ИИ, като **GPT-3** (2020) на *OpenAI*, допълнително задвижва търсенето на рамки на ИИ с отворен код. Тези модели са използвани в различни приложения, включително създаване на чатботове, генериране на код, демонстрирайки широките възможности на системите с ИИ.

През 2024 г. *Meta* публикува колекция от големи езикови модели, включително ***Llama 3.1 405B***, които са сравними с най-модерните конкуренти със затворен код. Компанията твърди, че нейният подход към ИИ за тези модели ще бъде придържане към стандартите за отворен код, за разлика от други големи технологични компании. Оказва се обаче, че моделът не е напълно отворен, защото *Meta* не публикува огромните масиви от данни, които използва

за тренирането на модела. Това не остава незабелязано от ***Open Source Initiative*** и други организации, които заявяват, че ***Llama*** не е с отворен код, въпреки че *Meta* го описва като такъв. Още една причина за това е и софтуерният лиценз на ***Llama***, който забранява използването му за някои специфични цели.

4. Предизвикателства пред системите с изкуствен интелект с отворен код

Въпреки че системите с изкуствен интелект с отворен код предлагат уникални възможности за силно въздействие върху повечето бизнеси, **предизвикателствата**, до които те могат да доведат, трябва да бъдат правилно идентифицирани и адресирани. В тази секция ще категоризираме различни видове **ограничения** от гледна точка на правото, технологиите, данните, етиката и др.

4.1 Правни предизвикателства

Основните **правни предизвикателства**, които могат да бъдат срещнати по отношение на ИИ с отворен код, са адресирани в **Регламент (ЕС) 2024/1689** на Европейския парламент и на Съвета от 13 юни 2024 г. за установяване на хармонизирани правила относно изкуствения интелект (**Акт за изкуствения интелект**). След години на преговори се приема нов европейски регламент за регулиране на тази нововъзникваща технология.

Актът за ИИ въвежда изчерпателен набор от правила и задължения за доставчици, внедрители, вносители, дистрибутори и производители на системи с ИИ. Регламентът групира системите с ИИ в **четири категории**: неприемлив риск (*системи за дистанционна биометрична идентификация в реално време, когнитивно-поведенческа обработка, разпознаване на емоции на работното място и в образователните институции, социално оценяване и др.*), висок риск (*за диагностициране на заболявания, автономно управление на превозни средства, биометрична идентификация на лица, миграция, застраховане и др.*), ограничен риск (*чатботове, генериращи съдържание системи с ИИ и др.*) и минимален риск (*игри, филтри за спам и др.*). Най-общо **Актът** забранява системите с ИИ с неприемлив риск, изисква високорисковите системи с ИИ да следват определени правила, включително свързани с управлението на данните, техническата документация, мониторинг на риска и оценки на въздействието, както и установява изисквания за прозрачност за системи с ИИ с ограничен риск. В допълнение, **Регламентът** добавя допълнителни правила за прозрачност и отчетност за моделите на ИИ с общо предназначение, известни също като основни модели (foundation models).

Една точка на **дебат** до последните часове е как ЕС трябва да се справи с ИИ с отворен код – модели на ИИ, които разработчиците предоставят безплатно на обществеността. От една страна политиците са особено загрижени за въздействието на Акта за ИИ върху отворения код, тъй като много от по-успешните стартиращи компании за ИИ в Европа публикуват модели с отворен код. Докато окончателният текст на **Регламента** изключва системите за ИИ с отворен код от някои задължения, все пак тези изключения се прилагат само при определени ограничени условия. Сега да разгледаме в детайли какви са те.

В съображение 102 от Регламента се дава своеобразна дефиниция за системите с ИИ с отворен код:

> ***Софтуер** и **данни**, включително модели, предоставени с **безплатен лиценз с отворен код**, който им позволява да бъдат **споделяни** свободно и чрез който ползвателите могат свободно да ги **достъпват**, **използват**, **променят** и **разпространяват** в **променен** или **непроменен** вид, могат да допринесат за **научните изследвания** и **иновациите** на пазара и да осигурят значителни възможности за растеж на икономиката на Съюза. Следва да се обмисли използването на модели на ИИ с общо предназначение, представени с безплатни лицензи с отворен код, за да се гарантират високи равнища на **прозрачност** и **отвореност**, ако техните **параметри**, включително **теглата**, информацията за **архитектурата** на модела и информацията за **използването** на модела са **публично достъпни**. **Лицензът** следва да се счита за **безплатен** и с **отворен код** и когато дава възможност на ползвателите да **използват**, **копират**, **разпространяват**, **изучават**, **променят** и **подобряват** софтуер и данни, включително модели, при условие че се посочва първоначалният доставчик на модела и се спазват еднакви или сходни условия на разпространение.*

Това съображение обхваща два важни въпроса. Първо, то дава определение на **безплатен лиценз с отворен код**. Второ, в него се посочва, че ако теглата, архитектурата и информацията за използването на модела са публично достъпни под безплатен лиценз с отворен код, то тогава тези модели гарантират високо ниво на прозрачност и отвореност.

Задълженията на доставчиците на модели на ИИ с общо предназначение се регламентират в **член 53, параграф 1** от **Регламента**. Изключение от част от тези задължения се дава в параграф 2 на същия член за доставчиците на модели на ИИ, предоставени с **безплатен лиценз с отворен код**, който позволява достъп, използване, изменение и разпространение на модела, и чиито параметри, включително теглата, информацията за архитектурата на модела и информацията за използване на модела, са публично достъпни. **Изключенията** са именно изготвянето и актуализирането на техническата документация на модела, включително процеса на неговото обучение и изпитване, както и резултатите от неговата оценка, която съдържа най-малко информацията, посочена в **приложение XI** от Акта (**буква а)**), както и предоставянето на информация и документация на доставчиците на системи с ИИ, които възнамеряват да интегрират модела на ИИ с общо предназначение в своите системи с ИИ (**буква б)**). Тези две изключения не се прилагат за модели на ИИ с общо предназначение, пораждащи системни рискове (**член 51 разяснява необходимите условия за пораждане на системен риск**). Към текущия момент се счита, че даден модел на ИИ с общо предназначение има способности с **висока степен на въздействие** (т.е. поражда системен риск), когато общото количество изчисления, използвани за неговото обучение, измерено при операции с плаваща запетая, е по-голямо от 10^{25} (**член 51, параграф 2**). Тези стойности подлежат на преразглеждане, като това става по по-бърза процедура чрез делегирани актове, приемани от Комисията (**член 51, параграф 3 на основание член 97**).

Остава отворен въпросът дали системите с ИИ с отворен код трябва да предоставят **достъп до набора от обучителни данни**, които използват. Сходно задължение може да бъде намерено в **член 53, параграф 1, буква г)** от **Регламента**.

> *представят и оповестяват публично достатъчно **подробно обобщение** на съдържанието, използвано за обучение на модела на ИИ с общо предназначение, по **образец**, осигурен от **Службата по ИИ**.*

Във всички случаи, предвиденото за моделите на ИИ с общо предназначение с отворен код изключение от съответствието с изискванията, свързани с **прозрачността** (посочени по-горе), не следва да се отнася до задължението за изготвяне на обобщение на съдържанието,

използвано за обучение на модели, и задължението за въвеждане на политики за спазване на правото на Съюза в областта на авторското право. Това се прави „като се има предвид, че предоставянето на модели на ИИ с общо предназначение с безплатен лиценз с отворен код не разкрива непременно съществена информация относно набора от данни, използван за обучението или финото регулиране на модела, и относно това как по този начин се гарантира спазването на авторското право“ (**съображение 104**).

В **преамбюла** на **Акта за ИИ** се посочва, че целта на това „**достатъчно подробно обобщение**“ е да се улесни упражняването и прилагането на права съгласно правото на Съюза от страни със законен интерес. Легитимният интерес може да се отнася до защитата на авторското право, което е изрично упоменато в **съображение 107**:

> *За да се повиши **прозрачността** по отношение на **данните**, които се използват при предварителното обучение и обучението на **модели на ИИ с общо предназначение**, включително **текст** и **данни**, защитени от правото в областта на **авторското право**, е целесъобразно доставчиците на такива модели да изготвят и оповестяват публично **достатъчно подробно обобщение** на съдържанието, използвано за обучение на модела на ИИ с общо предназначение. Като се отчита надлежно необходимостта от защита на **търговските тайни** и **поверителната търговска информация**, това обобщение следва като цяло да бъде **изчерпателно** по своя обхват, а **не технически подробно**, за да улесни лицата със **законни интереси**, включително носителите на **авторски права**, да упражняват и прилагат правата си съгласно правото на Съюза, например чрез изброяване на основните събрани данни или набори от данни, които са използвани при обучението по модела, като например **големи частни или публични бази данни** или **архиви с данни**, и чрез предоставяне на **описателно разяснение** за други използвани източници на данни. Целесъобразно е Службата по ИИ да осигури **образец на обобщението**, който следва да бъде **опростен**, **ефективен** и да позволява на доставчика да предостави изискваното обобщение в **описателна форма**.*

Разпоредбата за прозрачност в **член 53, параграф 1, буква г)** произтича от измененията, предложени от Европейския парламент. Тези изменения са въведени в отговор на **опасения**, повдигнати от различни организации, представляващи **автори** и други носители на права в рамките на **културните** и **творческите индустрии**, тъй като много модели на ИИ с общо предназначение, особено генеративни езикови модели, се обучават върху резултатите от човешко творчество, голяма част от които са **защитени с авторски права**. Прозрачността на данните за обучение е необходима, за да се позволи на създателите да определят дали техните произведения са включени в данните за обучение на съответния модел на ИИ с общо предназначение.

Съгласно **член 4, параграф 3** от **Директивата за авторското право и сродните му права в цифровия единен пазар от 2019 г.** (**Директива ЕС 2019/790**) авторите и другите носители на права **имат право** да се откажат от използването на техните произведения за извличане на текст и данни (**транспонирано в член 26е, алинея 4 и изключението за научни цели в член 26ж от ЗАПСП**). **Регламентът за ИИ** пояснява, че обучението на генеративни модели на ИИ е форма на извличане на текст и данни (**съображение 105**). На този фон **член 53, параграф 1, буква в)** от него изисква доставчиците на модели на ИИ с общо предназначение да „въвеждат политика за спазване на правото на Съюза в областта на авторското право и сродните права, и по-специално с цел установяване и съобразяване, включително чрез най-съвременните технологии, на случаите на запазване на права съгласно **член 4, параграф 3 от Директива (ЕС) 2019/790**“. Следователно, в контекста на авторското право, разпоредбата за прозрачност в **член 53, параграф 1, буква г)** има за цел да позволи на

авторите и другите носители на права да проверят дали доставчиците на генеративни модели спазват двете условия за законно извличане на текст и данни, съдържащи се в изключението в **член 4 от горецитираната Директивата** – от една страна, че произведенията, използвани като данни за обучение, са били законно достъпни (**член 4, параграф 1 – Директива ЕС 2019/790**) и от друга страна, че не са били „изключени“ (**член 4, параграф 3 – Директива ЕС 2019/790**).

В допълнение към защитените с авторски права материали, съдържанието, използвано за обучение, валидиране или тестване на системи с ИИ с общо предназначение, може да включва **лични данни**. Съгласно **европейското законодателство за защита на личните данни** лицата (субектите на данни) имат редица права на достъп. Сред тях се включват възможност за **достъп** до личните данни, **коригиране** на неточности в личните данни, **искане** за **коригиране** или **изтриване** на личните данни и др. (**чл. 15 от Регламент ЕС 2016/679 – ОРЗД**) Утвърждаването на тези права може да бъде трудно в случай на работа с обучителни данни при системите с ИИ. Необходима е **прозрачност**, за да може да се позволи на субектите на данни първо да определят дали доставчиците на модели **обработват** техни лични данни и второ, да имат възможност да **упражняват** правата си съгласно **Общия регламент относно защитата на данните**.

В случаите, когато доставчиците искат да пуснат **модел на ИИ с общо предназначение** на пазара на Европейския съюз, но са установени в **трети държави**, те трябва с **писмено пълномощно** да определят **представител**, който е установен на територията на Съюза. Доставчикът осигурява възможност на упълномощения си представител да изпълнява задачите, посочени в пълномощното, получено от доставчика. Тези задължения **не се прилагат** за доставчиците на модели на ИИ с общо предназначение, предоставени с **безплатен лиценз с отворен код**, който позволява достъп, използване, изменение и разпространение на модела, и чиито параметри, включително теглата, информацията за архитектурата на модела и информацията за използването на модела, са публично достъпни. (**член 54, параграф 6 – Регламент ЕС 2024/1689**). Задълженията все пак се прилагат, ако говорим за модел на ИИ с общо предназначение, пораждащ системни рискове (**член 54, параграф 6 – Регламент ЕС 2024/1689**).

Що се отнася цялостно до обхвата на **Регламента** за ИИ по отношение на системите с ИИ с отворен код, това се регламентира в **член 2, параграф 12**:

> *Настоящият регламент **не се прилага** за системи с ИИ, предоставени с **безплатен лиценз с отворен код**, освен ако не са пуснати на пазара или пуснати в действие като **високорискови системи** с ИИ или като система с ИИ, попадаща в обхвата на **член 5** или **50**.*

Регламентът не прави изключения за системи с ИИ с отворен код, когато става дума за забрани за ИИ с неприемлив риск и ограничения за високорискови системи с ИИ. За други системи, с изключение на такива с общо предназначение, **Актът за ИИ** не се прилага за трети страни, които правят публично достъпни продукти с ИИ с отворен код (**съображение 89 и член 25, параграф 4 – Регламент ЕС 2024/1689**). Това изключение обаче се прилага само ако те не монетизират своите продукти. В резултат на това всяка компания, която се опитва да монетизира своите продукти с ИИ с отворен код, като предлага платена техническа поддръжка за модела с отворен код или използва насочени реклами за покриване на разходи, няма да може да се възползва от това изключение (**съображение 103**). Законът също така посочва, че разработчиците с отворен код „следва да се насърчават да прилагат широко възприети практики за документиране, например **карти за модели** и **информационни фишове**, като

начин за ускоряване на обмена на информация по веригата за създаване на стойност в областта на ИИ“ (**съображение 89**), но не предоставя подробности за това как трябва да изглежда това насърчаване на практика.

Въпреки че европейските политици се опитват да отговорят на някои от опасенията на **общността на отворения код** (***open-source community***), много проекти, включващи системи с ИИ с отворен код, все още попадат под правилата на **Акта за ИИ**. В някои случаи може да се окаже, че когато една компания едностранно разработва модел на ИИ с отворен код, съответствието няма да е по-различно от това, ако компанията разработва собствен (частен) модел на ИИ. Всъщност няма конкретна причина правилата да благоприятстват или наказват бизнес моделите с отворен код. Въпреки това, за проекти с ИИ с отворен код, които се основават предимно на децентрализиран принос на отделни разработчици, които не са подкрепени от една компания, сложността на тези нови правила може да **затрудни** въвеждането на такъв тип софтуер в рамките на Европейския съюз.

4.2 Технически предизвикателства

Наследените традиционни системи за **бази данни**, използвани в публичния или частния сектор, невинаги са подходящи за управление и съхраняване на данни от **неструктурирани източници**. Правителствата и другите организации трябва да приемат подходяща **техническа инфраструктура** за управление на **големи данни** и това идва с предизвикателството за инвестиране и модернизиране на ИТ инфраструктурите на публичните администрации. Тъй като има много налични **технически решения**, вземането на такива относно отделните системи, както и комбинациите, които да се приемат, става все по-сложно. Трябва да се вземе предвид **съвместимостта** на системите с отворен код с текущите такива, както и възможността за поддръжка на технологията след като бъде внедрена. Интегрирането на огромното количество данни и оперативната съвместимост между различни ИТ системи е важно условие за успешна **цифрова трансформация**.

Софтуерът с отворен код не е по своята същност по-малко сигурен от частния софтуер със затворен код. Въпреки това, една основна разлика е, че когато се открият **уязвимости** в софтуер с отворен код, е много по-вероятно те да бъдат **разкрити публично**. Злонамерените участници често следят за тези разкривания на уязвимости с надеждата да **злоупотребят** със системи на компании, които зависят от податливи версии на тези библиотеки с отворен код и компилации на приложения. Това е по-малък проблем със затворения код, тъй като компаниите обикновено избягват да разкриват подробности за уязвимостите в своя софтуер пред обществеността или това става едва след като вече софтуерът е коригиран и актуализиран.

За да справят по-бързо с това предизвикателство, организациите, които използват отворен код, трябва внимателно да следят кои библиотеки с отворен код, приложения или други ресурси използват. Това трябва да включва проверка дали съществуват известни уязвимости в тези ресурси. **Инструментите за анализ на състава на софтуера** (*Software Composition Analysis*) са полезни за тази цел, тъй като те автоматично **сканират** кодови бази, за да **идентифицират** всички компоненти с отворен код, които се намират в тях, и да **маркират** всичко, за което е известно, че съдържа потенциални уязвимости в сигурността.

Софтуерът със затворен код обикновено се поддържа, актуализира и коригира от доставчиците на софтуер, което може да бъде голяма полза за екипите по разработка, които нямат време, ресурси или опит да го направят сами. Съществуват платформи с отворен код,

които получават активна поддръжка от доставчици на софтуер, като ***Red Hat Enterprise Linux*** и търговски дистрибуции на ***Kubernetes***.

В по-голямата си част обаче организациите, които внедряват софтуер с отворен код, са отговорни да гарантират, че той остава **актуализиран**. Неспазването на това крие риск от изпълнение на остарял код, който може да бъде с **грешки** или да има **уязвимости** в **сигурността**. Това предизвикателство се изостря още повече в случаите на липса на централизирани конзоли за управление или автоматизирани процеси за актуализиране, които могат да гарантират, че всички използвани компоненти с отворен код са актуални – нещо, което често се изтъква като предимство на частните софтуерни пакети. За щастие, някои видове софтуер с отворен код, като ***Linux*** дистрибуциите, предоставят мениджъри на пакети, които автоматично **актуализират** библиотеките и приложенията, инсталирани на тях, всеки път, когато бъдат пуснати нови версии.

4.3 Предизвикателства, свързани с данните

Дори преди ***ChatGPT*** да излезе на сцената, има безброй примери за ИИ, склонен към **пристрастия** (biases) и **несправедливост** (inequities). Част от решението на този проблем е предоставянето на по-добър **набор от данни за обучение**. Но ако не знаем на какво се обучава даден модел с ИИ, то не знаем какъв вид пристрастия възпроизвежда. Това всъщност е проблем с данните, а не проблем с теглата на модела. Дори и най-напредналата система с ИИ в света винаги ще произвежда предубедени резултати, ако се обучава на предубедени източници.

Без да се изисква данните да са отворени, не е възможно никой, който не разполага с тях, да проучи напълно или модифицира системи с ИИ с общо предназначение. Можем само да ги **използваме**, да ги **настройваме**, но не можем да се гмурнем надълбоко в тях, за да разберем защо правят определено действие.

Пристрастията живеят в данните. Вярно е, че алгоритмите и настройките на параметри играят роля, но ако тренираме модел, използвайки само снимки на мишки например, той ще бъде наистина предубеден (и недобър) в създаването на стихотворения. Ако по-голямата част от тези мишки са бели, то е очевидно какъв цвят мишки ще генерира. Големите езикови модели не могат сами да си съчинят напълно нова информация. Те са обучени да вземат данните, на които са били тренирани, и да „произвеждат“ повече неща, които приличат на тях. В този дух е и стремежът дефиницията за системи с ИИ с отворен код да изисква самите тренировъчни данни, а не само информация за тях, както е към текущия момент във **версия 1.0** на ***The Open Source AI Definition*** от ***Open Source Initiative***.

4.4 Предизвикателства по управление на риска

Самият **Акт за ИИ**, както стана ясно в секцията за правните предизвикателства, разчита на **риска** като основен критерий за категоризиране на системите с ИИ. От друга страна, ако физическите лица или предприятията не защитават или управляват добре своите системи с отворен код, съществува повишен риск от **пробив в сигурността**. Една такава ситуация се наблюдава в системата ***Apache Struts***, в която е била открита уязвимост, която засяга правителствени и бизнес организации, които са били свързани към системата. Нарушителите на системата в този случай са успели да манипулират засегнатото приложение, сякаш имат пълни потребителски права: преглед, промяна, изтриване на данни или създаване на нови акаунти. Повечето системи с отворен код често нямат специална техническа поддръжка и без такъв екип **актуализациите** и **корекциите** за сигурност не могат да бъдат налични навреме.

Ако се открият уязвимости в софтуера, участниците в **киберзаплахите** могат да използват тези уязвимости, за да получат достъп до мрежата, системите и информацията на организацията. Поради публичния характер на системите с ИИ с отворен код, проблемите с киберсигурността също са предизвикателство. Системите, които публичните и частните институции избират да внедрят, може да се окажат уязвими и податливи на атаки от други системи с ИИ. Всъщност ИИ може да се използва както от хакери, така и от защитници в сценарии на киберотбрана. Например, когато се разглеждат правителствени системи с модул за ИИ, които са изградени за управление и контрол на достъпа, противниците могат да компрометират много техники само и единствено чрез кражба на ключове за оторизация. Наблюдението на поведенческите модели от системи с ИИ също може да доведе до нарушения на поверителността и поради тази причина трябва да бъде забранено, ако има за цел манипулиране на поведението на потребителите, както е и е в Акта за изкуствения интелект, освен ако не се изисква за обществена сигурност (**член 5 – Регламент ЕС 2024/1689**).

Един друг риск може да бъде неволното **споделяне на частен (proprietary) код** или **чувствителни вътрешни данни** в публично хранилище с отворен код. Това може да се случи по много начини, като например, когато разработчиците споделят по-голяма част от своята база, отколкото са възнамерявали, или дори може да се случи да оставят **коментари** вътре в кода, които включват поверителна информация за вътрешни системи. Един от начините, по който този риск може да бъде митигиран, е чрез установяването на ясни **политики за управление**, които регулират кога и как разработчиците могат да добавят код към проекти с отворен код. Тези политики трябва да включват **сканиране на кода**, преди да бъде направен публично достояние, както и налагане на цялостни процеси на преглед, които проверяват за потенциални рискове за поверителността и сигурността във всичко, качено в хранилище с отворен код.

4.5 Обществени и етични предизвикателства

От **обществена** и **етична гледна точка** отвореният код може да доведе до редица предизвикателства в публичния сектор. Първото от тях е получаването на **социално приемане** и **доверие** в технологията както от гражданите, така и от държавните служители. Ако гражданите се чувстват **дискриминирани** от ИИ или чувстват, че той заплашва тяхната безопасност, неприкосновеност на личния живот или работата им, като по този начин противоречи на техните очакванията, то е много малко вероятно те да го приемат или да му се доверят. Веднъж установено, доверието става **крехко** и трябва да се поддържа чрез **надеждност**, **сътрудничество** и много важно **комуникация**. В последните месеци се наблюдава масово говорене по темата за ИИ от коментатори и журналисти, които не са експерти в конкретната област. Това само по себе си крие висока степен на опасност, защото широката общественост остава с погрешно впечатление за технологията. В резултат на това, различните поколения имат различни представи за това какво е ИИ. **Негативните примери** от развитието на ИИ биват много по-често отразявани, отколкото добрите. Аргументите в полза на ИИ остават **нечути**. В България отсъства и дискусията за системите с ИИ с отворен код, такава каквато трябва да има поне за редица софтуерни системи за управление на процеси в административни структури най-малкото в столицата и други големи градове в страната.

Докато системите с отворен код могат да увеличат прозрачността, трудното разбиране на тези алгоритми ще изисква **специфичен надзор** чрез модела на „човек в цикъла“ (***human-in-the-loop***), „обществото в цикъла“ (***society-in-the-loop***) или дори спазването на цял рамков модел. Освен това, точността при анализа на данни се е подобрила до степен, в която дори и за по-сложни задачи е по-добра от човешките способности. Тъй като системите с ИИ с отворен код ще поемат все по-голям брой задачи, предизвикателството е как да се помогне на

работниците да придобият необходимите умения чрез **трансформация на работната сила**, за да се възползват от новите възможности за работа, които тези системи предоставят. Въпреки че има възможности за поддържащите отворен код да получат финансова подкрепа за своя принос, от съществено значение е правителствата да внедрят **ясна структура по финансиране** на системи с ИИ с отворен код. Изключителната зависимост от доброволци може да доведе до проблеми с управлението на риска, ако се стигне до ситуация да липсват такива или да са прекалено малко. Правителствата трябва да осигурят **адекватно възнаграждение** за работата, извършена от разработчиците на ИИ с отворен код, и да предоставят допълнителна подкрепа, като инвестиране в преглед на коментари, въпроси и предложения за подобряване на интелигентните системи.

Етичните проблеми също все повече излизат на преден план, тъй като системите с ИИ стават неразделна част от вземането на решения в критични области като здравеопазване, наказателно правосъдие и финанси. Тези опасения включват различни **пристрастия**, **отчетност**, **прозрачност** и **потенциална злоупотреба**. Системите с отворен код могат да бъдат приведени в съответствие с принципите на етичния ИИ чрез следните методи:

- **Прозрачност и отчетност** – системите с ИИ с отворен код да гарантират, че алгоритмите и наборите от данни са достъпни за проверка. Тази прозрачност е от решаващо значение за идентифициране и коригиране на пристрастия в системите с ИИ. Така се насърчава и отчетността, тъй като разработчиците и потребителите разбират как системата взема решение (***eXplainable AI***);
- **Разнообразно и безпристрастно развитие** – съвместният характер на проектите с отворен код осигурява разнообразен набор от сътрудници. Това разнообразие е от решаващо значение за разработването на безпристрастни системи с ИИ;
- **Иновация с отговорност** – общностите, занимаващи се с развитието на софтуер с отворен код, насърчават иновациите с акцент върху етичните съображения. Тъй като тези общности често включват специалисти по етика, социални учени и различни групи крайни потребители, те са по-склонни да анализират общественото въздействие от системите с ИИ;
- **Демократизиране на ИИ** – системите с ИИ с отворен код демократизират достъпа, тъй като гарантират, че те не са пряка собственост на няколко мощни субекта. Тази демократизация е от съществено значение за предотвратяване на монополите и насърчаване на етичната употреба;
- **Насърчаване на глобални стандарти** – чрез насърчаване на международната общност от разработчици и потребители, системите с ИИ с отворен код могат да помогнат в установяването на глобални стандарти за етичен ИИ, което е от решаващо значение предвид безграничния характер на технологията и данните.

5. Предимства и недостатъци на системите с изкуствен интелект с отворен код

5.1 Предимства

Системите с ИИ с отворен код дават на потребителите пълен **контрол** върху модела, което им позволява да го **притежават** и да го **променят** завинаги според техните условия. Затворените модели са по-ограничени, а в немалко случаи и напълно изчезват. Такъв е случаят с компанията *OpenAI*, която спира да развива своя модел ***GPT-3.5***, което се отразява на качеството на всички проекти, стъпили върху него.

Моделите с отворен код също така предоставят по-голям контрол върху това как и къде се внедрява даден модел, което може да подобри аспектите, свързани с поверителността на данните. Чрез изпълнение на модели локално или в частна облачна инфраструктура, организациите могат да защитят **чувствителни данни**, без да разчитат на облачни услуги на трети страни, както често се изисква при затворените модели.

Друго **предимство** на отворения код е, че се позволява на потребителите да **адаптират** своите модели за **конкретни случаи на употреба** и **нужди на индустрията**. Компаниите могат да коригират своите параметри, да ги прецизират с допълнителни данни и да ги оптимизират за поставени задачи, като прогнозни анализи, автоматизация на бизнеса, езикови преводи и др.

Отворените модели позволяват на потребителите да **изследват** кода, данните за обучение и структурите на модела, които определят как работят. При затворените модели, от друга страна, тази информация е скрита в т.нар. „***черна кутия***“, което затруднява потребителите в идентифицирането на начина на работа и откриването на потенциални слабости. Тази прозрачност обаче не прави непременно отворените модели по-обясними. Дори, ако целият код, данни и тегла са напълно достъпни за проверка, все още е възможно да бъде трудно, често дори и невъзможно, да се разбере точно как и защо даден модел се държи по определен начин.

Положителна страна на системите с ИИ с отворен код е, че те помагат да направят разработката на ИИ **по-достъпна**, тъй като предоставят немалко ресурси и инструменти, които премахват **бариерите** за навлизане в областта. Разполагайки с модели с отворен код, библиотеки и рамки, практически всеки, от любител до професионалист, може да създаде свои собствени системи с ИИ без значителни финансови инвестиции или специализиран опит в тази сфера. Големите известни платформи като ***Hugging Face*** и ***TensorFlow*** предлагат солидна документация и поддръжка от общността, която стои зад тях, позволявайки на всеки начинаещ да учи и експериментира с ИИ със свое собствено темпо.

Общността играе жизненоважна роля в системите с ИИ с отворен код, подхранвайки **иновациите** чрез **сътрудничество** и **колективно решаване на проблеми**. Когато използваме ИИ с отворен код, ние ефективно се включваме в голяма, разнообразна мрежа от разработчици, които непрекъснато допринасят за непрекъснатото подобряване на тези инструменти, споделяйки информация и надграждайки взаимно работата си.

5.2 Недостатъци

За компаниите, които се занимават със системи с ИИ с отворен код, **реализирането на печалби** от тях може да бъде предизвикателство, тъй като тяхното разработване е **скъпо** и може да бъде **трудно** да се възстановят тези разходи, тъй като обикновено моделите се предлагат **безплатно**. За да се справят с предизвикателствата на монетизацията, някои компании продават услуги и приложения от бизнес клас в допълнение към своите отворени модели, като таксуват клиентите, които използват моделите за корпоративни функции и поддръжка. Някои компании предлагат комбинация от безплатни отворени модели и по-мощни затворени модели, които изискват такса или платен абонамент за достъп.

Друг **недостатък** е, че въпреки че отварянето на модели предоставя на потребителите по-голям контрол, то ефективно **намалява количеството контрол**, което организациите, създаващи тези модели, имат върху тях. **Лицензирането с напълно отворен код** позволява

на потребителите да променят и разпространяват свободно модели, което прави невъзможно компаниите да налагат ограничения върху използването.

Пускането сред обществеността на мощни големи езикови модели и мултимодални модели не е напълно без риск. Тези модели могат да бъдат по-лесно адаптирани за злонамерени цели, като например генериране на **подвеждаща информация**, създаване на **дълбоки фалшификати** (*deepfake*) и автоматизиране на **фишинг атаки**. Решение на този проблем може да бъде дадено от общността на разработчиците на системи с ИИ с отворен код. Като насърчава **сътрудничеството** и **прозрачността**, тази общност може бързо да реагира на възникващи заплахи, дори по-бързо от хората, работещи върху собствени модели в големи технологични компании.

Предизвикателство пред отворените модели с ИИ е и, че всяка промяна подлежи на **одобрение от общността**, което в някои моменти може да забави интервалите за иновации в сравнение с частните модели, които имат силно централизирано управление. В подобно отношение са и въпросите за **сигурността**, тъй като комерсиално разработените модели работят със строги и целенасочени процеси.

6. Примери за системи с изкуствен интелект с отворен код

6.1 Hugging Face

Hugging Face е платформа и общност, която помага на потребителите да изграждат модели за **машинно самообучение**, като предоставя инфраструктура за **обучение**, **стартиране** и **внедряване** на ИИ приложения. Платформата поддържа **библиотека** от повече от милион модели с отворен код с най-разнообразни задачи, сред които обработка на естествен език, компютърно зрение и др. ***Hugging Face*** разработва собствен езиков модел с отворен код, наречен ***BLOOM***, който основно се справя със задачи за създаване на междуезично съдържание и превод. Общността улеснява потребителите да споделят открито ресурси, модели и изследвания, помагайки за намаляване на времето и ресурсите за обучение на модели.

6.2 TensorFlow

TensorFlow е софтуерна библиотека с отворен код, създадена от *Google*, която помага на разработчиците да изграждат и внедряват модели за **машинно самообучение** на настолни, мобилни, уеб, облачни и IoT устройства. Предлага селекция от предварително **обучени** и **изследователски модели**, които потребителите могат да прецизират и персонализират с допълнителни данни за изпълнение на нови задачи. ***TensorFlow*** поддържа множество езици за програмиране, сред които *Python, JavaScript* и др. Платформата предлага безплатни **уроци**, **курсове** и **сертификати**, за да помогне на обучаващите се да научат основите на разработването на такъв тип системи с ИИ.

6.3 PyTorch

PyTorch е рамка, базирана на езика за програмиране *Python* и библиотеката *Torch*, която се използва за обучение на **невронни мрежи**. Първоначално е разработена от *Meta AI*, но сега е част от *Linux Foundation* – организация с нестопанска цел, която поддържа проекти за

софтуер с отворен код. Има **широка екосистема** от инструменти и модели като *TorchVision* (за задачи на компютърното зрение), *TorchText* (за задачи за обработка на естествен език) и *TorchAudio* (за задачи на аудио обработката). Използва **тензори** (*Tensors*) (специализирани структури от данни, които работят на графични процесори) за кодиране на **входовете**, **изходите** и **параметрите** на модел, като помага за **ускоряване** на **изчислителния процес**.

6.4 Scikit-learn

Scikit-learn е библиотека на *Python* с отворен код, предназначена за **машинно самообучение**, **прогнозни анализи** и **статистическо моделиране**. Включва различни алгоритми за *класификация*, *регресия* и *клъстеризация*. Осигурява различни инструменти за **предварителна обработка на данни**, **избор на модел**, **оценка на модела** и др. Проектирана е за взаимодействие с цифрови и научни библиотеки на *Python*, включително ***Pandas***, ***NumPy***, ***SciPy*** и др. Разполага с активна **общност** и предоставя множество **ресурси** за обучение.

6.5 Keras

Keras е базирана на *Python* библиотека за **невронни мрежи**, фокусирана върху изграждането и обучението на модели за **дълбоко обучение**. Работи върху различни рамки като *TensorFlow*, *PyTorch*, *JAX*. Предлага десетки модели за **дълбоко обучение**, заедно с предварително обучени тегла, които могат да се използват за **прогнозиране**, **извличане** на характеристики и фина **настройка**.

6.6 ClearML

ClearML е платформа с отворен код, предназначена да **автоматизира**, **наблюдава** и **организира** развитието на алгоритми за **машинно самообучение** – от такива с изследователски цели до комерсиални. Позволява на потребителите да **интегрират** модели за машинно самообучение, дълбоко самообучение или езикови модели във всеки голям набор от данни във всяка архитектура с вече съществуваща технологична рамка или стек. Предлага се с допълнителни **платени добавки** като приоритетна поддръжка, управлявани услуги, управление на разрешения и др. Платформата поддържа **локални** (*on-premise*), **въздушни** (*air-gapped*), **облачни** (*cloud*) и **хибридни** (*hybrid*) среди.

7. Заключение

Системите с ИИ с отворен код олицетворяват духа на **сътрудничеството** и **демократизацията** на технологиите, но тяхната стойност надхвърля простата **достъпност**. Те са катализатор за **иновации**, **прозрачност** и **етично развитие**, като позволяват на **общности** по целия свят да **допринасят** към напредъка на ИИ. Въпреки това тяхното **ефективно** и **отговорно използване** изисква осъзнаване на **рисковете**, като злоупотреба или липса на **контрол**, и поставяне на ясни **рамки**, включително такива от **правна** гледна точка чрез **Акта за ИИ** на ниво **Европейски съюз**. **Отвореният код** е едновременно **предимство** и **предизвикателство** – кое до каква степен ще надделее, зависи от нас самите.

8. Използвани литературни източници

[1] **Директива (ЕС) 2019/790 на Европейския парламент и на Съвета** от 17 април 2019 година относно авторското право и сродните му права в цифровия единен пазар [2019]. *Официален вестник на Европейския съюз* L 130. [Прегледан 8.01.2025]. Достъпно от: https://eur-lex.europa.eu/legal-content/BG/ALL/?uri=CELEX:32019L0790

[2] **Закон за авторското право и сродните му права**. Обнародван: ДВ, бр. 56 от 29 Юни 1993 г.; изменен: ДВ, бр. 70 от 20 Август 2024 г. [Прегледан 8.01.2025]. Достъпно от: https://lex.bg/laws/ldoc/2133094401

[3] **Регламент (ЕС) 2016/679 на Европейския парламент и на Съвета** от 27 април 2016 година относно защитата на физическите лица във връзка с обработването на лични данни и относно свободното движение на такива данни (***Общ регламент относно защитата на данните***) [2016]. *Официален вестник на Европейския съюз* L 119. [Прегледан 08.01.2025]. Достъпно от: https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=CELEX:32016R0679

[4] **Регламент (ЕС) 2024/1689 на Европейския парламент и на Съвета** от 13 юни 2024 година за установяване на хармонизирани правила относно изкуствения интелект (***Акт за изкуствения интелект***) [2024]. *Официален вестник на Европейския съюз* OB L. [Прегледан 08.01.2025]. Достъпно от: https://eur-lex.europa.eu/legal-content/BG/TXT/?uri=CELEX:32024R1689

[5] **Arsanjani**, Ali. *Comparative Analysis of the Pros and Cons of Open-source, Permissive and Proprietary Foundation Models* [online]. Medium, 19.02.2024. [Viewed 8.01.2025]. Available from: https://dr-arsanjani.medium.com/comparative-analysis-of-the-pros-and-cons-of-open-source-permissive-and-proprietary-foundation-f34e6cdad09b

[6] **Castro**, Daniel. *The EU's AI Act Creates Regulatory Complexity for Open-Source AI* [online]. Published 4.03.2024. [Viewed 8.01.2025]. Available from: https://datainnovation.org/2024/03/the-eus-ai-act-creates-regulatory-complexity-for-open-source-ai/

[7] **Glover**, Ellen. *Open Source AI: Definition and 10 Platforms to Know* [online]. Built In, 6.11.2024. [Viewed 8.01.2025]. Available from: https://builtin.com/artificial-intelligence/open-source-ai

[8] **Haddad**, Ibrahim. *Artificial Intelligence and Data in Open Source* [online]. The Linux Foundation, March 2022. [Viewed 8.01.2025]. Available from: https://www.linuxfoundation.org/research/artificial-intelligence-and-data-in-open-source

[9] **Nair**, Vandana. *The Hidden Risks in Open-Source AI Models* [online]. AIM, 31.07.2024. [Viewed 8.01.2025]. Available from: https://analyticsindiamag.com/ai-trends/the-hidden-risks-in-open-source-ai-models/

[10] **OpenCV: Thoughts on Open Source & AI Ethics** [online]. Published 10.01.2024. [Viewed 8.01.2025]. Available from: https://opencv.org/blog/thoughts-on-ai-ethics/

[11] **Open Source Initiative: The Open Source AI Definition – 1.0** [online]. Version 1.0. [Viewed 8.01.2025]. Available from: https://opensource.org/ai/open-source-ai-definition

[12] **Open Source Initiative: The Open Source Definition** [online]. Version 1.9, last modified 22.03.2007. [Viewed 8.01.2025]. Available from: https://opensource.org/osd

[13] **Robison**, Kyie. *Open-source AI must reveal its training data, per new OSI definition* [online]. The Verge, 28.10.2024. [Viewed 8.01.2025]. Available from: https://www.theverge.com/2024/10/28/24281820/open-source-initiative-definition-artificial-intelligence-meta-llama

[14] **Tarkowski**, Alek. *AI Act fails to set meaningful dataset transparency standards for open source AI* [online]. Open Future, 7.03.2024. [Viewed 8.01.2025]. Available from: https://openfuture.eu/blog/ai-act-fails-to-set-meaningful-dataset-transparency-standards-for-open-source-ai/

[15] **Theben**, A., **Gunderson**, L., **López Forés**, L., **Misuraca**, G., **Lupiáñez Villanueva**, F., *Challenges and limits of an open source approach to Artificial Intelligence*, study for the Special Committee on Artificial Intelligence in a Digital Age (AIDA) [online], Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2021. [Viewed 8.01.2025].
Available from: https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2021)662908

[16] **Tozzi**, Chris. *6 common problems with open source code integration* [online]. Published 22.11.2023. [Viewed 8.01.2025]. Available from: https://www.techtarget.com/searchapparchitecture/tip/Common-problems-with-open-source-code-integration

[17] **Warso**, Zuzanna, **Gahntz**, Maximilian, **Keller**, Paul, et. al. *Sufficiently detailed? A proposal for implementing the AI Act's training data transparency requirement for GPAI* [online]. Open Future Foundation, June 2024. [Viewed 8.01.2025]. Available from: https://openfuture.eu/wp-content/uploads/2024/06/240618AIAtransparency_template_requirements-2.pdf

[18] **Wikipedia: Open-source artificial intelligence** [online]. [Viewed 8.01.2025]. Available from: https://en.wikipedia.org/wiki/Open-source_artificial_intelligence