

FACTCHECK.BG

Асоциация на европейските журналисти – България

КОДЕКС ЗА УПОТРЕБА НА ИЗКУСТВЕН ИНТЕЛЕКТ В РЕДАКЦИЯТА НА [FACTCHECK.BG](https://factcheck.bg)

Предговор

Factcheck.bg е платформа за проверка на информацията и борба срещу дезинформацията, член на Европейската мрежа за стандарти при проверката на факти (EFCSN). [Factcheck.bg](https://factcheck.bg) е проект на сдружение „Асоциация на европейските журналисти-България“.

Мисията на редакцията е да противодейства на дезинформацията в България. Factcheck.bg носи отговорност пред читателите, партньорите и журналистическата общност.

Изкуственият интелект навлиза трайно в ежедневната работа на медиите. За Factcheck.bg той е инструмент с реален потенциал за по-ефективна работа на редакцията, за достигане до по-широка аудитория и за създаване на по-достъпно проверено съдържание. Същевременно, технологията носи рискове за независимостта, точността и доверието в журналистиката. Настоящият кодекс не ограничава употребата на ИИ, а я регулира с ясни правила за това кое е допустимо, кое е забранено и при какви условия може да се експериментира в работата с инструменти с изкуствен интелект.

Документът е разработен на основата на политиките на Агенция Франс Прес, Асошиейтед прес, Ройтерс и БиБиСи, Парижката харта за ИИ в журналистиката, насоките на EFCSN и Съвета на Европа, както и практиките на водещи организации, занимаващи се с проверка на фактите. Настоящият правилник включва и специален раздел за проекта VALIDAI - пилотна инициатива за адаптиране на проверено съдържание за различни аудитории с помощта на ИИ инструменти.

Основен принцип

Технологията служи на журналистиката, а не обратното. Изкуственият интелект е инструмент и не може да бъде считан за автор, редактор или носител на отговорност. Всяко публикувано съдържание остава изцяло отговорност на хората в редакцията на [Factcheck.bg](https://factcheck.bg).

I. Обхват и основни понятия

Настоящият кодекс установява правната и етичната рамки за внедряването на системи с изкуствен интелект (ИИ) в дейността на платформата [Factcheck.bg](https://factcheck.bg), управлявана от Сдружение „Асоциация на европейските журналисти – България“ с ЕИК 176007906 (наричана по-долу АЕЖ).

Настоящият правилник се прилага за всички членове на редакцията на Factcheck.bg, включително щатни журналисти, редактори, технически сътрудници, автори на свободна практика (фрийлансъри) и партньори, които работят по проекти на „Асоциация на европейските журналисти-България“.

Ключови понятия

Понятие	Определение
Изкуствен интелект (ИИ)	Система с изкуствен интелект: машинно базирана система, която е проектирана да работи с различни нива на автономност и която може да прояви адаптивност след внедряването си, и която, с явна или подразбираща се цел, въз основа на въведените в нея входящи данни, извежда начина на генериране на резултати като прогнози, съдържание, препоръки или решения, които могат да окажат влияние върху физическа или виртуална среда.
Генеративен ИИ (ГИИ)	Специфичен вид ИИ, който не само анализира, но и създава ново съдържание - текстове, изображения, аудио и видео. <i>Примери: Copilot, Claude, ChatGPT, DALL-E, Gemini, Midjourney.</i>
Одобрен инструмент	ИИ инструмент, изрично разрешен от редакционното ръководство за употреба в конкретни задачи.
Пилотен проект	Контролиран експеримент с ИИ инструменти в рамките на дефиниран обхват, с изрично одобрение и задължителна оценка.

Съществено влияние	Случай, при който ИИ е изиграл значима роля в създаването, редактирането или разпространението на публикувано съдържание.
Член на екипа	Всеки служител на трудов договор или сътрудник на граждански договор, работещ в рамките на проекта на АЕЖ-България - Factcheck.bg .
Грамотност в областта на ИИ	Набор от умения и знания, позволяващи на служителите да вземат информирани решения за използването на ИИ и да разпознават неговите ограничения.
Партньорски организации на АЕЖ	Юридически лица, които от свое име, под свой контрол и изцяло на своя отговорност създават и разпространяват сред обществото медийно съдържание чрез свои собствени медийни канали, като – но не само – печатни и електронни медии, аналогово и/или цифрово телевизионно и радиоразпръскване, и които се намират в договорни отношения с АЕЖ, съгласно които договорни отношения АЕЖ по редакционен, творчески, организационен, финансов или какъвто и да е друг начин участва в създаването на тяхното медийно съдържание.

II. Основни принципи

Всяка употреба на ИИ във [Factcheck.bg](https://factcheck.bg) трябва да бъде съобразена с **деветте основни принципа на Кодекса**. Те са обвързващи и са в основата на всички правила в следващите раздели.

Принцип 1: Човешки контрол и отговорност

Всички редакционни решения се вземат от хора. **ИИ не замества журналистическата и редакторската преценка**. Всяко публикувано съдържание, независимо дали ИИ е участвал в създаването му, е изцяло отговорност на журналиста/ите и редакцията.

Принцип 2: Проверка преди публикуване

Всеки резултат, генериран от ИИ инструмент, се третира като непроверен материал по подразбиране. Той трябва да бъде верифициран от журналист по

стандартните редакционни процедури преди да бъде използван или публикуван.

- ИИ **не** е надежден или авторитетен източник на информация;
- Резултатите от ИИ подлежат на същите стандарти за проверка, приложими към всеки друг непроверен материал;
- При съмнение в достоверността на дадена информация и липса на доказателства в нейна подкрепа тя не се публикува;
- В процеса на работата си по създаване на медийно съдържание, Членовете на екипа спазват протокол за Етично и безопасно използване на системи с ИИ. Протоколът е даден като **Приложение №2** на настоящия Кодекс;
- Всеки член на екипа е длъжен да уведоми редакторите как и за какво е използвал ИИ в работата си.

Принцип 3: Прозрачност към аудиторията

Когато ИИ е оказал съществено влияние върху създаването или разпространението на публикувано съдържание, това се декларира ясно и разбираемо за читателите.

Деклариране се изисква при:

- Съдържание, създадено основно или изцяло с помощта на генеративен ИИ;
- Преводи, извършени с ИИ инструменти;
- Адаптации за различни аудитории, при които ИИ е изпълнявал съществена роля;
- Всеки случай, при който ИИ има значително влияние върху тона, структурата или съдържанието на публикацията.

Factcheck.bg се ангажира да предоставя при поискване информация за вида и целта на използване на системи с ИИ.

Примерна формула за деклариране

Тази публикация е написана и проверена от журналист на [Factcheck.bg](https://factcheck.bg). За [посочете задачата – превод / адаптация на стил / резюме] е използван ИИ инструмент. Съдържанието е преминало през проверка от журналист и редактор.

Принцип 4: Защита на поверителна информация и на източници

Забранено е въвеждането на поверителна, чувствителна или позволяваща идентификацията на лични данни информация в инструменти с изкуствен интелект.

Това включва:

- Данни, свързани с идентичността на източници;
- Поверителни документи, изтекли материали, правни документи;
- Лични данни на трети лица;
- Записи и транскрипции от интервюта с потенциален риск за сигурността на източниците;
- Непублична информация, свързана с текущи разследвания.

Членовете на екипа следва да имат предвид, че при безплатните нива на ИИ инструменти правилата за използване на данни могат да са по-гъвкави, дори когато инструментът твърди, че не използва входните данни за обучение. Затова всеки въведен текст следва да се третира като потенциално публичен.

Принцип 5: Автентичност и забрана за манипулация

- [Factcheck.bg](#) **не** манипулира снимки, видео или аудио материали с генеративен ИИ, не добавя и не премахва елементи от реални записи чрез ИИ инструменти;
- [Factcheck.bg](#) **не** публикува съдържание с неустановена автентичност;
- Синтетичното съдържание (изображения, видео, аудио, генерирани от ИИ), е ясно обозначено, когато е допустимо за публикуване.

Принцип 6: Само одобрени инструменти

Членовете на екипа и използват **само** ИИ инструменти, изрично одобрени от редакционното ръководство за конкретни задачи. Нови инструменти се тестват и одобряват преди широко внедряване в работата на редакцията.

Принцип 7: Без директна публикация на ГИИ съдържание

- [Factcheck.bg](#) **не** публикува съдържание, създадено директно от генеративен ИИ, без съществена човешка намеса. Това правило важи за текстове, изображения, видео и аудио материали;
- Изключения са допустими единствено в рамките на обозначени пилотни проекти;
- Дори при пилотни проекти публикуваното съдържание подлежи на задължително деклариране и редакторски преглед.

Принцип 8: Справедливост и защита от дискриминация

При употребата на ИИ редакцията полага усилия за идентифициране и минимизиране на алгоритмична пристрастност (**algorithmic bias**). ИИ инструментите не трябва да водят до дискриминация по пол, етническа принадлежност, политическа ориентация или друг признак в съответствие с правото на ЕС и българското законодателство.

Принцип 9: Принцип на Обществено и екологично благосъстояние

- Системите с ИИ се използват по устойчив и екологосъобразен начин, както и така, че да са от полза за всички хора, като същевременно се наблюдават и оценяват дългосрочните въздействия върху отделните лица, обществото и демокрацията.
- Приоритет е използването на ИИ за обработка на данни и съдържание, която би била невъзможна или значително по-сложна без участието на ИИ.
- АЕЖ и Партньорските организации предпочитат използването на оптимизирани, „леки“ ИИ модели пред енергоемки гигантски системи, когато задачата го позволява.

III. Възможни употреби на ИИ

Следните задачи могат да включват употреба на ИИ инструменти, при спазване на общите принципи и конкретните условия, описани по-долу.

3.1. Проучване и подготовка

- Търсене и обобщаване на публично достъпна информация за начална ориентация по дадена тема;
- Генериране на предложения за въпроси за интервюта;
- Обобщаване на дълги документи и доклади за вътрешна употреба;
- Анализ на бази данни и голям масив от публично достъпни данни.

Важно: Резултатите от тези задачи не се публикуват директно. Те служат само за вътрешна работа и изискват журналистическа проверка.

3.2. Транскрипция и превод

- Транскрипция на аудио и видео материали - резултатите се проверяват спрямо оригиналния запис;
- Превод на собствени материали - задължителен преглед от журналист, владеещ съответния език;
- Превод не се препоръчва за спешно съдържание или чувствителни материали.

3.3. Редакционна поддръжка

- Предложения за заглавия – задължителни преглед и одобрение от редактор;
- Корекция на стил и правопис при запазване на авторския текст като основа;
- Генериране на резюмета на материали, написани от журналист - с редакционен преглед преди публикуване.

3.4. Мониторинг и анализ на дезинформация

- Използване на ИИ инструменти за наблюдение и разпознаване на циркулиращи дезинформационни наративи;
- Анализ на разпространение и мрежи за дезинформация;
- Детекция на потенциално манипулирано аудиовизуално съдържание - резултатите се използват като индикация, не като окончателно заключение.

3.5. Технически и административни задачи

- Разработка на код и автоматизация на ниско-рискови административни процеси;
- Обработка и анализ на метаданни;
- Подпомагане при създаване на образователни и комуникационни материали (с редакционен преглед).

3.6. Верификация на съдържание и дигитално разследване

ИИ инструментите са съществена помощ при проверката на съдържание и дигиталното разследване. Следните употреби са разрешени при задължителна редакторска преценка. **Резултатите от тези инструменти са индикативни, а не окончателни и никога не се публикуват като самостоятелно доказателство без допълнителна проверка.**

Разпознаване на синтетично и манипулирано съдържание

- **Разпознаване на изображения, генерирани с ИИ.** Инструменти за детекция на синтетични снимки (напр. Hive Moderation, AI or Not, Illuminarty); резултатите изискват кръстосана проверка с поне един допълнителен метод;
- **Разпознаване на deepfake/cheapfake видео.** Инструменти за анализ на манипулирано видео съдържание (напр. Microsoft Video Authenticator, Sensity, FakeCatcher); поради бързото развитие на технологията резултатите се третират с повишена критичност;
- **Разпознаване на звук и глас, генерирани с ИИ.** Инструменти за детекция на синтетично аудио (напр. ElevenLabs Speech Classifier, Resemble Detect, AI Voice Detector); особено важно при верификация на аудиозаписи с твърдения от публични личности.

Верификация на произход и локация

- **Обратно търсене на изображения.** За установяване на произход и контекст на снимки (Google Lens, TinEye, Yandex Images); **задължително** при всяко изображение от непотвърден източник;
- **Геолокация и верификация на локации.** ИИ-асистирани инструменти за определяне на географски контекст на изображения и видеа (напр. GeoSpy, GeoGuessr за верификация, Google Earth, SunCalc за анализ на сенки и осветление); резултатите се потвърждават с допълнителни визуални и картографски данни;
- **Анализ на метаданни на файлове.** Извличане и интерпретация на EXIF данни и технически метаданни от изображения, видео и документи;

метаданните могат да бъдат изтрети или манипулирани и не са достатъчно доказателство сами по себе си.

Анализ на мрежи, профили и разпространение на дезинформация

- **Анализ на мрежи и профили в социалните мрежи.** Инструменти за установяване на координирано неавтентично поведение, ботове и дезинформационни мрежи; данните от анализа се третират **като журналистическа улика, не като окончателен извод**;
- **Транскрипция и анализ на видео съдържание за верификация.** Автоматично извличане на текст от видеа и аудиозаписи с цел проверка на цитати и твърдения; резултатът се проверява спрямо оригиналния запис;
- **Автоматизирано търсене в архиви и бази данни.** ИИ-асистирано претърсване на публично достъпни архиви, регистри и бази данни за подкрепа на фактчекинг разследвания; използва се само за предварителен анализ, не като краен извод.

Визуална комуникация и достъпност

- **Създаване на субтитри за видео съдържание.** Автоматично генериране на субтитри с последваща редакторска проверка; задължителна е корекцията на имена, термини и цитати.
- **Създаване на дизайни и анимации за илюстративни цели.** ИИ-асистирано визуално съдържание за онагледяване на теми, не за информационни цели; **задължително обозначаване като генерирано с ИИ.**

IV. Правила за деклариране

Прозрачността пред аудиторията е ключова за редакционния интегритет на [Factcheck.bg](https://factcheck.bg). Следната таблица описва кога и как се декларира употребата на ИИ.

Случай	Изискване
Съдържание, създадено в съществена степен с помощ от ИИ	Задължително деклариране с обяснение как е използван ИИ и как е верифицирано съдържанието.
Превод в съществена степен с ИИ инструмент	Деклариране: Преведено с помощта на ИИ и проверено от журналист.
Адаптация в съществена степен на стил/тон за различна аудитория (пилотни проекти)	Деклариране в публикацията; обяснение на целта на адаптацията.
ИИ-генерирана илюстрация (когато е тема на материала)	Ясно обозначаване в надписа/caption-a.
ИИ за помощни задачи (транскрипция за вътрешна употреба, проверка на правопис)	Не се изисква публично деклариране.
ИИ за мониторинг и сигнализиране	Не се изисква публично деклариране, но се документира вътрешно.

Формула за деклариране (по модела на „БиБиСи“)

При деклариране се препоръчва следната структура:

- Вид на съдържанието;
- Роля на журналиста/редактора (включително надзор и верификация);
- Роля на ИИ инструмента (каква задача е изпълнявал);
- Връзка към допълнителна информация (ако е приложимо).

Примерна формула за деклариране

Тази статия е написана от журналист на [Factcheck.bg](https://factcheck.bg). За превода от английски е използван ИИ инструмент; преводът е проверен за точност от журналист, владеещ съответния език. Научете повече за политиката ни за употреба на ИИ на [\[линк\]](#).

V. Пилотни проекти и контролирано експериментиране

[Factcheck.bg](#) е отворена за иновации и контролирано експериментиране с нови ИИ инструменти, при спазване на следните условия:

Условия за стартиране на пилотен проект

- Изрично одобрение от редакционното ръководство преди стартиране;
- Ясно дефиниран обхват, цели и времева рамка;
- Определен отговорен редактор или координатор;
- Задължителен редакционен преглед на всяко публикувано съдържание в рамките на проекта;
- Деклариране на употребата на ИИ при всяка публикация в рамките на проекта;
- Оценка на резултатите и документиране на научените уроци.

Управление на риска при пилотни проекти

Ниво на риск	Мерки за управление
Неприемлив риск	Забранено: всяка употреба на ИИ, която може да подкопае доверието в редакцията или да разпространи дезинформация.
Висок риск	Пристрастност на алгоритмите; заплаха за защита на източници; размиване на границата реално/синтетично → изисква одобрение от ръководството, правен преглед, психологическа оценка.
Приемлив/ограничен риск	Халюцинации и грешки в ранните проучвания → изисква задължителна проверка от редактор преди публикуване.

VI. Корекции, актуализации и обучение

8.1. Политика за корекции

При установяване на грешка, произтичаща от употребата на ИИ инструмент в публикувано съдържание:

- Грешката се коригира незабавно и с видимост, съизмерима с тази на оригиналната публикация;
- Коригираната версия ясно посочва извършената корекция;
- Инцидентът се документира вътрешно за бъдещо обучение.

8.2. Актуализация на Кодекса

Настоящият кодекс е жив документ. Той ще се преглежда и актуализира периодично в отговор на:

- Технологични промени и нови ИИ инструменти;
- Нови регулаторни изисквания (AI Act, GDPR и др.);
- Промени в стандартите на EFCSN и международните практики;
- Научени уроци от пилотни проекти и редакционната практика.

8.3. Обучение на екипа

Редакцията провежда редовни обучения за всички членове на екипа във връзка с:

- Възможностите и ограниченията на ИИ инструментите;
- Прилагането на настоящия кодекс в ежедневната работа;
- Идентифициране и минимизиране на алгоритмична пристрастност;
- Нови инструменти и практики в областта на ИИ в журналистиката.

IX. Ресурси и нормативна база

За неуредените от този кодекс въпроси се прилагат разпоредбите на българското и европейското законодателство. Настоящият кодекс е разработен на базата на следните документи и стандарти:

Организация / Документ	Основен принос
AFP - Editorial Standards (февруари 2025 г.)	Забрана за директна публикация на ГИИ; защита на поверителни данни; само одобрени инструменти
Associated Press - Generative AI Standards	Непроверени по подразбиране резултати; ограничено експериментиране с верификация
Reuters - Reuters and AI	Деклариране при основна/изключителна употреба на ГИИ
BBC - AI Principles and Transparency	Формула за деклариране; директна/непряка прозрачност; пилотни проекти
Paris Charter on AI and Journalism (RSF)	10 принципа за ИИ в журналистиката; приоритет на човешкия контрол
EFCSN - AI Working Group	Специфични стандарти за фактчек организации; управление на риска
Council of Europe - Guidelines on Responsible AI in News Media	Правова и регулаторна рамка; защита на правата
Chequeado & Factchequeado AI Policies	Практически протоколи; деклариране; забранени употреби
Thomson Reuters Foundation - AI-ready Newsroom Guide	Практически наръчник за редакционна интеграция на ИИ
VALIDAI Project Documentation (2025)	Специфичен контекст за адаптация на проверено съдържание

Приложение №2: Протокол за Етично и безопасно използване на системи с ИИ

Преди да публикувате на какъвто и да медиен материал, създаден с участието на система с ИИ, членът на екипа следва да премине през следните стъпки:

1. ПРОВЕРКА НА ДАННИТЕ

- ☐ **Анонимизирах ли запитването (prompt)?** Уверих се, че в чата с ИИ не присъстват имена на защитени източници, лични телефони, адреси или непублични документи.

- ☐ **Използвам ли одобрен от редакцията инструмент?**

2. ВЕРИФИКАЦИЯ НА ФАКТИТЕ

- ☐ **Проверих ли „фактите“ на ИИ?** Намерих поне един независим, надежден източник (официален документ, държавна агенция, авторитетна медия), който потвърждава извода на алгоритъма.
- ☐ **Проверих ли цитатите и линковете?** Уверих се, че източниците, които ИИ посочва, действително съществуват и казват точно това, което ИИ твърди.
- ☐ **Има ли логически грешки?** Прочетох текста критично за противоречия, които ИИ често допуска при по-дълги анализи.

3. ДЕТЕКЦИЯ НА МАНИПУЛАЦИИ

- ☐ **Използвах ли повече от един инструмент?** (Например ако Hive казва "AI-generated", проверих ли и през InVID или чрез обратен анализ на изображения?)

4. ПРОЗРАЧНОСТ

- ☐ **Маркирах ли съдържанието?** Ако материалът съдържа изображение, генерирано с ИИ, или е преведен от ИИ, поставих ли съответния етикет (Disclaimer)?
- ☐ **Описах ли метода?** В текста е ясно обяснено: „Използвахме [Инструмент X], за да анализираме [Данни Y]“.

5. КРАЙНА ОТГОВОРНОСТ

- ☐ **Готов ли съм да защита този резултат?** Ако утре се окаже, че ИИ ни е подвел, мога ли да докажа, че съм положил необходимите усилия (due diligence) за проверка?
- ☐ **Човешки контрол.** Редакторът прегледа ли финалната версия на материала?

[Factcheck.bg](https://factcheck.bg) | Асоциация на европейските журналисти – България
Версия 1.0 | 2025
www.factcheck.bg