

# Как да разпознавате съдържание, създадено от изкуствен интелект или манипулирано по цифров път

Достъпните и лесни за използване модели изкуствен интелект (ИИ) могат да бъдат полезни в ученето или създаването на съдържание, но могат и да улеснят **дезинформацията** и по този начин да застрашат демократичния дебат. Ето защо е важно информационната среда, която всички използваме, да не бъде претрупана с дезинформация, създадена от ИИ или цифрово манипулирана.

Част от решението са нови технологии, позволяващи да открием първоизточника или детектори - инструменти, засичащи манипулации или съдържание, създадено. **Но технологичните решения далеч не са свършени и ние се нуждаем от работата на независими проверители на факти, за да осигурим на обществото проверена информация.**

Ето кратко обобщение за това как работят професионалните и независими проверители на факти (факт-чекъри), за да идентифицират и опровергават дезинформацията, създавана от ИИ.

## Съдържанието, създадено от ИИ е във възход

Днес създадената от ИИ дезинформация е малък процент от всички твърдения, които независимите проверители на факти проверяват. Съдържанието, манипулирано с помощта на цифровите технологии, е често срещано от факт-чекърите.

Но: във вътрешно проучване сред членовете на EFCSN **повечето проверители на факти се съгласиха, че значението на генерираното от ИИ и цифрово манипулираното съдържание ще нараства.** И последните [примери](#), свързани с европейските избори, потвърждават тази прогноза.

**НАУЧЕТЕ:** Цифрова манипулация е всяко значително изменение на съдържанието спрямо оригинала. За целта могат да се използват и инструменти на ИИ. Това не включва корекции за постигане на по-добро качество.

Създадено с изкуствен интелект : под това се разбира всяка форма на съдържание, създадена с ИИ.



## Технологиите се развиват бързо, но не можем да разчитаме само на тях.

Експертите по ИИ и факт-чекърите са съгласни: Само инструментите за откриване на ИИ (детектори) не са достатъчни, за да се потвърди, че съдържанието е генерирано от изкуствен интелект.

Преди да използвате детектор за ИИ, експертите препоръчват да се запознаете с различни подобни инструменти, но и с такива, които генерират ИИ. Когато разберат как се обучават моделите и с помощта на малко статистика, факт-чекърите могат да започнат да разпознават предимствата и слабостите на различните инструменти.

**Инициативи за удостоверяване на произхода на съдържанието, като например спецификациите C2PA, могат да помогнат, но водният знак и проверката не са безгрешни.**

## Как дезинформацията с ИИ влияе на хората

*“Всеки път, когато следвате инстинкта си, пропускате да помислите.”*

Кристин Дюгоан\*

**ПСИХОЛОГИЯ:** Операциите за оказване на влияние често са разработени така, че да се използват психологическите предразсъдъци.

Ако познавате добре вашите и тези на публиката, ще можете да противодействате по-добре на дезинформацията.

**ЦЕЛИ:** Защо някой, който е недобронамерен би ползвал ИИ, за да създава или разпространява дезинформация? Какво цели?

- Да увеличат влиянието си в друга държава или общност?
- Да не бъдат разкрити или да затруднят проверителите на факти, като създадат много варианти на подобни твърдения?
- Да повлияят, като създадат привидна автентичност чрез мрежи от неавтентични профили?

Кристин Дюгоан е изследовател по въпросите на информационното влияние в Сорбоната.

# Опровергаването изисква многостранен подход и разбиране.

И така, ако инструментите за откриване не работят, какво правим? Важно е да се разбере контекстът на дадено твърдение. Факт-чекърите имат необходимите умения за разследване. Ето няколко съвета,

*“Инструментите за детекция  
никога няма да работят  
стопроцентово”  
-Хенк ван Ес\*\**



**ОБЪРНЕТЕ ВНИМАНИЕ НА ИЗТОЧНИКА:** Можете ли да потвърдите самоличността му? Какви теми публикува и споделя? Кой реагира на това съдържание? Какъв ефект може да има това съдържание върху читателите?



**ПРОВЕРКА ЗА ДОСТОВЕРНОСТ:** Направете независима проверка. Може да проверите чрез надеждни източници като експерти с практически опит в областта. Попитайте ги дали според тяхната експертиза твърдението звучи достоверно.



Използвайте **ИНСТРУМЕНТИТЕ** на онлайн разследващата журналистика като извличане на данни, геолокация, биометрично разпознаване, анализ на модели и др.



**УЧЕТЕ И СЕ АДАПТИРАЙТЕ:** Създателите на дезинформация, генерирана от ИИ се адаптират постоянно. Ето защо и вие трябва да се адаптирате към променящата се среда.

## БЪДЕТЕ ПРОЗРАЧНИ В ПРОВЕРКАТА СИ

Експертите препоръчват, ако опровергавате дадено твърдение, да бъдете прозрачни за това как сте стигнали до извода и да дадете връзки към източниците, на които се позовавате. Това може да помогне на читателите да проследят разследването и да се убедят сами.

В някои случаи разследването е по-важно, отколкото само да докажете, че съдържанието е създадено с ИИ.

\*\* Хенк ван Ес е експерт в областта на OSINT и техниките за проверка на факти.

# Няколко кратки съвета - Какво да правите и какво не

Ето някои подробности, които могат да ви подсказат, че дадено съдържание е създадено от ИИ или е цифрово манипулирано. Заедно с другите съвети, споменати тук (контекст, техники за разследване или детектори), те могат да ви помогнат да разберете какво се крие зад това, което виждате:

## Текст

- Често (но не винаги) **правописът** е по-добър от човешкия.
- Вероятно е стилът да е **сух и сложен**, което е необичайно за социалните мрежи.
- **Прекалено много наречия или прилагателни.**
- Липса на **човешки емоции**, хумор, сарказъм и метафори...
- Може да липсват конкретни детайли (имена, дати, места).
- Най-важното: точни ли са фактите?

## Видео

- Не проверявайте стопкадри от видеото с детектор за изображения, генерирани от ИИ.
- Обърнете внимание на **движението на лицето** и дали движението на устата съвпада с говора.
- Може да забележите **резки прекъсвания.**

## Аудио

- **Сравнете** съмнителния аудиозапис с автентичен образец с помощта на инструменти, които могат да открият разлики в речта и дишането, интонацията...
- Ако използвате **детектор**, то аудиото не трябва да е с ниско качество или със странични звуци.
- Може да забележите неестествен говор, липса на паузи или поемане на въздух.

## Снимки

- Търсете **неестествени детайли**: перфектна кожа, размазан фон, неестествена светлина, както и странности, като например допълнителни пръсти.
- Потърсете **логото** на някой от известните генератори на изображения с ИИ.
- Има ли логика в това, което виждате?
- Когато използвате детектори, изберете снимка с **високо качество**, а не такава, която вече е била споделяна.