

НОВ БЪЛГАРСКИ УНИВЕРСИТЕТ

# BREAK THE FAKE

Наръчник за разпознаване на AI-генерирано съдържание

## ПРАКТИЧЕН НАРЪЧНИК

*Не вярвай на това, което виждаш. Ела да провериш.*

**Автор: Етиен Янев**

Лектор по приложен изкуствен интелект и дигитален маркетинг

*В партньорство с Yettel*

София, 2026 г.

## Съдържание

---

За какво е този наръчник .....	3
Защо точно сега .....	3
Признаци за разпознаване .....	4
<i>Изображения</i> .....	4
<i>Видео</i> .....	5
<i>Аудио</i> .....	5
<i>Текст и фалшиви профили</i> .....	5
30 секунди преди да споделиш .....	6
Какво да правиш в конкретни ситуации .....	7
Инструменти за проверка .....	9
Запомни тези пет правила .....	10

## За какво е този наръчник

Преди три години създаването на убедителна фалшива снимка, видео или глас изискваше експертни умения и скъп софтуер. Днес отнема няколко минути и безплатен инструмент. Това променя из основи правилата на играта в онлайн пространството.

Този наръчник ти дава практически работещи методи как да разпознаеш AI-генерирано или манипулирано съдържание. Без технически жаргон. Без сложни инструменти. Само неща, които можеш да приложиш в следващите 30 секунди, когато попаднеш на нещо съмнително.

### За кого е

- За студенти и преподаватели, които искат да изградят критично отношение към онлайн съдържанието.
- За журналисти и медийни професионалисти, които работят с източници всеки ден.
- За всеки, който използва социални мрежи и иска да не се превърне в неволен разпространител на дезинформация.

### Как да го използваш

Прочети го веднъж от началото до края. След това го запази като референция и се връщай към конкретен раздел, когато попаднеш на съдържание, което ти изглежда странно. Раздел 4 е чеклист, който можеш да ползваш буквално преди да натиснеш „Сподели“.

**⚠ Важно:** Никой признак не е стопроцентово сигурен. AI моделите се подобряват всеки месец и това, което днес е надежден червен флаг, утре може вече да не работи. Затова основното правило е едно: при съмнение, провери преди да повярваш.

## Защо точно сега

Манипулираното съдържание не е ново явление. Фотомонтажи съществуват от поне сто години. Каквото е различно днес е достъпността.

Един човек със смартфон и интернет може за по-малко от пет минути да:

- Генерира фотореалистична снимка на несъществуващо събитие.
- Клонира гласа на друг човек от 30 секунди записан звук.
- Създаде кратко видео с реалистична сцена и герои.
- Произведе хиляди коментари или ревюта в стил, неразличим от човешки.

Това означава, че всяка снимка, всяко видео и всяко текстово съобщение, което виждаш онлайн, теоретично може да е изкуствено създадено. Не съветваме да станеш параноичен. Съветваме да изградиш един прост рефлекс: пауза преди реакция, проверка преди споделяне.

## Кой може да бъде засегнат

Тип риск	Какво се случва
<b>Финансови измами</b>	Обаждане с гласа на роднина, който иска пари спешно. Имейл от „шефа“ за неотложен превод.
<b>Кражба на идентичност</b>	Фалшив профил с твоите снимки в социална мрежа или среща за запознанства.
<b>Дезинформация</b>	Фалшиво видео с „изявление“ на политик или фалшива новина за бедствие.
<b>Репутационни атаки</b>	Поставяне на конкретен човек в компрометираща ситуация чрез deepfake.

## Признаци за разпознаване

Този раздел е сърцето на наръчника. За всеки тип съдържание ще намериш конкретните неща, които AI моделите правят зле. Този списък е валиден към момента на публикуване и подлежи на актуализация.



### Изображения

Снимки на несъществуващи събития, фалшиви документи, портрети на хора, които никога не са съществували. Тук признаците са най-разнообразни и често най-достъпни за разпознаване.

#### На какво да обърнеш внимание

- Ръце и пръсти. Грешен брой пръсти, неестествено усукани, някои по-дебели от други. Това остава най-надеждният признак.
- Текст в изображението. Ако има надпис, табела, заглавие на книга или меню, погледни внимателно. AI генерира букви, които изглеждат като букви, но често не образуват реални думи.
- Симетрия в детайлите. Уши, очила, обици, копчета. Често единият детайл не съответства на другия или се „слива“ с кожата.
- Светлина и сенки. Сенките падат от различни посоки върху едно лице. Очите са осветени от източник, който не съществува в сцената.
- Кожата е твърде перфектна. Без пори, без бръчки, без естествени несъвършенства. AI „идеализира“ повърхностите.
- Фон. Размазани периферии, повтарящи се елементи или хора в дъното, които изглеждат „размити“ по странен начин.

**💡 Бърз тест:** Увеличи изображението на 200% и фокусирай вниманието си върху ръцете, ушите и фона. Ако нещо в детайлите изглежда „неправилно“, дори да не можеш да обясниш точно защо, доверчете се на тази интуиция.


## Видео

Видеото е най-сложното за качествено генериране, но именно там разпознаването е и най-важно, защото видеото има най-силно емоционално въздействие.

РЕАЛНО	AI
<b>15-20</b>	<b>0-2</b>
<b>мигвания/мин</b>	<b>мигвания/мин</b>
<ul style="list-style-type: none"> <li>• Естествени паузи</li> <li>• Дишане се чува</li> <li>• Микродвижения на лицето</li> <li>• Кожа с пори и тон</li> </ul>	<ul style="list-style-type: none"> <li>• Гладка реч без паузи</li> <li>• Без фонов шум</li> <li>• Застинало изражение</li> <li>• Восъчна, перфектна кожа</li> </ul>

### На какво да обърнеш внимание

- Мигането. Героят мига твърде рядко, твърде механично или изобщо не мига. Реалните хора мигат 15 до 20 пъти в минута. Преброй за 30 секунди и сравни.
- Устни и реч. Движенията на устните не съвпадат точно със звука. Често разминаването е минимално, но забележимо в края на думите.
- Несъответствие между глас и изражение. Гласът звучи разтревожен, но лицето е спокойно. Усмивка с равен тон.
- Линията на косата и ушите. При движение на главата, ръбовете на косата и ушите се „разпадат“ или размазват.
- Промяна на обекти между кадрите. Предмети, които героят държи, променят размера, формата или цвета си. Дрехите променят детайли.
- Неестествена физика. Тялото се движи по начин, който нарушава законите на гравитацията. Дрехи, които „следват“ грешни закони.

 **Бърз тест:** Гледай видеото два пъти. Първи път с фокус върху лицето. Втори път с фокус върху обектите и фона. Ако забележиш несъответствия в който и да е от двата прохода, видеото е съмнително.


## Аудио

Гласовото клониране е една от най-бързо развиващите се технологии в момента. Само 30 секунди записан глас са достатъчни за създаване на убедителна имитация.

### На какво да обърнеш внимание

- Липса на естествени паузи. Реалната реч има запъвания, поемане на дъх, „ъм“-ове. AI-генерираният глас често тече прекалено гладко.

- Липса на фонов шум. Реалните записи винаги имат шум от средата: стая, улица, офис. AI глас често звучи като в стерилна студийна среда.
- Плоски емоции. Интонацията не варира естествено според смисъла на казаното.
- Монотонен ритъм. При повторно слушане ритъмът се усеща предсказуем и циклично равен.
- Странно произношение на специфични думи. Имена, термини и редки думи често звучат механично или неправилно ударени.


 **Бърз тест:** Затвори очи и слушай само звука. Ако нещо в гласа те кара да се напрегнеш, без да можеш да обясниш защо, не пренебрегвай това усещане.

## Текст и фалшиви профили

Това е най-зрелият формат за AI генериране. Качеството на текста отдавна надмина прага на разпознаваемост за повечето хора. Затова тук фокусът се измества от самия текст към контекста около него.

### На какво да обърнеш внимание

- Перфектен, но безличен език. Граматически безупречен, но без характерни стилови маркери и личностни нюанси.
- Профилна снимка, твърде перфектна. Изглежда като студийна снимка или като реклама на стокова фотография.
- Нов акаунт с висока активност. Профилът е създаден скоро, но вече има голям брой публикации и последователи.
- Постове в редовни интервали. Публикациите излизат на еднакви часови интервали, сякаш генерирани от планировчик.
- Липса на лични спомени. Профилът няма публикации с конкретни места, реални хора или лични истории.
- Безсмислени взаимодействия. Коментарите под публикациите са общи („чудесно!“, „браво!“) и не показват реално разбиране на съдържанието.

 **Бърз тест:** Кликни върху профилната снимка и я провери чрез Google Reverse Image Search. Ако снимката се появява на много други места с различни имена, профилът е почти сигурно фалшив.

## 30 секунди преди да споделиш

Този чеклист е създаден така, че да можеш да го прилагаш буквално за 30 секунди, преди да натиснеш бутона „Сподели“. Ако дори една от точките те смущава, спри и провери допълнително.

## Бърз чеклист

- Кой е първоначалният източник? Не междинният разпространител, а първоначалният.
- Появява ли се същата информация в поне един друг независим източник?
- Кога е създаден акаунтът или сайтът, който я публикува?
- Информацията е ли „твърде хубава“ или „твърде ужасна“, за да бъде истина?
- Какъв е тонът? Цели ли да предизвика силна емоционална реакция?

⚠ **Правило на петте секунди:** Преди да реагираш емоционално, направи пауза от поне пет секунди. Емоционалната реакция е основният механизъм, по който дезинформацията се разпространява в социалните мрежи.

## Кога си струва да отделиш повече от 30 секунди

Има ситуации, в които едно бързо проверяване не е достатъчно. Отдели повече време, ако съдържанието:

- Касае публична личност с конкретно изявление или действие.
- Описва събитие с конкретни жертви, дати или места.
- Те призовава да направиш нещо: да дариш, да гласуваш, да споделиш, да преведеш пари.
- Може да повлияе на твоето решение в нещо важно: финанси, здраве, гласуване.

## Светофар за решения

Когато прецениш съдържанието, попадаш в една от трите зони. Това е визуално обобщение на всичко по-горе.

### ЗЕЛЕНО — споделяй



- Източникът е известна и доверена медия
- Информацията се потвърждава от поне 2 независими източника
- Снимките минават проверка с Reverse Image Search

### ЖЪЛТО — провери още



- Един източник, без независимо потвърждение
- Силна емоция в заглавието ("шок", "скандал")
- Профилът е нов или с подозрителна активност

### ЧЕРВЕНО — НЕ споделяй



- Видими признаци на манипулация по съдържанието
- Източникът е анонимен или непознат
- Целта е финансово или политическо действие

## Какво да правиш в конкретни ситуации


Признаците са едно. Реалните ситуации, в които ще ти се наложи да ги използваш, са друго. Ето няколко от най-често срещаните и какво да направиш в тях.

### Получаваш обаждане от близък в спешна ситуация

Гласът е разпознаваем. Историята звучи правдоподобно: катастрофа, болница, неотложен превод. Емоцията е силна и натискът за бързо действие е голям.

#### Какво да направиш

1. Не предприемай нищо в момента. Кажи че ще се обадиш отново след минута.
2. Затвори. Обади се на същия човек на номера, който знаеш, че е негов. Ако не отговаря, обади се на друг близък, който може да го намери.
3. Ако ситуацията се потвърди от втори източник, тогава действай.

 **Превенция:** Договори се с близките си за „кодова дума“ за критични ситуации. Дума, която само вие знаете и AI клонинг не би могъл да я познае. Ако обажданият се не я знае, обаждането е фалшиво.

### Виждаш шокираща новина в социалните мрежи

Заглавието е голямо. Снимката е силна. Първата ти реакция е да я споделиш на всички. Но точно в този момент е най-важно да спреш.

#### Какво да направиш

1. Провери кой е публикувал. Истински медиен сайт или непознат акаунт.
2. Намери оригиналната снимка чрез Google Reverse Image Search или TinEye.
3. Провери дали поне още една утвърдена медия е публикувала същата информация.
4. Ако след 60 секунди проверки нищо не потвърждава новината, не я споделяй.

### Получаваш съмнителен имейл или съобщение

Имейл от „шефа“ или от банката с искане за бързо действие. Спешен превод, актуализация на парола, потвърждение на данни.

#### Какво да направиш

1. Не кликай върху никакви линкове в самия имейл.
2. Провери имейл адреса на изпращача. Често е почти същият като оригиналния, но с малка промяна.
3. Ако се изисква действие, обади се на изпращача на познат телефонен номер за потвърждение.

4. За банкови имейли влез в банковото приложение директно, не през линка в имейла.

## Намираш интересен профил в социална мрежа или среща за запознанства

Снимките изглеждат добре. Историята звучи добре. Но нещо в чувството ти ти казва, че може и да не е истинско.

### Какво да направиш

1. Запази една от снимките и я провери чрез Google Reverse Image Search.
2. Виж кога е създаден профилът и колко публикации има.
3. Виж дали приятелите и последователите са истински хора с истински профили.
4. Поискай видеоразговор. Реален човек ще приеме, фалшив профил ще намери извинение.

## Инструменти за проверка

---

Изброените тук инструменти са безплатни и не изискват регистрация (с малки изключения). Препоръчваме да запазиш този списък като отметка в браузъра си.

### За снимки и изображения

- Google Reverse Image Search ([images.google.com](https://images.google.com)). Десен бутон върху снимка → „Търсене с Google Lens“. Показва къде другаде се появява тази снимка.
- TinEye ([tineye.com](https://tineye.com)). Алтернатива на Google, която проследява историята на изображението и неговите модификации.

### За проверка на факти и новини

- AFP Провери ([proveri.afp.com/afp-bulgaria](https://proveri.afp.com/afp-bulgaria)). Български раздел на международната агенция AFP за проверка на факти.
- Factcheck.bg. Български сайт, специализиран в проверка на факти и разобличаване на дезинформация.
- БНР Фактчек ([bnr.bg/factcheck](https://bnr.bg/factcheck)). Проверка на факти от Българското национално радио.

### За AI-генериран текст

- GPTZero ([gptzero.me](https://gptzero.me)). Дава процентна индикация дали даден текст е генериран от AI. Резултатът е ориентир, не присъда.

### За deepfake видео

- Към момента не съществува безплатен и надежден публичен инструмент за откриване на deepfake видео. Технологията за откриване изостава от

технологията за генериране. Затова основното остава да познаваш признаците и да гледаш с критично око.

⚠ **Важно:** Нито един инструмент за откриване не е стопроцентово надежден. Резултатите от автоматичните проверки трябва да се интерпретират като допълнителна информация, а не като окончателна присъда. По-надеждни са структурните проверки: кой е първоначалният източник и потвърждава ли се информацията независимо.

## Запомни тези пет правила

Ако трябва да отнесеш само пет неща от целия този наръчник, нека са следните.

### Правило 1

Манипулацията не изисква умения, а само достъп до съответните AI инструменти за създаване на подвеждащи изображения и видео. Това означава, че всеки може да я приложи и всеки може да стане жертва.

### Правило 2

Никой признак не е стопроцентово сигурен. Затова не разчитай само на един. Гледай за съчетание от няколко признака преди да направиш заключение.

### Правило 3

Емоционалната реакция е сигнал да спреш, не да действаш. Колкото по-силна е емоцията, която съдържа нието предизвиква, толкова по-важно е да направиш пауза.

### Правило 4

Първоначалният източник винаги е по-важен от междинния разпространител. Преди да повярваш или да споделиш, намери откъде идва информацията.

### Правило 5

Ти си последната линия на защита. Алгоритмите на платформите няма да защитят аудиторията ти от дезинформация. Ти си този, който решава какво да сподели нататък.

*Не вярвай на всичко, което виждаш. Критичното мислене е твоята защита.*

### Откъде идва този наръчник

Наръчникът е създаден на базата на лекцията „Break the Fake“, проведена в Нов български университет в партньорство с Yettel. По време на лекцията студенти от НБУ участваха в практически експеримент: гледаха AI-генерирано видео и без предварителна подготовка идентифицираха над 30 различни признака, че видеото е изкуствено създадено. Голяма част от наблюденията в раздел „Признаци за разпознаване“ са формулирани директно на базата на техните открития.

Изводът от експеримента е окуражителен: критичното мислене не е специализирано умение, а тренируем навик. Всеки може да го развие за минути на насочено внимание.

## Наръчник изготвен от Етиен Янев

За Нов български университет, в партньорство с Yettel

### Break the Fake.

## Приложение



### КАК ДА РАЗОБЛИЧИМ AI СЪДЪРЖАНИЕТО?

**КАКВО Е ФЕЙК?**  
Фейк съдържание е изображение, видео или текст, създадени или манипулирани с помощта на изкуствен интелект, които изглеждат реалистични, но всъщност не са истински.

**КАК СЕ РАЗОБЛИЧАВА?**

- ✓ Проверявай детайлите – необичайни неща като допълнителни пръсти, размивки, деформирани обекти.
- ✓ Внимавай с текстовете в изображението – често има правописни грешки или неестествен шрифт.
- ✓ Използвай обратното търсене на изображения (Google Images, TinEye), за да видиш дали снимката съществува и другаде.
- ✓ Провери източника – достоверни ли са профилът, сайтът или авторът?

**⚠ НЕ ВСИЧКО, КОЕТО ИЗГЛЕЖДА ИСТИНСКО, Е ИСТИНСКО.**  
AI технологиите се развиват бързо – информираността е най-добрата защита.

**💡 БЪДИ КРИТИЧЕН, ПРОВЕРЯВАЙ И НЕ ВЯРВАЙ НА ВСИЧКО, КОЕТО ВИЖДАШ В ИНТЕРНЕТ!**