unicef | за всяко дете

КРАТЪК АНАЛИТИЧЕН ДОКЛАД – ФЕВРУАРИ 2026 г.

Изкуственият интелект и сексуалното насилие и експлоатацията на деца

Този кратък обзор изследва как генеративният изкуствен интелект (ИИ) създава нови заплахи за правото на децата на защита от сексуално насилие и експлоатация. В него се разглеждат нововъзникващите рискове и доказателства и се очертават неотложни приоритети за действие.

ОСНОВНИ ПОСЛАНИЯ

1. Все по-широкото разпространение на инструменти за генериране на изображения или видеоклипове с интегриран изкуствен интелект, които създават материали, свързани със сексуално насилие над деца, бележи значително увеличаване на рисковете за децата посредством дигиталните технологии.
2. Необходими са спешни действия от страна на правителствата и индустрията, за да се предотврати създаването и разпространението на генерирано от ИИ сексуално съдържание с деца.
3. Родителите, педагозите, социалните служби, специалистите по психично здраве и правоприлагащите органи се нуждаят от ресурси и непрекъснато обучение, за да подкрепят засегнатите деца.

Предизвикателството

Възходът на мощните инструменти за генериране на изображения с ИИ трансформира рисковата среда в областта на закрилата на детето. Преди по-малко от 5 години висококачествените генеративни модели бяха до голяма степен патентовани и изискваха значителна изчислителна мощ и експертиза.[1] Днес моделите с отворен код могат да работят на хардуер от потребителски клас, което значително улеснява извършителите да създават съдържание със сексуална злоупотреба.[2]

Дори когато не е замесено пряко „истинско“ дете, вредата за индивида и обществото е реална – подобно съдържание нормализира сексуализацията на деца, може да увеличи търсенето и да превърне в жертва децата, чийто образ е наподобен.[3] То също така усложнява идентифицирането на жертвите, като затруднява определянето на това дали даден образ изобразява истинско дете, което има спешна нужда от помощ. Това може да доведе до разхищение на и без това оскъдните ресурси на правоприлагащите органи и да забави реакциите за защита на децата.[4]

Тези събития бележат сериозно нарастване на рисковете, пред които са изправени децата в рамките на дигиталната среда. Извършителите могат да създават реалистични сексуални изображения на дете без неговото участие или знание, което означава, че правото на закрила на детето вече може да бъде нарушено, без то да е изпратило съобщение или дори да знае, че това се е случило. След това децата се сблъскват с цялата тежест на последствията: срам, заклеймяване, морално осъждане от страна на връстници и възрастни, социална изолация и дългосрочни емоционални вреди.

Какво знаем?

По-голямата достъпност на инструментите за генериране на изображения или видеоклипове с интегриран изкуствен интелект доведе до увеличаване на производството и разпространението на материали, съдържащи сексуално насилие над деца (CSAM). Тези рискове не са хипотетични – те са документирани и нарастват с бързи темпове. Например за един месец британската Фондация за наблюдение на Интернет (Internet Watch Foundation, IWF) откри в рамките на един форум в тъмната мрежа, посветен на материали за сексуално насилие над деца, почти 14 000 изображения, за които се предполага, че са генерирани от изкуствен интелект. Почти една трета от тях бяха потвърдени като криминални. Бяха открити и първите реалистични видеоклипове с ИИ, съдържащи сексуално насилие над деца.[5] IWF също така идентифицира генерирани от ИИ материали, съдържащи сексуално насилие над деца, в популярните платформи.

Освен съдържанието, създадено от пълнолетни извършители, генерираните от ИИ материали, съдържащи сексуално насилие над деца, включват и „дийпфейк“ голи снимки, създадени от връстници, предимно насочени към момичета. В Корея например правоприлагащите органи съобщават за десетократно увеличение на сексуалните престъпления, включващи технологии с ИИ и дийпфейк, в периода 2022 – 2024 г., като по-голямата част от обвиняемите са тийнейджъри.[6] В САЩ проучването на Thorn установява, че 1 от 10 тийнейджъри знае за случаи, в които приятели или съученици са създали синтетични интимни изображения на други деца без тяхното съгласие, използвайки генеративни инструменти с ИИ.[7]

Неотдавнашно широкомащабно проучване на УНИЦЕФ, ЕКПАТ и Интерпол в рамките на проекта „Прекъсване на вредата“ (Disrupting Harm)[8] показа, че в 11 държави най-малко 1,2 милиона деца са съобщили, че през изминалата година техни изображения са били манипулирани в сексуални „дийпфейк“ фалшификати чрез инструменти с ИИ. В някои от страните това се равнява на 1 на 25 деца или еквивалент на едно дете в стандартен училищен клас.

Самите деца вече добре осъзнават тази заплаха: в някои от 11-те държави до две трети от тях са заявили, че се притесняват, че изкуственият интелект може да бъде използван за създаване на фалшиви сексуални изображения, въпреки че нивата на тревога в отделните държави варират значително, което подчертава необходимостта от мерки за осведоменост и закрила, съобразени с контекста.

Използването на инструменти за генериране на изображения и видеоклипове с интегриран ИИ все по-често излиза извън рамките на създаването на материали, съдържащи сексуално насилие над деца. Според доклад на базирания в САЩ Национален център за изчезнали и експлоатирани деца (NCMEC) генеративният изкуствен интелект е свързан с различни форми на сексуална експлоатация на деца, включително склоняване към сексуални действия и изнудване.[9]

В обобщение, тези технологични постижения означават, че тялото, самоличността и репутацията на детето могат да бъдат осквернени дистанционно, невидимо и трайно. Освен това децата могат да бъдат заплашвани, изнудвани или шантажирани с помощта на изображения, които са изфабрикувани, но изглеждат като истински, възприемат се от обществеността като достоверни и могат да ги следват десетилетия наред.

Ключови приоритети за действие

Тази безпрецедентна ситуация поставя нови предизвикателства пред превенцията и образованието, правните рамки, както и пред услугите за реакция и подкрепа. Настоящите усилия за превенция често се съсредоточават върху обучението на децата относно безопасността в интернет и рисковете от създаването или споделянето на сексуални изображения. Те са все така важни, но недостатъчни, когато сексуалното съдържание може да бъде изкуствено генерирано.

- **Родителите и лицата, полагащи грижи**, трябва да бъдат информирани относно сексуалната експлоатация и злоупотреба с помощта на изкуствен интелект и да имат разбирането и капацитета да подкрепят децата, които са засегнати.
- **Училищата**, които са на първа линия, следва да информират учениците за рисковете, свързани с ИИ, като например за съществуването на „дийпфейк“ голи снимки и за инструментите с ИИ за „разголване“/„съблича не“, както и за значителните вреди, които те причиняват на засегнатите. Те следва да насърчават докладването на опасно поведение, нанасящо вреда и да обучават преподавателите да реагират правилно.[10]

- **Социалните служби, специалистите в областта на психичното здраве и правоприлагащите органи** се нуждаят от ресурси и непрекъснато обучение, за да могат да окажат подкрепа на засегнатите деца, като отчитат конкретното психологическо въздействие на материалите за сексуално насилие над деца, генерирани от ИИ, включително когато са създадени от връстници.[11]

- **От съществено значение е, че заплахата от сексуално съдържание, генерирано от изкуствен интелект, изисква спешна законодателна намеса.** Националното наказателно право трябва да бъде актуализирано и прилагано в съответствие с ерата на дигиталните технологии.[12] Това включва гарантиране, че дефинициите за материали, съдържащи сексуално насилие над деца (CSAM), обхващат материали, генерирани с ИИ, привеждане на националните рамки в съответствие с международните стандарти и осигуряване на правна защита на децата жертви. Държавите следва също да гарантират наличието на набор от наказателни, граждански и административни санкции за юридически лица при престъпления, свързани със сексуална експлоатация на деца, както и при нарушения на задълженията за защита на децата от подобни вреди.[13]

- И накрая, **от съществено значение са отчетността и прозрачността на индустрията, както и засилените действия от страна на самата индустрия.** Държавите следва да изискват от компаниите да извършват надлежна проверка за спазване на правата на детето, по-специално последващи оценки на въздействието върху правата на детето.[14] Компаниите следва да включат в редовните си отчети категорично оповестяване на въздействието върху правата на детето.[15] Всеки участник във веригата за създаване на стойност в областта на ИИ, от доставчиците на данни до разработчиците на модели, трябва да внедри принципа за безопасност при проектирането. Това включва предварителни тестове за безопасност на моделите с отворен код, за да се намали злоупотребата или незаконната употреба[16], като се използват насоки от нововъзникващите стандарти.[17] Индустрията също има неизмерима роля за принос към разработването на решения в подкрепа на правата и благосъстоянието на децата в ерата на ИИ, включително чрез принос към междуотраслови инициативи и научноизследователски действия.

Крайни бележки

1 Caoilte Ó Ciardha et al., 'AI Generated Child Sexual Abuse Material -- What's the Harm?,' arXiv:2510.02978, preprint, arXiv, October 3, 2025.

2 Ibid.

3 Ibid.; UN Committee on the Rights of the Child, Guidelines regarding the implementation of the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography, CRC/C/156, 10 September 2019, para. 63.

4 Caoilte Ó Ciardha et al., p. 17.

5 IWF, '2024 Update: Understanding the Rapid Evolution of AI-Generated Child Abuse Imagery,' 2024; IWF, 'What has changed in the AI CSAM landscape?,' July 2024; IWF, 'AI becoming 'child sexual abuse machine' adding to 'dangerous' record levels of online abuse, IWF warns,' 16 January 2026.

6 'Deepfake, other digital sex crimes by teens nearly double in South Korea in four years,' The Straits Times, 15 September 2025.

7 Thorn, 'REPORT: 1 in 10 Minors Say Peers Have Used AI to Generate Nudes of Other Kids,' 14 August 2024.

8 „Прекъсване на вредата" е многонационален изследователски проект за сексуалното насилие над деца, улеснено от технологиите, който се провежда в 25 държави от УНИЦЕФ, ECPAT International и Интерпол в периода 2019 – 2026 г. с финансиране от Safe Online. През 2025 г. на приблизително 11 000 деца, живеещи в 11 държави, бяха зададени въпроси относно сексуалното насилие над деца, генерирано от изкуствен интелект, чрез национално представително проучване сред домакинствата. Повече данни и констатации от „Прекъсване на вредата" ще бъдат публикувани през 2026 г. тук: https://safeonline.global/ disrupting-harm-2/

9 NCMEC, 'Spike in online crimes against children a "wake-up call",' 4 September 2025.

10 Вижте например eSafety Commissioner (Australia), 'Deepfake damage in schools: How AI-generated abuse is disrupting students, families and school communities', 27 June 2025. See also: Grossman, S., Pfefferkorn, R., & Liu, S, 'AI-Generated Child Sexual Abuse Material: Insights from Educators, Platforms, Law Enforcement, Legislators, and Victims', Version 1, Stanford Digital Repository, 2025.

11 Вижте например Internet Watch Foundation and CEOP Education (UK), 'Child sexual abuse material generated by artificial intelligence: An essential guide for professionals who work with children and young people', Internet Watch Foundation, 2025.

12 United Nations Children's Fund (2022) 'Legislating for the digital age: Global guide on improving legislative frameworks to protect children from online sexual exploitation and abuse' UNICEF, New York (updated version to be released in late 2026).

13 Ibid., pp. 106-107.

14 UN Committee on the Rights of the Child, General comment No. 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, 2 March 2021, para. 38. UNICEF has developed a toolbox for businesses with step-by-step guidance on conducting child rights impact assessments in relation to the digital environment. See UNICEF, D-CRIA Toolbox, 2025.

15 UNICEF, 'Corporate reporting on child rights in relation to the digital environment: Disclosure recommendations and guidance for business', 2025.

16 Вижте още: UNICEF Innocenti – Global Office of Research and Foresight, UNICEF Guidance on AI and Children 3.0, UNICEF Innocenti, Florence, December 2025, p. 18.

17 Thorn, 'Safety by Design for Generative AI: Preventing Child Sexual Abuse,' 2024.

Благодарности

Настоящият кратък обзор е изготвен от Афруз Кавиани Джонсън, дирекция „Глобални програми" на УНИЦЕФ, и Даниъл Кардфелд Уинтър и Аниша Баргава от Центъра за стратегии и доказателства на УНИЦЕФ – Иноченти. Изказваме благодарност на Жозиан Галеа Барон от дирекция „Глобални програми" на УНИЦЕФ за предоставените данни по проекта.

Отказ от отговорност: Това е работен документ. Той е изготвен, за да улесни обмена на знания и да предизвика обсъждане. Текстът не е редактиран в съответствие с официалните стандарти за публикуване и УНИЦЕФ не поема отговорност за грешки. Твърденията в тази публикация представляват вижданията на автора(ите) и не отразяват непременно политиките или вижданията на УНИЦЕФ. Географските обозначения в тази публикация не съдържат по подразбиране становища относно правния статут на която и да било държава или територия, нейните власти или установените държавни граници.

За нас

УНИЦЕФ, агенцията на Организацията на обединените нации за децата, работи за закрила на правата на всяко дете, навсякъде, особено за децата в най-неравностойно положение и в най-трудните за достигане места по света. В над 190 държави и територии по света правим всичко необходимо, за да помогнем на децата да оцелеят, да се развиват и да разгърнат потенциала си.

Публикувано от:

Отдел „Закрила на детето и миграция“ на УНИЦЕФ, дирекция „Глобални програми“ и Център за стратегии и доказателства на УНИЦЕФ – Иноченти.

Имейл: childprotection@unicef.org | innocenti@unicef.org

Уебсайт: unicef.org | unicef.org/protection | unicef.org/innocenti | data.unicef.org

Социални мрежи: @UNICEF | @UNICEFInnocenti в Bluesky, Instagram, LinkedIn и YouTube | UNICEF-Child Protection в LinkedIn

Препоръчителен начин на цитиране

United Nations Children’s Fund, Artificial Intelligence and Child Sexual Abuse and Exploitation, UNICEF, February 2026.

© Детски фонд на ООН (УНИЦЕФ), [февруари] [2026 г.] С изключение на логото на УНИЦЕФ, снимките и местата, където е посочен друг знак ©, тази публикация е лицензирана под CC BY 4.0. Логото на УНИЦЕФ и приложените снимки могат да се използват единствено като част от точно възпроизведени копия на публикацията. Когато е посочен друг знак ©, трябва да поискате от посочения носител на авторските права разрешение да използвате съответния материал извън точно възпроизведени копия на публикацията.

unicef
за всяко дете