

СИНТЕТИЧНО МЕДИЙНО СЪДЪРЖАНИЕ И ПРОМЕНЯЩАТА СЕ КОМУНИКАЦИОННА СРЕДА

Доц. д-р Стойко Петков

Резюме: В публикацията са разгледани бързо развиващите се възможности и все по-голямото присъствие на системи, базирани на изкуствен интелект (ИИ), чрез които се създава синтетично медийно съдържание. Въпреки че много творци използват възможностите за генериране на синтетично медийно съдържание за законна употреба, в същото време се наблюдава увеличение на публикуваните манипулативни и подвеждащи материали, предназначени за измама, изнудване или с други неетични цели. Изкуствено създаденото съдържание, от една страна, е полезно при проекти, в които се използва за възстановяване на гласове или липсваща информация, а от друга, е опасно, когато се използва за подмяна на обективна реалност или за разпространението на дезинформация.

Ключови думи: синтетично медийно съдържание, синтетични медии, изкуствен интелект, дигитални комуникации.

SYNTHETIC MEDIA AND THE CHANGING COMMUNICATION ENVIRONMENT

Assoc. Prof. Stoyko Petkov, PhD

e-mail: spetkov@nbu.bg

Summary: The article discusses the rapidly evolving capabilities and growing presence of Artificial Intelligence (AI) based systems through which synthetic media content is created. Although many organizations use the ability to generate synthetic media content for legitimate use, at the same time, there has been

<https://doi.org/10.33919/ydmc.19.1.5>

an increase in published manipulative and misleading media content intended for fraud, extortion or other unethical purposes. Artificially created content is useful, on the one hand, for projects in which it is used for voice recovery or missing information, and on the other hand it is dangerous when it is used to replace objective reality or to spread disinformation.

Key words: syntetic media, synthetic media content, artificial intelligence, digital communications

Синтетичното медийно съдържание (в някои публикации – синтетични медии) е термин за създадени или модифицирани от алгоритъм видео, графични изображения, текст и аудио. Въпреки че възможността за създаване на синтетично съдържание и синтетични персонажи съществува вече десетилетия, едва през последните няколко години тези технологии са достъпни за масовите потребители/автори на дигитално съдържание и това неминуемо оказва влияние на комуникационната среда.

Доскоро синтетичното медийно съдържание се създаваше само от висококвалифицирани програмисти, монтажисти и творци на специални ефекти, но новите технологии „демократизираха“ достъпа на масовия потребител до този висок клас програмни продукти. Днес почти всеки може да използва специализирани инструменти и софтуер за промяна на изображения, аудио или видео. В отговорни ръце тази технология може да бъде използвана за създаване на произведения на изкуството, безпроблемно превеждане на аудио и видео съдържание на различни езици или дори за създаване на реалистични, персонализирани аватари в цифрови пространства. В неправилни ръце синтетичните медии биха могли да причинят дълбоки разделения в обществото и в държавното управление, тъй като става все по-трудно да се разбере, кое е реално и кое не е. С известно приближение, може да се каже, че живеем в епоха, в която на практика не на всичко, което виждаме или чуваме може да се вярва.

Появата на синтетично медийно съдържание е свързано с дълбочинното обучение – поле в машинното обучение, което обучава компютрите да учат чрез пример. По-конкретно, синтетичните медии разчитат на генеративно-състезателните мрежи (Generative Adversarial Network, GAN). Това са алгоритми за машинно обучение, които работят без учител на база на две невронни мрежи: генератор и дискриминатор. Алгоритъмът на генератора създава модели, а алгоритъмът на дискриминатора определя отнася ли се всеки модел към тренировъчния набор данни или не. Едната мрежа, „генераторът“ се опитва да произведе изкуствени резултати (като изображения), които приличат на реални примери (реални снимки), а другата, „дискриминаторът“ се опитва да различи реалните от фалшивите. Те си помагат една на друга, докато се усъвършенства генераторът така, че да стане опитен фалшификатор.¹ Тук е добре да направим уточнението, че не всички методи за създаване на синтетично съдържание включват машинно обучение. Например традиционните компютърни графични програми също могат да се използват за подправяне на изображения и видео.

Има различни форми на синтетично медийно съдържание, например:

Дийп-фейк (англ. Deepfake, т. нар. „дълбоки лъжи“): подправени видео или аудио записи, които изглеждат и звучат автентично. Те обикновено се създават от злонамерени автори.

Според речника на Министерство на транспорта, информационните технологии и съобщенията, Дийп-фейк е термин, използван за описване на медийно съдържание, произведено чрез изкуствен интелект. Чрез синтезиране на различни елементи от съществуващи видео или аудио файлове, изкуственият интелект позволява сравнително лесни методи за създаване на „ново“ съдържание, в което индивидите изглежда, че говорят думи и изпълняват действия, които не се основават на реалността. Въпреки че все още са в ранна фаза на развитие, вероятно ще видим примери за този вид синтетични медии, използвани по-често в кампании за дезинформация, тъй като тези техники стават все по-сложни.²

„Вирусното“ разпространение на дийп-фейк започва през 2017 година, когато анонимен потребител на новинарския агрегатор Reddit, наричащ себе си „deepfakes“, публикува наложени изображения на актрисите Гал Гадот, Майси Уилямс и Дейзи Ридли върху лицата на жени, участващи в порнографски филми. Т.е.,създателите на съдържанието „подменят“ лицето на участника във видеото.

При случаите на подмяна на гласа и промяна на мимиката на лицето на участник, един от първите примери предизвикал огромен интерес, е кратко видео, в което бившият президент на САЩ, Барак Обама, отправя нападки към Доналд Тръмп. И в този случай е почти невъзможно да се отгатне, че видеото е фалшиво – виждаме, че говори Обама, прави го с интонация, жестове, мимики, като на пръв поглед няма и следа от компютърна намеса.

И ако в споменатите дотук примери имаме недоброжелателно използване на технологиите, то в практиката вече има и достатъчно примери с положителен знак.

В началото на 2019 година британската спортна звезда Дейвид Бекъм бе представен във видео на кампанията „Маларията трябва да умре“³ на организацията Malaria No More. В няколко минутния материал Бекъм отправя призив за набиране на средства на девет различни езика. Движенията на устата и мимиката на лицето са променяни за всеки от езиците и се създава впечатление, че Дейвид Бекъм говори с няколко гласа.

Също чрез изкуствен интелект музеят на Салвадор Дали във Флорида и агенция GS&P създадоха „Дали е жив“⁴. Това е инсталация, която „възкресява“ художника и го превръща в домакин на посетителите. Виртуалният Дали поздравява гостите и посетителите на посветения на личността му музей от големи екрани, на които в цял ръст говори, обяснява и споделя истории за своето изкуство и „мисли“ за съвременните събития.

Виртуални персонажи

При новинарските емисии успешно се експериментира със синтетични говорители, а за целите на ПР вече имаме и синтетични инфлуенсъри.

Хацуне Мику, Лил Микела, Син Сяомен и Чжан Чао. Ако не сте чували тези имена, то вероятно няма да е за дълго. Те са привлекателни, фотогенични, говорят добре и всеки от тях има публика от милиони. Особеното при тях е, че не са истински хора, но и не са механични роботи или холографски изображения. Хацуне е японска поп звезда, на чиито музикални концерти присъстват десетки хиляди фенове. Лил Микела е виртуален моден инфлуенсър с близо 2 милиона последователи в Инстаграм, а синтетичните китайски говорители на новините на информационна агенция „Синхуа“ са Син Сяомен и Чжан Чао.⁵

Разработват се и механизми за персонализация в „общуването“ с тези синтетични персонажи и това е отново чрез генеративно-състезателни мрежи.

Синтетични изображения и дигитални персонажи в киното (напр. CGI – Computer-generated imagery персонажи)

Петко Якимов твърди, че „дигиталните персонажи могат да присъстват в кино разказа като главни герои, спомагателни лица, или като епизодици от масовката. Те могат да бъдат съвършено човешко копие, страдащо от раздвоение на личността и преминаващо през различни трансформации. Могат да бъдат и абстрактни или сюрреалистични образи, излезли от салоните за модерна скулптура или биеналетата на печатната графика.“⁶

Напредъкът при ИИ сега прави процеса на създаване на човешки глас и видео не само много по-бърз, но и с много по-висока степен на правдоподобност.

Според Уилям Осмон някои от новите дигитални възможности за манипулиране на изображение в киното са:

- Разпознаване на лица
- Моделиране на изкуствени светове
- Имитация на реален свят
- Имитация на поведение
- Копиране на филмови звезди от миналото
- Превръщане на текст в изображение
- Превод на филми

- Копиране на стил на режисьор
- Моделиране на реални обекти
- Пресъздаване на исторически събития⁷.

Привържениците на синтетичните медии твърдят, че това е естествено развитие на медиите, което обществото трябва да възприеме. В крайна сметка има много примери за използване на синтетично съдържание за „добро“. Например феновете на „Междувездни войни“ по целия свят бяха доволни, когато разбраха, че в „Междувездни войни: Последният джедай“, са включени кадри с вече починалата актриса Кари Фишър и това е нещо, което би било по-трудно да се направи без най-новите постижения в синтетичните технологии.

Как можете да проверите дали съдържанието е синтетично или не е? В епохата на синтетичните медии и дезинформацията как репортерите и ПР-специалистите могат да са сигурни, че съдържанието, което са получили или разпространяват, всъщност е „истинско“?

Разработват се различни инструменти, включително от български специалисти, които да са в услуга на медиите. Например стартъпът Deertrace разработва инструменти за засичане на фалшифицирани видеа като подобно на други компании, те планират да сътрудничат на новинарски организации и онлайн платформи.

Въпреки че досега няма подход, който да е доказано точен, експертите призовават медийните специалисти да се запознаят с нововъзникващото поле на автоматизирана проверка, където да проверят получената информация. При анализа на изследваното съдържание се използват същите системи за машинно обучение, чрез които се създава синтетично съдържание, и той включва:

- Откриване на „топлинна карта“ на фалшиви пиксели в изображения на лица, създадени с помощта на FaceSwar
- Идентифициране откъде са взети елементите на фалшивото изображение чрез филогенеза на изображението

- Използване на невронни мрежи за откриване на физиологични несъответствия в синтетичното съдържание, например липсата на премигване на очите
- Използване на самите GAN за откриване на фалшиви изображения въз основа на данни за обучение на синтетични видео изображения, създадени с помощта на съществуващи инструменти (като базата данни на FaceForensics или други).⁸

Един от най-новите методи за откриване на подправено съдържание се основава на естественото кръвообращение на хората. Алгоритъмът може да изчисли пулса на човека чрез измерване на честотата на фините промени в цвета на тъканта на лицето на заснет участник. Тази технология ни позволява да измерваме пулса на човек в аудио-визуален материал.⁹ Следователно, чрез алгоритъм можем да засечем променени или генерирани от компютър видеоклипове, тъй като при тях ще липсват тези фини промени.

При една подобна дискусия няма как да не засегнем въпросите за въздействието, управлението и етиката при тези технологии. Най-общо казано, необходимо е внимателно да прилагаме потенциала на ИИ системите при създаването на синтетично медийно съдържание, като същевременно гарантираме, че те не задълбочават съществуващите неравенства и пристрастия или че не създават нови.

Разпространението на дезинформация руши доверието, необходимо за ефективното функциониране на демократичните общества по две причини. Първо, информационният поток е „инфектиран“ с особено опасна форма на манипулация и лъжа. Второ, възможно е обществото да престане да вярва на истински, но неудобни факти.

Според изследователя Авив Овадя синтетичните медии са предизвикателство за способността ни да осмисляме света и да вземаме компетентни решения. Особено тревожно, според него, е нарастването на безразличието към реалността – как хората да разграничават фактическата реалност от подправената реалност и как да избират в какво да вярват, образувайки все по-големи и по-големи клъстери от съмишленици. Овадая твърди, че това

са много по-широки обществени проблеми, но те биха могли да бъдат сериозно повлияни от възможностите за манипулиране на аудио и видео съдържание.¹⁰

Вече имаме достатъчно опит и примери как дийп-фейк материали могат да се превърнат в заплаха за обществото и за сигурността. Това са например случаите, когато троловете се опитват да създадат паника с невярна информация за екологични бедствия, или както при създадената през април 2017 година криза между Саудитска Арабия и Катар, когато хакер проникна в системата на информационната агенция на Катар и публикува подправени и враждебни цитати от емира на Катар.¹¹

Един от примерите от Европа е масова психоза във Франция в началото на 2019 година, предизвикана от фалшиви твърдения в интернет, че роми отвличат момичета с цел проституция и търговия с органи. В резултат на това, от полицията са регистрирани множество нападения срещу роми в няколко североизточни предградия на Париж.¹²

Тъй като медиите, журналистиката и качеството на информацията играят важна роля за насърчаването на здрави демократични общества, доверието в дигиталните платформи, които сами по себе си все повече разчитат на ИИ за подбор на съдържанието, което виждаме, предизвиква загриженост относно ролята, която автономните системи играят за въздействие върху човешката преценка, мнения и решения.

За да избегнат подобни ситуации в бъдеще, много информационни агенции и новинарски сайтове, и изобщо медиите като цяло, започнаха да развиват инструменти, базирани на изкуствен интелект, за да идентифицират фалшиво съдържание. Така тяхното съдържание ще бъде безопасно, достоверно и практично за зрители и читатели. Защо ще се нуждаят от изкуствен интелект? Отговорът е прост – количеството от новини и информация, включително фалшивите, расте с невиджани досега темпове и автоматизираната проверка е единственото възможно решение.

Заклучение

Синтетичните медии имат силата да променят изцяло потребителската медийна среда. Въвеждането на синтетично съдържание ще доведе до дългосрочни последици върху доверието, фактичката реалност и изкуството. Синтетичните медии значително ще ускорят творческата изява, като ще съкратят времето за реализация в етапа между възникването на идеята до публикуването на съдържание.

Една очевидна опасност от създаването на синтетични медии е потенциалът им да разрушат доверието в медиите. Повече от век аудиото и видеото служат за документиране на реалността и като основа на фактичката реалност. Текстът и аудио-визията не само са записали нашата история, но и са информирали и оформяли нашето възприятие за реалността. Какво би станало, ако вече не можем да се доверим на зрението си или на слуха си? Напредъкът на изкуствения интелект може да направи създаването на убедителни, но подправени аудио и видео материали сравнително лесно. Това ще изведе противопоставянето и търсенето на политическо и икономическо надмощие чрез дезинформация на съвсем ново ниво.

Въпреки че можем да се надяваме, че по-голямата част от синтетичното медийно съдържание ще бъде използвано за добро, а не за лошо, истината е, че синтетичните медии са мощно оръжие, което може да бъде изключително опасно, когато е в ръцете на злонамерени хора.

Бележки:

1. Panetta, Kasey. 5 Trends Appear on the Gartner Hype Cycle for Emerging Technologies, 2019, <https://www.gartner.com/smarterwithgartner/5-trends-appear-on-the-gartner-hype-cycle-for-emerging-technologies-2019/> (27.11.2019)
2. Дийп фейк, Министерство на транспорта, информационните технологии и съобщенията, <https://www.mtitc.government.bg/bg/content/deep-fakes-diyp-feyk> (20.11.2019)
3. <https://www.malariamustdie.com/> (26.11.2019)
4. <https://thedali.org/exhibit/dali-lives/> (21.11.2019)

5. Vallor, S. Technology and the virtues: A philosophical guide to a future worth living. New York: Oxford University Press. 2016
6. Якимов, Петко. Дигиталната сценография и персонаж в киното, Автореферат, НБУ 2016, <http://eprints.nbu.bg/3187/1/avtoreferat.pdf> (28.11.2019)
7. Osmon, William. Is AI the Future of Hollywood? Medium, 2019, <https://medium.com/deskree-studio/is-ai-the-future-of-hollywood-9ff5011c4be1> (23.11.2019)
8. Gregory, Sam. Deepfakes and Synthetic Media: What should we fear? What can we do?, Witness, 2018 <https://blog.witness.org/2018/07/deepfakes/> (27.11.2019)
9. Turek, Matt. Media Forensics, (MediFor), DEF. ADV. RES. PROJECTS AGENCY, <https://www.darpa.mil/program/media-forensics> (28.11.2019)
10. Ovadya, Aviv. Deepfake Myths: Common Misconceptions About Synthetic Media, The Alliance for Securing Democracy (ASD), 2019, <https://securingdemocracy.gmfus.org/deepfake-myths-common-misconceptions-about-synthetic-media/> (28.11.2019)
11. Salisbury, Peter. The fake-news hack that nearly started a war this summer was designed for one man: Donald Trump, 2017, Quartz, <https://qz.com/1107023/the-inside-story-of-the-hack-that-nearly-started-another-middle-east-war/> (22.11.2019)
12. Йотов, Деян. Нападения срещу роми заради фалшиви новини във Франция, БНТ, 27.03.2019. <http://news.bnt.bg/bg/a/napadeniya-sreshchu-romi-zaradi-falshivi-novini-vv-frantsiya> (30.11.2019)