ПРОГРАМА НА КУРСА

Финансирано от Европейския съюз. Изразените възгледи и мнения обаче принадлежат изцяло на техния(ите) автор(и) и не отразяват непременно възгледите и мненията на Европейския съюз или на Европейската изпълнителна агенция за образование и култура (EACEA). За тях не носи отговорност нито Европейският съюз, нито EACEA.
Пр. Но: 2025-2-LU01-KA210-VET-000371932

Съфинансирано от Европейския съюз

Deep Tech

Съфинансирано от Европейския съюз

ПРОГРАМА НА КУРСА

Приложно откриване на дийпфейк и аудио-визуална верификация

DeepTech за творческо доверие: Борба с дийпфейковете и насърчаване на етична изкуствена интелигентност в аудиовизуалната индустрия (DeepTech)

Номер на проекта: 2025-2-LU01-KA210-VET-000371932

Финансирано от Европейския съюз. Изразените възгледи и мнения обаче принадлежат изцяло на техния(ите) автор(и) и не отразяват непременно възгледите и мнението на Европейския съюз или на Европейската изпълнителна агенция за образование и култура (EACEA). За тях не носи отговорност нито Европейският съюз, нито EACEA.

Deep Tech

Съфинансирано от Европейския съюз

Име на документа	Учебна програма, дейност 2: Пилотно внедряване и създаване на ресурси за преподаватели
Тип на ревизия	Окончателна
Дата на ревизия	юни 2026 г.
Автори	XU Exponential University of Applied Sciences (Германия), Luxembourg Creative Lab (Люксембург), Budakov Films (България)

Декларация за авторски права:

Този документ е защитен чрез лицензна Creative Commons Attribution-Non-Commercial-Share-Alike 4.0 International. Имате право да:

- Споделяте — копирате и разпространявате материала във всякакъв вид медия или формат
- Адаптирате — ремиксирате, трансформирате и надграждате материала при следните условия:
- Приписване — трябва да посочите подходящ източник, да предоставите линк към лицензна и да посочите дали са направени промени. Можете да го направите по всякакъв разумен начин, но не по начин, който предполага, че лицензодателят ви подкрепя или одобрява вашата употреба.

- Некомерсиално — не можете да използвате материала за комерсиални цели.
- Споделяне при същите условия — ако ремиксирате, трансформирате или надграждате материала, трябва да разпространявате вашите приноси под същия лиценз като оригинала.

Всяко неразрешено използване или възпроизвеждане на съдържанието на този обучителен модул ще се счита за нарушение на закона за авторското право и ще бъде предмет на съдебни действия.

1. Обща информация

Наименование на курса	Приложно откриване на дийпфейк и аудио-визуална верификация
Модул / Единица	Приложно откриване на дийпфейк и аудио-визуална верификация
Институция / Организация	(Да се попълни от изпълнителната организация)
Преподавател(и)	(Да се попълни от изпълнителната организация)
Контакт	(Попълва се от изпълнителната организация)

Работно време / Поддръжка	По предварителна уговорка; поддръжка по имейл по време на курса; насрочени консултации по време на провеждането на курса.
ECTS / Натоварване	0,5–1 ECTS (общо 10–12 часа натоварване за учащия)
Ниво по ЕКР	Ниво 5 по ЕКР
Форма на провеждане	Смесено обучение (хибриден формат – онлайн и на място)
Продължителност	7 сесии (общо 10–12 часа)
Целева група	Медийни професионалисти, журналисти, създатели на съдържание, обучители в професионалното образование и обучение, преподаватели в културния и творческия сектор (КТС), проверяващи факти и студенти във висшето образование, които се нуждаят от приложни умения за откриване и проверка на манипулирано аудиовизуално съдържание. Участниците с основни познания по генеративна изкуствена интелигенция ще извлекат най-голяма полза.

Основна област на компетентност	Умения за работа с и анализ на дийпфейк
Вторична(и) област(и) на компетентност	Умения за защита на данните и управление на личните данни; умения за интегриране в учебната програма
Предпоставки	Силно се препоръчва завършване на Силабус 1 (Основи на генеративната ИИ и грамотност в областта на дийпфейк) или еквивалентни базови знания по генеративна ИИ и концепции за синтетични медии от .

Описание на курса

Този приложен курс предоставя на участниците техническите знания и практическите умения, необходими за откриване на манипулирано аудиовизуално съдържание, оценка на автентичността на дийпфейк и прилагане на професионални инструменти за проверка в контекста на медиите, журналистиката, образованието и творческата индустрия. Участниците проучват техническите механизми, стоящи зад създаването на дийпфейк, усвояват техники за откриване, обхващащи анализ на визуални артефакти, временни несъответствия, методи в честотната област и инструменти, подпомагани от изкуствен интелект, и придобиват практически опит с водещи платформи за верификация, включително Microsoft Video

Authenticator, Reality Defender, Hive Moderation, Sensity AI, Deepware Scanner и InVID. Етичните, правните и медийните перспективи са интегрирани в целия курс.

2. Картографиране на компетенциите

Този курс допринася за следните компетенции, определени в рамката за компетенции „DeepTech for Creative Trust“. Учащите ще развият и демонстрират умения в изброените по-долу области.

Област на компетентност	Роля в този курс	Описание
Умения за работа с Deepfake и анализ	Основна	Приложно откриване, проверка и оценка, основана на доказателства, на синтетични аудиовизуални медии с помощта на професионални инструменти и методологии.
Умения за защита на данните и	Средно	Поверителност, съгласие и правни задължения при работа с манипулирано

управление на поверителността		съдържание, съдържащо лични данни или биометрична информация.
Умения за интегриране в учебната програма	Средно	Разработените тук методики за проверка и етични рамки могат да бъдат интегрирани в образователни и обучителни програми.

3. Общ преглед на курса

Настоящата програма предлага приложно, практическо обучение по откриване на дийпфейк и аудиовизуална верификация, като се основава директно на основните концепции, представени в Програма 1. Учениците развиват практически умения за идентифициране на манипулирано видео и аудио съдържание, разбиране на техническата архитектура зад съвременните дийпфейк системи и прилагане на професионални инструменти, използвани в медийната криминалистика и работните процеси по верификация.

Курсът обхваща целия процес на откриване: от разбирането на това как технически се създават дийпфейковете (размяна на лица, клониране на глас, GAN, дифузионни модели, мултимодална ИИ) до прилагането на анализ на артефакти, техники във времевата и честотната област, инструменти за откриване с помощта на ИИ и методи за аудио верификация. Участниците придобиват практически опит с водещи търговски и отворени платформи за верификация и развиват умения за комуникиране на оценки за автентичност в професионална среда. Етичните, правните и регулаторните аспекти — включително Законът за изкуствения интелект на ЕС, рамките за съгласие и отговорностите на платформите — са интегрирани в целия курс.

4. Резултати от обучението

До края на този курс участниците ще могат:

• Да обяснят как се създават технически съвременните дийпфейкове с помощта на GAN, дифузионни модели, архитектури за размяна на лица, системи за клониране на глас и мултимодални технологии за изкуствен интелект.

• Да откриват манипулирано видеосъдържание чрез прилагане на анализ на визуални артефакти, откриване на временни несъответствия и техники за изследване във честотната област.

• Удостоверяват автентичността на аудио съдържание, като използват методи за откриване на клониране на глас, спектрограмен анализ и подходи за акустично вземане на отпечатъци.

• Прилагат професионални инструменти за проверка — включително Microsoft Video Authenticator, Reality Defender, Hive Moderation, Sensity AI, Deepware Scanner и InVID — за оценка на автентичността на аудиовизуалните медии.

• Оценяване на етичните, правните и социалните последици от дейностите по откриване на дийпфейк, включително правата на неприкосновеност на личния живот, съгласието и разпоредбите на Закона за изкуствения интелект на ЕС.

• Изготвяне на структурирани, основани на доказателства доклади за оценка на автентичността и съобщаване на резултатите от проверката на професионална и широка аудитория.

5. Съдържание на курса

Курсът се провежда в рамките на 7 сесии (общо 10–12 часа) и се състои от 6 тематични модула. Всеки модул съчетава концептуално разбиране, практическо приложение и етична рефлексия.

Модул 1 – Въведение в приложното откриване на дийпфейк

Този модул установява приложния контекст за откриване на дийпфейк чрез разглеждане на еволюцията на дийпфейк технологиите и очертаване на съвременния пейзаж на заплахите, пред които са изправени медийните професионалисти, журналистите и творците. Участниците изследват екосистемата на синтетичните медии: взаимосвързания пейзаж от инструменти за създаване, платформи за разпространение, системи за откриване и регулаторни субекти.

Участниците разглеждат знакови инциденти с дийпфейк в журналистиката, политиката, развлекателната индустрия и личните вреди, като развиват разбиране за реалните последствия от неоткритите синтетични медии. Модулът въвежда основните принципи на откриването и концепцията за работния процес на верификация, която е в основата на всички последващи практически дейности. Етичните съображения включват отговорностите на платформите, задълженията на медиите и задълженията на практикуващите в борбата срещу злоупотребата със синтетични медии.

Модул 2 – Технически основи на създаването на дийпфейк

Дълбокото разбиране на това как технически се създават дийпфейковете е от съществено значение за ефективното им откриване. Този модул разглежда основните архитектури на изкуствения интелект, които стоят в основата на съвременните синтетични медии: Генеративни съпернически мрежи (GAN) и техните варианти (StyleGAN, FaceSwap-GAN); Дифузионни модели (Stable Diffusion, видеосистеми от типа Sora); Процеси за замяна на лица (архитектури енкодер-декодер, откриване на ориентири, смесване); Системи за клониране на глас (невронен TTS, преобразуване на глас); Модели за синхронизация на устните (Wav2Lip, SadTalker); и мултимодални системи за изкуствен интелект, комбиниращи визуална, аудио и текстова синтеза.

Участниците анализират характерните свойства на всеки метод за генериране и научават как те се проявяват като откриваеми артефакти — формирайки техническата основа за всички последващи дейности по откриване на „ “. Етичните съображения засягат двойното предназначение на генеративната ИИ, отговорното разкриване и последствията от демократизираните инструменти за създаване на „deepfake“ за медийните екосистеми.

Модул 3 – Техники за откриване на дийпфейк

Този модул развива практически умения за откриване чрез четири допълващи се аналитични подхода: Анализ на визуални артефакти (неестествени граници на смесване, несъответстваща текстура на кожата, нередовни отражения в очите, аномалии в зъбите, грешки в границите на косата); Откриване на временни несъответствия (трептене между кадрите, неестествени модели на мигане, нередности в микроизраженията, трептене на позата на главата); Анализ на честотната област (DCT артефакти, GAN отпечатъци, спектрални несъответствия с помощта на инструменти, базирани на FFT); и Откриване с помощта на ИИ (невронни класификатори, оценка на достоверността, обяснимост на резултатите от откриването).

Участниците работят с реални и синтетични медийни образци, като прилагат всяка методология за откриване в структурирани лабораторни упражнения. Модулът разглежда ограниченията и потенциалните пристрастия на автоматизираните системи за откриване, като набляга на верификацията, ориентирана към човека, която съчетава технически инструменти с информирана експертна преценка. Етичните съображения включват риска от фалшиви положителни резултати и техните последствия за лицата, репутацията и организациите.

Модул 4 – Аудиоверификация и гласова автентификация

Този модул развива способността на обучаващите се да откриват манипулирано аудио съдържание, като разглежда откриването на клонирани гласове, спектрограмен анализ, акустични отпечатъци и оценка на автентичността на речта. Участниците изследват характерните белези на клонираните с изкуствен интелект гласове и проучват как артефактите на вокодера, неестествената прозодия и спектралните аномалии сочат за синтетична реч.

Практическите упражнения с използване на инструменти за аудиокриминалистика позволяват на обучаващите се да развият практически умения за верификация в реалистичен контекст на медийна продукция. Участниците проучват също така пресечната точка между аудио и визуална верификация в мултимодалните

дийпфейкове, където синхронизацията на устните (lip-), гласът и видеото трябва да се анализират съвместно. Етичните съображения засягат клонирането на глас при измами, несъгласувано представяне за друго лице и работни процеси в творческата индустрия.

Модул 5 – Инструменти за верификация в професионалната практика

Този модул предлага структурирано, практическо обучение с водещите професионални платформи за проверка, използвани в медийната криминалистика, журналистиката и модерацията на съдържание: Microsoft Video Authenticator (криминалистична оценка на достоверността); Reality Defender (мултимоделно ансамблово откриване); Hive Moderation (автоматизирано пресяване на съдържание); Sensity AI (изкуствен интелект за deepfake на корпоративно ниво); Deepware Scanner (откриване на видео с отворен достъп); InVID / WeVerify Plugin (обратно търсене на видео, анализ на метаданни, извличане на ключови кадри); и криминалистични инструменти с отворен код (FaceForensics++, FotoForensics, ExifTool).

Участниците прилагат всеки инструмент към стандартизиран набор от тестови медийни образци и развиват умения за структурирана сравнителна оценка. Модулът разглежда критериите за избор на инструменти, интегрирането им в професионалните работни процеси и разбирането на ограниченията и праговете на достоверност на автоматизираните системи. Практически резултат: доклад за сравнителна оценка на инструменти за верификация.

Модул 6 – Етични, правни и регулаторни аспекти

Този модул поставя откриването на дийпфейк в по-широкия му етичен, правен и регулаторен контекст. Участниците разглеждат правата на неприкосновеност на личния живот и задълженията за съгласие при работа с потенциално манипулирано съдържание; съображенията за авторско право и интелектуална собственост при разследване на дийпфейк; разпоредбите на Закона за ИИ на ЕС, свързани с етикетирането на синтетични медии и ИИ системи с висок риск; отговорностите на платформите и задълженията за модерация на съдържанието; както и подходите за

медийна грамотност, които дават възможност на аудиторията да оценява критично синтетичното съдържание.

Казусите разглеждат знакови съдебни дела, регулаторни интервенции и реакции на платформите на злоупотребата с дийпфейк. Участниците разработват рамки за етично вземане на решения за сложни сценарии на проверка, включващи уязвими лица, политическо съдържание или оспорвани доказателства. Модулът подчертава, че техническото откриване трябва винаги да бъде допълнено от правно съзнание, етична преценка и разбиране на контекста.

6. Методи на преподаване и учене

Този курс възприема приложен подход в стил лаборатория, който се фокусира върху практическа работа с реални медийни образци и професионални инструменти за верификация. Учебните дейности развиват както технически компетенции, така и професионална преценка, подготвяйки участниците за автентични предизвикателства при верификацията в медиите, журналистиката, образованието и творческите индустрии.

Методите на преподаване включват структурирани семинари за инструменти, упражнения за анализ на медии с ръководство, сравнителни оценки на платформи, учене въз основа на конкретни случаи с използване на документирани инциденти с дийпфейк и съвместни разследвателни дейности, симулиращи реални работни процеси по верификация. Участниците работят както с търговски, така и с отворени системи за откриване, развивайки гъвкави компетенции за верификация, независими от конкретния инструмент.

Критичното размишление е вплетено в целия курс, като след всяко практическо упражнение се провеждат структурирани обсъждания, посветени на ограниченията на инструментите, рисковете от фалшиви положителни резултати, етичните дилеми и важността на човешкия надзор в работните процеси по верификация, подпомагани от изкуствен интелект.

Примерна програма на курса

Сесия	Тема / Фокус	Формат	Оценка / Етап	Бележки
1	Въведение в приложението за откриване на фалшиви видеоклипове	Онлайн / На място	Проверка на знанията	Общ преглед на заплахите и концепции за верификация
2	Технически основи: Как се създават дийпфейк	Онлайн / На място	Дейност за размисъл	Анализ на GAN и дифузионния модел
3	Техники за визуално и времево откриване	Онлайн / На място	Практическо упражнение за откриване	Семинар за идентифициране на артефакти
4	Аудио верификация и гласова автентификация	Онлайн / На място	Задача за спектрограмен анализ	Акустичен анализ с помощта на аудио инструменти

5	Инструменти за верификация: практическо упражнение	Онлайн / На място	Доклад за оценка на инструментите	Сравнение и оценка на платформи
6	Етични, правни и регулаторни аспекти	Онлайн / На място	Анализ на казуси	Законът на ЕС за изкуствения интелект и преглед на политиката за платформите
7	Съвместна проверка Разследване и обобщение	Онлайн / На място	Представяне на окончателния проект	Доклад за оценка на автентичността

Стратегия за оценяване

Оценяването е предназначено да подпомогне както ученето, така и демонстрирането на компетентност чрез комбинация от формативни и сумативни дейности:

Тест за проверка на знанията *(формативна оценка)*

Въпроси с избор от няколко отговора, обхващащи концепции за откриване, възможности на инструментите и нормативни рамки от модули 1–3.

Упражнение за анализ на спектрограми и артефакти *(формативно)*

Практически упражнения с ръководство, при които участниците трябва да анотират и обяснят откритите аномалии в предоставените медийни проби.

Доклад за проверка *(сумативна оценка – 40 %)*

Участниците използват най-малко два професионални инструмента, за да анализират три предоставени аудиовизуални проби и да изготвят структуриран доклад за оценка на автентичността с доказателства и обосновка.

Съвместно разследване за верификация *(обобщаващо – 60%)*

Малки групи разследват реален инцидент с дийпфейк, като прилагат техники за откриване, инструменти за верификация и етичен анализ, за да изготвят доклад, основан на доказателства, и групова презентация.

Препоръчителни ресурси и инструменти:

- Microsoft Video Authenticator, Reality Defender, Hive Moderation, Sensity AI, Deepware Scanner.
- Браузър плъгин InVID / WeVerify; FotoForensics; ExifTool.
- Набор от данни FaceForensics++ (с отворен код, за образователни цели).
- Официален текст и насоки на Закона за изкуствения интелект на ЕС.
- Ръководства за проверка и методологии за проверка на факти на First Draft.
- Подбрана библиотека с примери на автентично и синтетично аудиовизуално съдържание за тренировъчни упражнения.

7. Валидиране и оценка на уменията

Общ преглед на сумативната оценка

Компонент на оценяването	Тежест	Описание
Доклад за проверка	40%	Структурирана оценка на автентичността на три аудиовизуални проби, като се използват минимум два професионални инструмента за проверка, с анотирани доказателства и писмена обосновка, обосноваваща заключенията.
Съвместно разследване за проверка	60	Групово разследване на документиран случай на „дипфейк“: прилагане на методи за откриване, инструменти за верификация и етичен анализ; представяне под формата на писмен доклад и групова презентация.

Дейности за формативна оценка (без тежест)

• Тест за проверка на знанията, обхващащ концепции за откриване, възможности на инструментите и регулаторни рамки.

• Упражнения за анотиране на спектрограми и визуални артефакти.

• Участие в групови дискусии, практически упражнения и дейности по взаимна оценка.

• Записи в дневник за размисъл, документиращи оценки на инструменти и опит в откриването.

Критерии за оценяване

Дейностите по оценяване оценяват способността на учащите да:

• Обясняват техническата архитектура и сигнатурите за откриване на съвременните системи за дълбоки фалшификати.

• Прилагат точно и систематично техники за визуално, времево, честотно и аудио откриване.

• Използват компетентно професионални инструменти за верификация и критично интерпретират резултатите от тях.

• Изготвят структурирани оценки на автентичността с ясно обосновано разсъждение.

• Оценяват етични и правни съображения, свързани с дейностите по верификация.

• Да комуникирате ясно резултатите от верификацията както на техническа, така и на нетехническа аудитория.

8. Приобщаване, достъпност и подкрепа за учащите

За да се гарантира приобщаващо учебно преживяване, курсът включва:

• Гъвкави темпове и формати на преподаване (онлайн, на място или хибридно).

• Достъпни учебни материали, използващи ясен език и визуални средства, където е уместно.

• Възможности за различни стилове на учене чрез комбинация от презентации, дискусии, практически упражнения и дейности за сътрудничество.

• Структурирана подкрепа от съученици и насочено фасилитиране, за да се насърчи участието на учащи с различни нива на дигитална грамотност и предварителни знания.

Този курс следва принципите на „Еразъм+“ за приобщаване:

- Равни възможности за участие за всички учащи.
- Адаптиране на материалите, където е необходимо.
- Насърчаване на разнообразни гледни точки и опит.
- Безопасна и уважителна среда за учене.

Механизмите за подкрепа включват консултации с инструктора, сътрудничество между съучениците и допълнителни материали за насоки, предоставяни по време на курса.

9. Очаквания към преподавателя и учащите

Отговорности на преподавателя

- Осигуряване на ясна структура и насоки.
- Дава навременна и конструктивна обратна връзка.
- Да улеснява приобщаващо и уважително учене.
- Поддържане на актуални познания за инструментите за изкуствен интелект и новите разработки.

Отговорности на учащите

- Активно участие във всички учебни дейности.
- Да съобщавате за предизвикателствата навреме.
- Да спазват етичните, академичните и професионалните стандарти.
- Подхождайте към инструментите за изкуствен интелект критично и отговорно.

10. Източници

Техническа литература – Откриване на дийпфейк

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). „Deepfakes и отвъд тях: преглед на манипулациите на лица и откриването на фалшиви изображения“. Information Fusion, 64, 131–148.

Rossler, A., Cozzolino, D., Verdoliva, L., et al. (2019). FaceForensics++: Обучение за откриване на манипулирани изображения на лица. ICCV 2019.

Guera, D., & Delp, E. J. (2018). Разпознаване на Deepfake видеоклипове с помощта на рекуррентни невронни мрежи. AVSS 2018.

Ли, Й., & Лю, С. (2018). Разкриване на DeepFake видеоклипове чрез откриване на артефакти от деформация на лицето. CVPR Workshops.

Аудио криминалистика и откриване на клониране на глас

Yi, J., Bai, Y., Zeng, J., et al. (2022). Half-Truth: Набор от данни за откриване на частично фалшиви аудио записи. Proc. Interspeech.

Todisco, M., Wang, X., Vestman, V., et al. (2019). ASVspoof 2019: Бъдещи хоризонти в откриването на фалшиви и подправени аудио записи. Proc. Interspeech.

Muller, N., Chadha, K., Riedl, P., et al. (2022). Може ли откриването на аудио фалшификати да се обобщи? ICASSP 2022.

Етика, право и регулаторни рамки

Европейски съюз. (2024). Регламент (ЕС) 2024/1689 – Закон за изкуствения интелект. Официален вестник на ЕС.

Чесни, Р., и Ситрон, Д. К. (2019). Дийпфейкове: надвисващо предизвикателство за неприкосновеността на личния живот, демокрацията и националната сигурност. California Law Review, 107(6).

Paris, B., & Donovan, J. (2019). Deepfakes и евтини фалшификати. Изследователски институт „Data & Society“.

ЮНЕСКО. (2021). Препоръка относно етиката на изкуствения интелект. Издателство на ЮНЕСКО.

Инструменти и ресурси за проверка

Microsoft Research. (2020). Microsoft Video Authenticator – Техническа документация.

First Draft. (2022). Ръководства и методология за визуална верификация.

Проект InVID / WeVerify. (2021). Документация за набор от инструменти за проверка. Европейска комисия H2020.

Sensity AI. (2023). Доклад за заплахите от дийпфейк.