EXChange AI · Модул 3: ИИ и обществото

Обучителни бележки за преподаватели · Проект **Expeditions for Change: Navigating the Skills & AI Landscape**

EXCHANGEAI.EU

2024-2-BG01-KA210-ADU-000296153

Съфинансирано от Европейския съюз

Финансирано от Европейския съюз. Изразените възгледи и мнения са единствено на автора(ите) и не отразяват непременно тези на Европейския съюз или на Центъра за развитие на човешките ресурси. Нито Европейският съюз, нито финансиращият орган носят отговорност за тях.

Как да използвате тези бележки

Предназначение

Бележките са написани така, че да могат директно да се копират в полето **Notes** на PowerPoint за съответния слайд.

Структура на сценария

Текстът е в ролята на **преподавателски сценарий**: какво да се обясни, как да се направи преход към темата и как да се активира аудиторията.

- Какво да обясните на участниците
- Как да направите преход към следващата тема
- Как да активирате аудиторията с въпроси

Целта на модула е **реалистичен поглед** върху ИИ: нито реклама, нито страхова картина — а балансирано разбиране на ползите, ограниченията и рисковете.

Слайд 1–2: Въведение и „Правилото на Златокоска"

Открийте модула, като представите темата: не просто какво е изкуствен интелект, а как той вече влияе върху обществото, работата, образованието, доверието и начина, по който вземаме решения.

1

Прекален оптимизъм

ИИ ще реши всички проблеми

2

Балансирана позиция

ИИ е мощен инструмент, но не е магия и не е самостоятелен морален агент

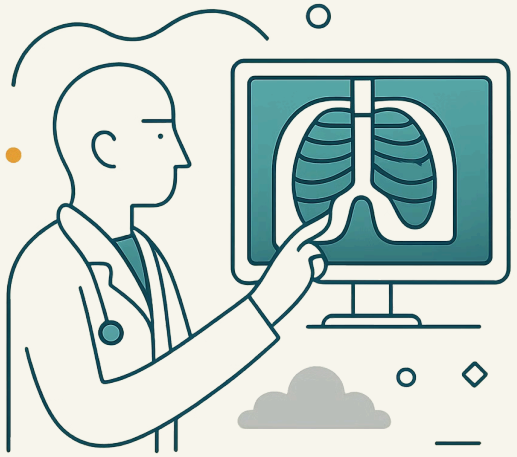
3

Прекален песимизъм

ИИ неизбежно ще доведе до катастрофа

💬 Въпрос към аудиторията: *„Коя от тези три позиции най-често срещате в медиите или сред колегите си?"*

Слайд 3: Ограничения на ИИ системите



Пример от медицинската образна диагностика: системи могат да се обучават да разпознават образи като пневмоторакс на рентгенови снимки.

Две ключови ограничения

Производителност

Системата може да се справя добре в тестови условия, но да греши при различни данни, различни пациенти или различен контекст.

Обяснимост

Понякога е трудно да разберем защо моделът е дал точно този резултат. В чувствителни области не е достатъчно да имаме отговор — важно е да можем да проверим логиката.

⚠ ИИ трябва да **подпомага**, а не безкритично да **замества** професионалната преценка.

Слайд 4–7: Дискриминация и предразсъдъци

ИИ не възниква в празно пространство. Той се обучава върху данни, създадени от хора и общества, а тези данни често съдържат исторически неравенства, стереотипи и непълно представяне на различни групи.

Произход на проблема

Предразсъдъците не са „грешка в компютъра" — те са социален проблем, усилен чрез автоматизация.

Пример: Словесни вектори

Системата може да научи: „мъж — програмист" и „жена — домакиня", защото подобни асоциации са чести в обучаващите текстове.

Реални последици

Подбор на кандидати, разпознаване на лица, кредитиране, оценяване на риск — системна вреда за цели групи.

Решения

Технически мерки + прозрачност + разнообразни екипи + периодични одити + човешка проверка.

💬 Въпрос: *„Кой е представен в данните? Кой липсва? Кой проверява резултатите? Какви последствия има грешката?"*

Слайд 8–12: Враждебни атаки срещу ИИ

Тук не говорим за обикновени грешки, а за **умишлени малки промени** в данните или средата, които могат да подведат модела. За човека промяната може да изглежда почти незабележима, но за системата да доведе до напълно различна класификация.

01

Дигитални атаки

Добавя се малко смущение към изображението и моделът променя отговора си драматично — например от „колибри" към „чук".

02

Физически атаки

Очила, стикери, пластири, промени по знаци — предмети, които объркват системите за разпознаване в реалния свят.

03

Уверен, но погрешен отговор

Системите често не казват „не знам", а дават уверен, но погрешен отговор. Увереността не е доказателство за истинност.

04

Защити и „надпревара"

Обучение с враждебни примери, мониторинг, филтриране. Когато защитите се подобряват, нападателите също се адаптират.

⚠ Сигурността на ИИ не е еднократна настройка, а **процес на непрекъснато тестване, наблюдение и актуализация**.

💬 Въпрос: *„Къде бихме били най-притеснени от подобна уязвимост — в социалните мрежи, в автомобилите, в банките, в училищата?"*

Слайд 13–14: Злоупотреби с ИИ

Разграничение между непреднамерена вреда и **съзнателна злоупотреба**. Мощните инструменти могат да се използват за различни цели — затова ИИ грамотността включва умението да разпознаваме, проверяваме и оценяваме критично.

Дийпфейкове

Видео или аудио, в което човек изглежда, че казва или прави нещо, което никога не е казвал. Подкопава доверието в медии, институции и лична комуникация.

Фалшиви коментари

Масовото генериране на фалшиви коментари може да създаде илюзия за обществено мнение. Автоматизираните измами стават по-убедителни и персонализирани.

Наблюдение чрез ИИ

Разпознаването на лица и анализът на поведение могат да имат полезни приложения, но в потиснически контекст могат да ограничат свободи и права.

Насърчете практики за **медийна грамотност**: проверка на източници, търсене на контекст, разпознаване на манипулативни сигнали и внимателно споделяне на съдържание.

Слайд 15–18: ИИ и работните места

ИИ не е само технологичен въпрос — той има **икономически измерения**. Може да повиши продуктивността и да създаде нови професии, но и да промени структурата на труда и да изисква нови умения.

Автоматизация

До 2030 г. автоматизацията може да засегне стотици милиони работни места. Фокусът е върху **трансформация**, не само заместване.

Задачи под ИИ

Писане на първи вариант, анализ на данни, обслужване на клиенти, обработка на документи — части от работата могат да се автоматизират.

Човешки умения

Критично мислене, етика, комуникация, педагогически дизайн и работа с хора остават ключово важни.

Учене през целия живот

Хората няма да могат да разчитат само на знанията от първоначалното си образование. Нужна е постоянна актуализация на уменията.

💬 Въпрос: *„Кои задачи във вашата работа могат да бъдат подпомогнати от ИИ и кои трябва да останат под човешки контрол?"*

Слайд 17–18: Потенциал за автоматизация и решения

Професионални групи с висок потенциал за промяна

72%

Офис подкрепа

Административни и секретарски задачи

60%

Обслужване на клиенти

Комуникация, отговори, обработка

55%

Бизнес и правни дейности

Документи, анализи, договори

48%

Производство и услуги

Рутинни производствени процеси

Политически и организационни решения

- **Условен базов доход** — предпазна мрежа, свързана със стимули за обучение
- **Преквалификация** — организационна и държавна подкрепа
- **ИИ грамотност** в учебните програми
- **Обучители като посредници** между технологичната промяна и реалните нужди

„Потенциал за автоматизация" не означава автоматично загуба на работа — означава, че задачи могат да бъдат **променени, ускорени или частично поети** от инструменти.

Слайд 19–20: Екологичният отпечатък на ИИ

Дигиталните услуги не са „безтегловни": те използват електроенергия, вода, материали и инфраструктура. Това не означава да спрем да използваме ИИ, а да го използваме **осъзнато и пропорционално на задачата**.

Практически съвети за обучители

- Формулирайте **ясни и точни промптове** — избягвайте излишни повторения
- **Групирайте въпросите** вместо да изпращате много отделни заявки
- Избирайте **подходящ инструмент** — не използвайте големи модели за тривиални задачи
- **Проверявайте преди** да генерирате много версии

Съчетаваме **продуктивност с отговорност** — осъзнатата употреба е ключова компетентност на ИИ грамотността.

Слайд 21: Заключение — ИИ грамотността като гражданска компетентност

ИИ грамотността не е само техническо умение, а **гражданска, професионална и етична компетентност**. Участниците трябва да могат да използват ИИ, но и да разбират ограниченията, рисковете, възможните злоупотреби и обществените последствия.

🔍 Предразсъдъци и дискриминация

Данните отразяват обществото — ИИ може да усили неравенствата

🛡️ Враждебни атаки и сигурност

Умишлени манипулации могат да подведат дори точни системи

⚠️ Злоупотреби и дезинформация

Мощните инструменти изискват медийна и критична грамотност

🌍 Работа и планета

Трансформация на труда и екологична отговорност

> Насърчете обучителите да въведат ИИ грамотност в своите курсове чрез дискусии, практически задачи, анализ на казуси и критична рефлексия.

Посетете exchangeai.eu за останалите модули