Как работи изкуственият интелект?

Етичен одит на ИИ

преподаваме.bg | Yettel.

8 - 12 клас

преподаваме.bg

Можете ли да се доверите на отговорите на ИИ?

Най-голямата платформа с практични и качествени обучения и ресурси за учители и директори

преподаваме.bg

Пример 1: Ученик пита ChatGPT за есе по биология

Промпт на ученика:

„Обясни ми механизма на CRISPR и дай научни източници."

Отговор на ChatGPT:

„CRISPR-Cas9 е революционна технология за редактиране на гени, открита през 2012 г. от Дженифър Дудна и Еманюел Шарпантие. Механизмът включва използването на РНК молекула като „навигатор", която насочва ензима Cas9 към точното място в ДНК веригата.

Източници:

— Petrov, I. & Ivanova, M. (2021). Gene Editing and Ethical Boundaries. Journal of Molecular Biology, 48(3), 112–128.

— Dimitrov, K. (2020). CRISPR Applications in Human Medicine. Sofia University Press.

Най-голямата платформа с практични и качествени обучения и ресурси за учители и директори

преподаваме.bg

Въпрос за обсъждане

Можем ли да се доверим на този отговор на ИИ?

преподаваме.bg

Пример 1: Ученик пита ChatGPT за есе по биология

Промпт на ученика:

„Обясни ми механизма на CRISPR и дай научни източници."

Източници:

— Petrov, I. & Ivanova, M. (2021). Gene Editing and Ethical Boundaries. Journal of Molecular Biology, 48(3), 112–128.

— Dimitrov, K. (2020). CRISPR Applications in Human Medicine. Sofia University Press.

⚠️ Журналът „Journal of Molecular Biology, 48(3)" не съдържа тази статия. Авторите „Petrov & Ivanova" не съществуват. „Sofia University Press" не е издал тази книга. Всичко е измислено – с пълна увереност.

Пример 2: Компания използва ИИ система за подбор на персонал.

Контекст: Двама кандидати с идентично образование и опит

	Кандидат А	Кандидат Б
Образование	Технически университет, 5.50	Технически университет, 5.50
Опит	3 години, 2 компании	3 години, 2 компании
Умения	Python, SQL, Machine Learning	Python, SQL, Machine Learning
Извънкласни	Шахматен клуб, математическа олимпиада	Женски шахматен клуб, математическа олимпиада
Оценка от ИИ	⭐⭐⭐⭐⭐ 92 точки	⭐⭐⭐ 67 точки

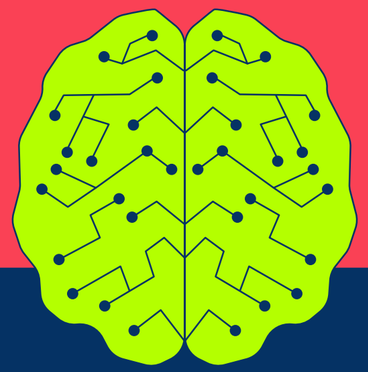
Най-голямата платформа с практични и качествени обучения и ресурси
за учители и директори

преподаваме.bg

Можем ли да се доверим на този резултат от ИИ?

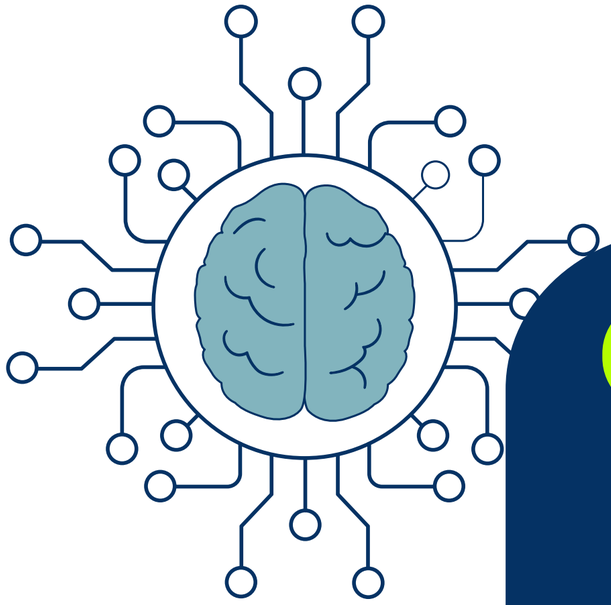
Пример 2: Компания използва ИИ система за подбор на персонал.

⚠️ Системата е обучена върху данни от последните 10 години – в които 90% от наетите в технологичния сектор са мъже. Тя е научила, че думата „женски" корелира с „по-рядко нает" – и автоматично понижава оценката. Не защото е „решила" да дискриминира. А защото данните са я научили точно на това.



Две важни понятия:

преподаваме.bg



Халюцинация

Примерът с есето по биология

Моделът генерира нещо, което звучи напълно правдоподобно – и не знае, че не е вярно.

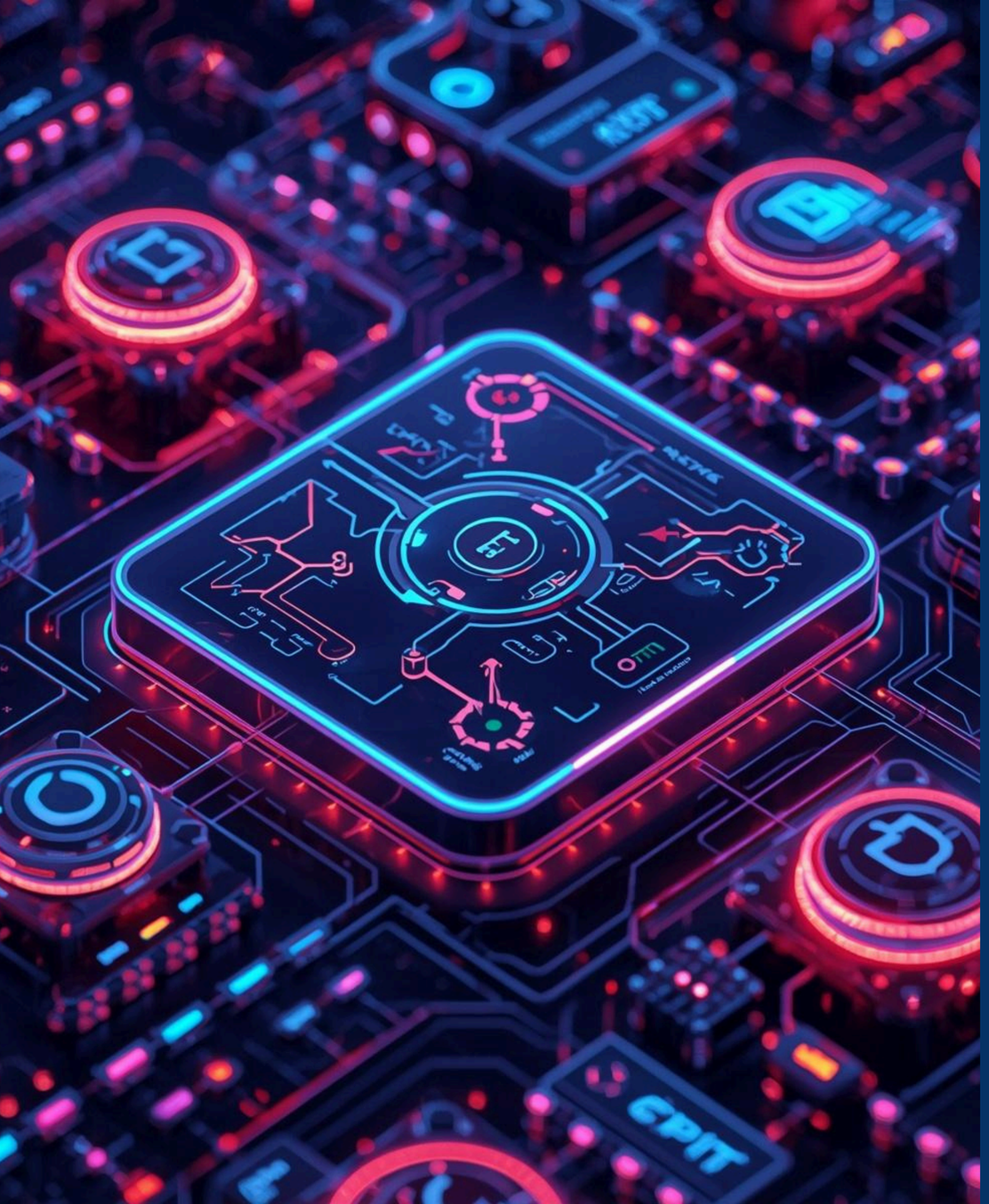
ХАЛЮЦИНАЦИЯ = правдоподобно, но невярно

Алгоритмичен предразсъдък

Примерът за подбор на персонал

Системата е обучена върху данни в технологичния сектор, които показват, че 90% от наетите са мъже.

АЛГОРИТМИЧЕН ПРЕДРАЗСЪДЪК = данните отразяват неравенство → моделът го научава → моделът го възпроизвежда



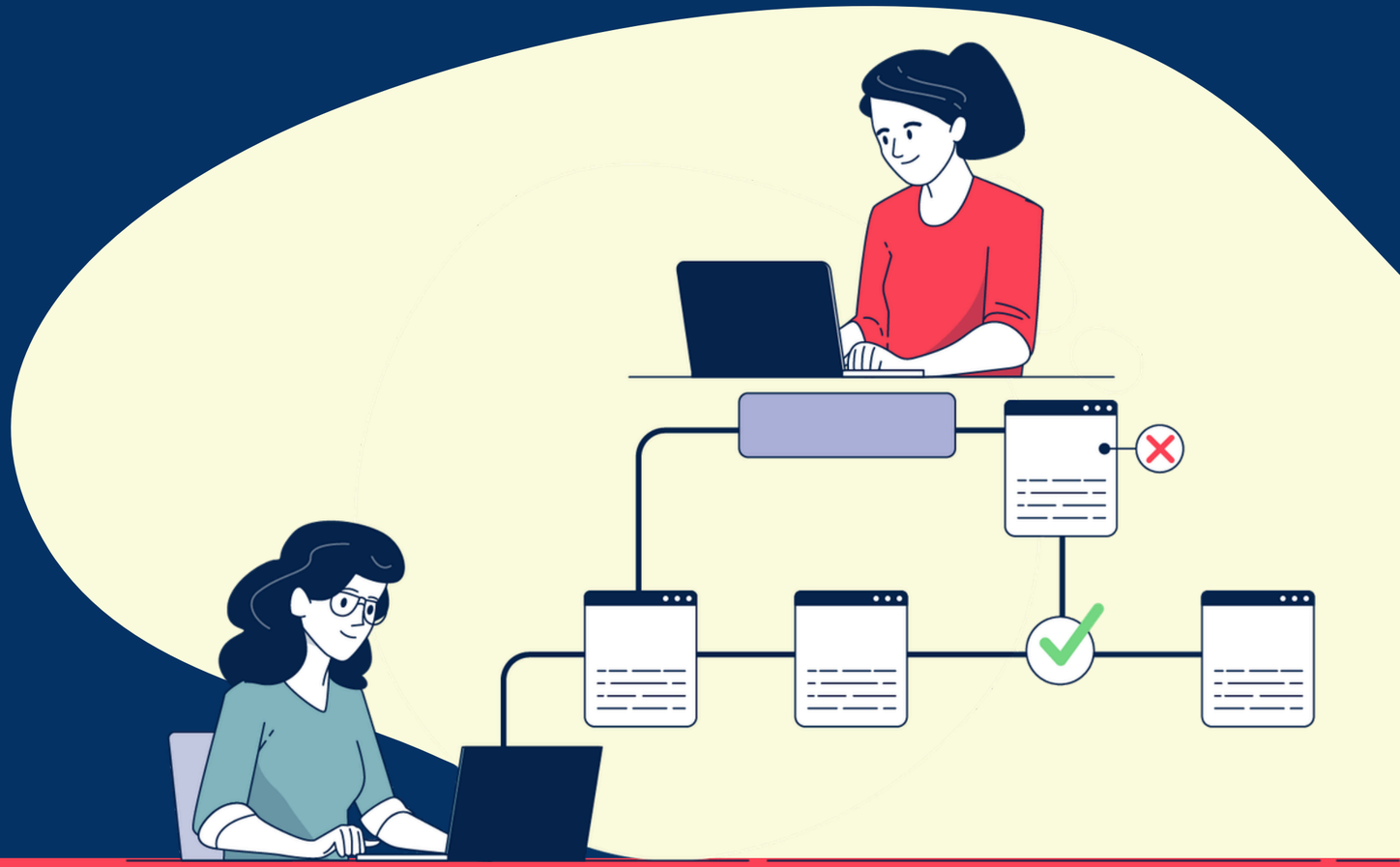
преподаваме.bg

Не всички грешки на ИИ са технически.

Някои грешки са етични.

преподаваме.bg

Етичен одит на ИИ



Ще тестваме реални ИИ инструменти

преподаваме.bg

1. Използвате работния лист Одитна карта

2. Тествате ИИ с четирите промпта от листа

3. За всеки отговор попълвате дали е точен, непълен, каква е грешката, кога греши?

4. Работите в групи по 3-4 души

5. Имате 15 минути за работа (10 минути за тестване и 5 за обобщение)

преподаваме.bg

Всяка група споделя само една находка.

Групираме ги в две колони - технически и етични.

преподаваме.bg

Дебат

Трябва ли ИИ да взема решения вместо хора?

преподаваме.bg

Дебат: Етап 1

Трябва ли ИИ да взема самостоятелни решения с последствия за хора – в медицината, в правото, в образованието?

За 2 минути прочетете ролята си и подгответе позицията си.

1. Какъв е най-силният аргумент в полза на вашата позиция?
2. Какво би казал опонентът ви – и как отговаряте?
3. Има ли граница, до която позицията ви е валидна?

преподаваме.bg

Дебат: Етап 2

Трябва ли ИИ да взема самостоятелни решения с последствия за хора – в медицината, в правото, в образованието?

Всяка роля представя позицията си в рамките на 1 минута.

- Не прекъсваме.
- Не коментираме.
- Само слушаме.

преподаваме.bg

Дебат: Етап 3

Трябва ли ИИ да взема самостоятелни решения с последствия за хора – в медицината, в правото, в образованието?

Имате право на реплика към всяка друга роля. Правилата са три:

1. Говорите от ролята, не от личното мнение.
2. Коментирате аргумента, не човека.
3. Ако някой е казал нещо, с което сте съгласни, можете да го признаете. Дебатът не е война.

преподаваме.bg

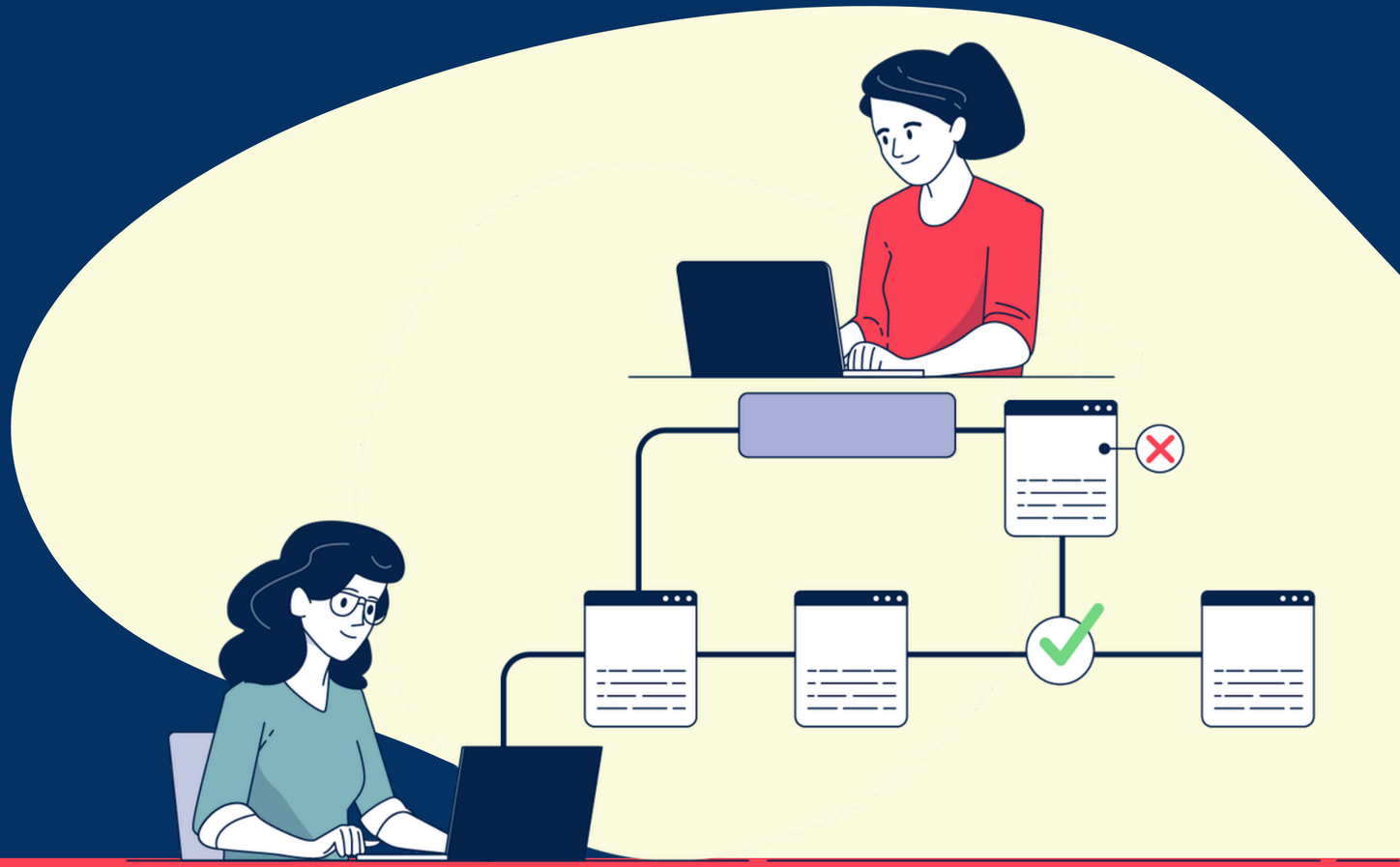
Дебат: Финален въпрос

Кой носи отговорността?

Когато ИИ система вземе решение – за лечение, за присъда, за оценка – и това решение наранява някого… кой носи моралната отговорност?

преподаваме.bg

Етичен доклад за изкуствен интелект



Кратък, структуриран анализ на ИИ инструмента, който тествахте

преподаваме.bg

1. Използвате работния лист Етичен доклад

2. Използвайте работните листове и наблюденията от предишния час и ги превърнете в конкретни препоръки

3. Работите в групи по 3-4 души

4. Имате 15 минути, за да отговорите на трите въпроса

преподаваме.bg

Всяка група споделя само заключението от доклада си.

Най-голямата платформа с практични и качествени обучения и ресурси за учители и директори

преподаваме.bg

За финал:

Запишете едно изречение, което започва с „Не очаквах, че..”

преподаваме.bg

Благодаря!