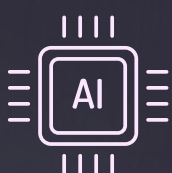


Етика и управление на изкуствения интелект на практика

~° материал на британския институт
"Алън Тюринг"

<https://www.turing.ac.uk/news/publications/ai-ethics-and-governance-practice-ai-safety-practice>



Документът представлява практическо ръководство за прилагане на етични принципи и ефективно управление на системи с изкуствен интелект в публичния и частния сектор, разработено от Института „Алън Тюринг“.

Моделът **SSAFE-D**

В рамките на документа е представен и SSAFE-D моделът. Целта му е да превърне етичните принципи в ясни, приложими практики за всеки проект с ИИ.



S Устойчивост / **Sustainability**

Проектите с ИИ трябва да бъдат чувствителни към реалния им социален и екологически ефект. Да включват заинтересованите страни и да се прави оценка на въздействието

S Безопасност / **Safety**

Проектите със системите с ИИ трябва да са сигурни и надеждни. Трябва да се документират потенциалните рискове и действията за справяне с тях.

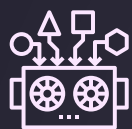


A Отчетност / **Accountability**

В проектите трябва да има ясни роли, отговорност и проследимост на решенията (одитируемост).

F Справедливост / **Fairness**

Проектите трябва да предотвратят дискриминация и неравенство (самооценка на предубежденията, управление на риска от предубеждения и т.н).



E Обяснимост / **Explainability**

Процесите на проектите за ИИ и техните резултати трябва да са обясними и обосновани. Решенията на ИИ трябва да са прозрачни и ясни за гражданите

D Управление на данните / **Data Stewardship**

Защита на данните, неразкриване на лична информация и добросъвестно обработване при всички етапи на проекта с ИИ.

