ЦЕНТЪР ЗА ОБУЧЕНИЕ – БАН

1000 София
ул. „Сердика“ № 4
http://edu.bas.bg

email: tdc-phd@cu.bas.bg
тел.: 02 987 31 67
02 979 52 60

Специализиран докторантски курс

Основна информация:

Име на курса: **Изкуствен интелект в хуманитарните и социалните науки**
Лектор: проф. д.н. Стоян Ставру
Телефон: 0888280888
Имейл: stoyan.stavru@ips.bas.bg
Хорариум: 30 часа лекции + 90 часа самостоятелна работа по идейния проект на дисертацията

Анотация:

Курсът изследва правните и етическите проблеми, свързани с използването на генеративния изкуствен интелект (ИИ) и големите езикови модели (ГЕМ) в хуманитарните и социалните науки. Направени са важни разграничения в начините на употреба на тези нови инструменти при извършването на научните изследвания, като специално внимание е обърнато на предизвикателствата пред докторантите предвид обстоятелството, че тяхното обучение е свързано с придобиването на образователна и научна степен „доктор“. Това изисква ясно разграничаване на етически допустимите употреби на ГЕМ в образованието и в науката, както и представяне на съществуващите регулации и практики, което са от значение за докторантите при тяхната работа. Разглежданите въпроси са обособени в две големи групи по линията на академичното четене (проучване на литература, запознаване с първоизточници, интерпретиране на оригиналния текст на използваните произведения и др.) и по линията на академичното писане (планиране и структуриране на изложението, писане и генериране на текстове, използване на ИИ за редактиране на научни текстове и др.).

Тематично съдържание на курса:

Тема 1: Специфики на хуманитарните и социалните науки.
Тема 2: Генеративният ИИ в образованието и в науката. Големи езикови модели. Значение на инструкциите (prompt engineering.). Агентен ИИ.
Тема 3: Генеративният ИИ като изследователски инструмент. Автоматизирани изследвания.
Тема 4: „Усилено“, „отдалечено“ и „слабо“ четене. Проучване на литературата. Рискове от делегирането на интерпретативни задачи на ИИ.
Тема 5: ИИ в емпиричните изследвания. Синтетични данни и симулации на социално поведение.
Тема 6: Плагиатство и използване на големи езикови модели. ИИ и академичното писане.
Тема 7: Генеративният „обрат“ в научното авторство. Авторство и генеалогия.
Тема 8: Изкуствен интелект и феноменология на авторството. Съавторство. Свръхавторство.
Тема 9: Генеративният ИИ в издателските политики. Рецензиране. Забрани и задължения за деклариране.
Тема 10: Инструменти за установяване на употребата на генеративен изкуствен интелект. Доказване на нарушение. Регулации и възможни санкции.

Тема 11: ИИ и оценяването на научните проекти. Неравенства и състезателност. Изкуствен интелект в работата на етическите комисии.
Тема 12: Влияние на големите езикови модели върху съдържанието на научните текстове. Хомогенизация и разнообразие.
Тема 13: Генеративният ИИ и критериите за научен принос.
Тема 14: Проследимост и защита на подпомогнатото от ИИ изследване. Изследователски дневник, възпроизводимост и отговорност.
Тема 15: Когнитивно натоварване и интелектуална зависимост. Загуба на изследователски умения. Хипотеза за „мъртвата“ наука.

Форми на обучение и оценяване:

Правен анализ на приложимото българско, европейско и международно законодателство.
Провеждане на дискусии и обсъждания.
Самостоятелно проучване на проблеми с проверка на първичните източници и подготовка на анализ, свързан с идейния проект на дисертацията.
Изпит чрез защита на самостоятелен анализ на казус или изследователски проблем.
Критерии за оценяване: точност на източниците, разграничаване на правни и етически аргументи, критическо обсъждане на алтернативите и обоснованост на изводите.

Компетентности, придобити в резултат на обучението:

- способност да се ориентират в основната етична и правна терминология и да я използват при обсъждането на етически допустимите употреби на генеративния изкуствен интелект в хуманитарните и социалните науки;
- умение да използват големи езикови модели при извършването на своите научноизследователски задачи по начин, който съответства на актуалните регулации;
- умение да прилагат приложимите издателски практики, свързани с етически допустимата употреба на генеративен изкуствен интелект в хуманитарните и социалните науки.

Литература:

Основна литература:

Bender, E. M., T. Gebru, A. McMillan-Major, S. Shmitchell On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? // Proceedings of FAccT ’21, 2021, p. 610-623, https://doi.org/10.1145/3442188.3445922;
Birhane, A., A. Kasirzadeh, D. Leslie, S. Wachter Science in the Age of Large Language Models. // Nature Reviews Physics, 2023, № 5, p. 277-280. https://doi.org/10.1038/s42254-023-00581-4;
Bjelobaba, S., L. Waddington, M. Perkins, T. Foltýnek, S. Bhattacharyya, D. Weber-Wulff Maintaining Research Integrity in the Age of GenAI: An Analysis of Ethical Challenges and Recommendations to Researchers. // International Journal for Educational Integrity, 2025, № 21, 18. https://doi.org/10.1007/s40979-025-00191-w;
Floridi, L., M. Chiriatti GPT-3: Its Nature, Scope, Limits, and Consequences. // Minds and Machines, 2020, № 30, p. 681-694. https://doi.org/10.1007/s11023-020-09548-1;

Gao, C., F. Howard, N. Markov, E. Dyer, S. Ramesh, Y. Luo, A. Pearson Comparing Scientific Abstracts Generated by ChatGPT to Real Abstracts with Detectors and Blinded Human Reviewers. // npj Digital Medicine, 2023, No 6, art. No 75. https://doi.org/10.1038/s41746-023-00819-6;
Islam, G., M. Greenwood Generative Artificial Intelligence as Hypercommons: Ethics of Authorship and Ownership. // Journal of Business Ethics, 2024, Vol. 192, p. 659-663. https://doi.org/10.1007/s10551-024-05741-9
Messeri, L., M. Crockett Artificial Intelligence and Illusions of Understanding in Scientific Research. // Nature, 2024, No 627, p. 49-58. https://doi.org/10.1038/s41586-024-07146-0;
Piper, A. A Theory-First Approach toward Using Generative AI for Humanities Research. // Computational Humanities Research, 2026, Vol. 2. https://doi.org/10.1017/chr.2026.10039
Mann, S., A. Vazirani, M. Aboy, B. Earp, T. Minssen, I. Cohen, J. Savulescu, J. Guidelines for Ethical Use and Acknowledgement of Large Language Models in Academic Writing. // Nature Machine Intelligence, 2024, No 6, p. 1272-1274. https://doi.org/10.1038/s42256-024-00922-7
van Dis, E., J. Bollen, W. Zuidema, R. van Rooij, C. Bockting ChatGPT: Five Priorities for Research. // Nature, 2023, No 614, p. 224-226. https://doi.org/10.1038/d41586-023-00288-7.

Институционални източници:

Европейски съюз Регламент (ЕС) 2024/1689 на Европейския парламент и на Съвета от 13 юни 2024 година (Актът за изкуствения интелект), публикуван в Официален вестник на Европейския съюз L, 2024/1689. https://eur-lex.europa.eu/legal-content/BG/TXT/HTML/?uri=OJ:L_202401689;
ALLEA The European Code of Conduct for Research Integrity. Revised Edition 2023. Berlin: All European Academies. https://doi.org/10.26356/ECOC;
European Commission Living Guidelines on the Responsible Use of Generative AI in Research. 2026, Third edition. https://tinyurl.com/2s48akjn;
Miao, F., W. Holmes Guidance for Generative AI in Education and Research. Paris: UNESCO. 2023, https://unesdoc.unesco.org/ark:/48223/pf0000386693.

Допълнителна информация:

Няма предварителни изисквания. Препоръчително е докторантът да може да работи с източници на английски език.