Aprobat: ***Consiliul ȘD Științe ale Naturii***
Proces verbal 9 din 27.10.25

FIȘA DISCIPLINEI
Ciclul III, DOCTORAT

Domeniul studii de doctorat	**Tehnologii ale informației și comunicațiilor**
Programul de doctorat/ specialitatea	**122.03. Modelare, metode matematice, produse program**
Codul și Denumirea disciplinei	**S.02.O.7 Studiu istoriografic și bibliografic: Inteligența Artificială Explicabilă (XAI) pentru detecția de fraudă și risc**
Titularul disciplinei	**Inga ȚIȚCHIEV , conf. univ., dr. în informatică**

Numărul de ore				Nr. de credite	Forma de evaluare
Total	**Prelegeri**	**Seminare**	**Lucrul individual**		
180	**4**	**6**	**170**	**6**	***Examen***

Fundamentare	Disciplina dezvoltă capacitatea de asimilare a cunoștințelor avansate în Inteligența Artificială Explicabilă (XAI) și aplicarea acestora pentru a spori transparența și încrederea în sistemele AI utilizate în domenii critice, cum ar fi detecția automată a fraudelor și riscurilor în asigurări.
Conținutul disciplinei	1. Fundamentele Inteligenței Artificiale Explicabile (XAI). Definire, obiective (încredere, etică, conformitate) și provocări. 2. Impactul reglementărilor (GDPR, EU AI Act) asupra transparenței și necesitatea explicabilității în AI. 3. Taxonomia metodelor XAI. Clasificarea după tip (global/local), scop (fidelitate/simplitate) și agnosticism (agnostic/specific modelului). 4. Algoritmi de explicabilitate locală. Implementarea și interpretarea LIME (Local Interpretable Model-agnostic Explanations). 5. Algoritmi de explicabilitate locală/globală. Implementarea și interpretarea SHAP (SHapley Additive exPlanations). 6. Utilizarea Permutation Feature Importance și a metodelor bazate pe surrogate models pentru explicabilitate globală. 7. Modele AI pentru detecția de anomalii și fraudă/risc (e.g., outlier detection, imbalanced learning). 8. Măsurarea și evaluarea calității explicabilității. Fidelitate, stabilitate, și intelizibilitatea pentru utilizatorul final. 9. Integrarea componentelor XAI în sistemul distribuit (pe bază de microservicii) pentru furnizarea de explicații în timp real. Utilizarea XAI pentru auditarea și identificarea bias-ului modelului, asigurând echitatea deciziilor în aplicații critice (asigurări).
Competențele obținute/ Rezultatele învățării	**CP 1.** Asimilarea metodelor de cercetare specifice Inteligenței Artificiale și dezvoltarea capacității de a implementa algoritmi XAI. **CP 2.** Proiectarea unui subsistem de detecție automată a fraudelor și riscurilor bazat pe AI explicabil. **CP 3.** Analiza și interpretarea deciziilor AI utilizând instrumente XAI pentru a identifica și reduce potențialul bias al modelului. **CP 4.** Elaborarea de publicații științifice care să demonstreze impactul creșterii transparenței AI în aplicații practice.
Bibliografia selectivă/ minimală	1. Molnar, Christoph. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (Second Edition). Independently published. 317 p. [Online: https://christophm.github.io/interpretable-ml-book/] 2. Wieringa, Roel J. (2014). Design Science Methodology for Information Systems and Software Engineering. Springer. 332 p. [Online: https://doi.org/10.1007/978-3-662-43839-8] 3. Ribeiro, Marco Tulio; Singh, Sameer; Guestrin, Carlos. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. În: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144). ACM. 4. Goodman, Bryce; Flaxman, Seth. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". AI Magazine, 38(3), 50-57.