UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA V KOŠICIACH
Prírodovedecká fakulta

INSTITUTE OF COMPUTER SCIENCE
Jesenná 5, 041 54 Košice, Slovakia
WWW: ics.upjs.sk

Názov odboru:	Informatika
Názov študijného programu:	Informatika
Názov dizertačnej práce:	**Vysvetliteľná umelá inteligencia**
Meno školiteľa:	prof. RNDr. Gabriel Semanišin, PhD. (gabriel.semanisin@upjs.sk)
Konzultant:	doc. RNDr. Ľubomír Antoni, PhD. (lubomir.antoni@upjs.sk)
Forma realizácie DŠ:	denná

Anotácia:
V posledných rokoch zažívame nebývalý rozvoj metód umelej inteligencie a ich intenzívne využívanie v rôznych oblastiach života ľudskej spoločnosti. Spolu s využívaním umelej inteligencie vyvstávajú aj otázky ohľadom spoľahlivosti a dôveryhodnosti týchto metód a ich potenciálneho zneužitia alebo neadekvátneho použitia. Zvlášť sú tieto otázky citlivé v oblastiach, kde sa vyžaduje vysoká spoľahlivosť alebo majú priamy vplyv na zdravie alebo sociálnu integritu človeka. Vo všeobecnosti totiž pri väčšine metód umelej inteligencie nie je úplne transparentný spôsob, akým prebieha výpočtový a rozhodovací proces. Ten je realizovaný pomocou veľkého množstva parametrov, ktorých interpretácia je veľmi komplikovaná. Tieto problémy sa snaží eliminovať vysvetliteľná umelá inteligencia (XAI), ktorej cieľom je vyvíjať modely a techniky na zabezpečenie vysvetliteľnosti jednotlivých algoritmov a metód. Cieľom dizertačnej práce je analyzovať známe postupy a pokúsiť sa vyvinúť nové vysvetliteľné systémy umelej inteligencie, ktoré bude možné považovať nielen za efektívne, ale aj dôveryhodné.

Field:	Computer Science
Study programme:	Computer Science
Title of thesis:	**Explainable artificial intelligence**
Supervisor:	prof. RNDr. Gabriel Semanišin, PhD. (gabriel.semanisin@upjs.sk)
Co-supervisor:	doc. RNDr. Ľubomír Antoni, PhD. (lubomir.antoni@upjs.sk)
Form of study:	internal

Annotation:
In recent years, we have experienced an unprecedented development of artificial intelligence methods and their intensive use in various areas of human society. Along with the use of artificial intelligence, questions arise about the reliability and trustworthiness of these methods and their potential misuse or inadequate use. These issues are particularly sensitive in areas where high reliability is required or where they have a direct impact on human health or social integrity. Indeed, in general, most AI methods are not fully transparent in the way in which the computational and decision-making process is carried out. It is implemented using a large number of parameters, the interpretation of which is very complicated. Explainable Artificial Intelligence (XAI), which aims to develop models and techniques to ensure the explainability of individual algorithms and methods, seeks to eliminate these problems. The aim of this dissertation is to analyse known approaches and try to develop new explainable artificial intelligence systems that can be considered not only efficient but also trustworthy.