

Vysvetliteľná umelá inteligencia pre moderáciu zmluvných pokút súdom

Explainable artificial intelligence for the moderation of contractual penalties by the court

Mgr. Andrej Oriňak

Katedra obchodného práva a hospodárskeho práva, Právnická fakulta Univerzity Pavla Jozefa Šafárika v Košiciach, <https://orcid.org/0009-0008-0123-8076>

Anotácia

Inštitút moderácie zmluvnej pokuty súdom predstavuje kľúčový nástroj na zabezpečenie spravodlivosti v záväzkových vzťahoch, pričom jeho aplikácia závisí od výkladu neurčitého právneho pojmu „neprimerane vysoká“ pokuta. Súčasný nástup umelej inteligencie do právnej oblasti ponúka možnosti na analýzu rozsiahlej judikatúry, no naráža na problém „čiernej skrinky“, ktorý je v rozpore s požiadavkou odôvodnenia súdnych rozhodnutí. V príspevku skúmame potenciál vysvetliteľnej umelej inteligencie pri objektivizácii kritérií pre moderáciu pokút v kontexte slovenskej právnej úpravy. Na základe analýzy právnej doktríny a praktickej skúšky technických možností vysvetliteľnej umelej inteligencie v predkladanom príspevku hodnotíme, či je možné pomocou algoritmov poskytnúť právne relevantné odôvodnenie pre moderáciu zmluvnej pokuty. V predkladanom príspevku dospievame k záveru, že najvhodnejším modelom je využitie umelej inteligencie ako pomocníka v hybridnom systéme, kde konečné posúdenie primeranosti ostáva v kompetencii ľudského sudcu.

Annotation

The court's authority to moderate contractual penalties is a key tool for ensuring fairness in contractual relationships, and its application depends on the interpretation of the vague legal concept of an "excessively high" penalty. The current emergence of artificial intelligence in the legal field offers opportunities for analyzing extensive case law, but it faces the "black box" problem, which conflicts with the requirement to justify judicial decisions. In this paper, we examine the potential of explainable artificial intelligence in objectifying the criteria for moderating fines in the context of Slovak law. Based on an analysis of legal doctrine and a practical examination of the technical capabilities of explainable artificial intelligence, this paper assesses whether it is possible to use algorithms to provide legally relevant justification for the mitigation of contractual penalties. In this paper, we conclude that the most suitable model is the use of artificial intelligence as an assistant in a hybrid system, where the final assessment of proportionality remains within the competence of a human judge.

Kľúčové slová

zmluvná pokuta, moderačné právo, interpretovateľnosť, SHAP, LIME, Grok

Key words

contractual penalty, moderation right, interpretability, SHAP, LIME, Grok

I. Úvod

Zmluvná pokuta predstavuje v slovenskom právnom poriadku jeden z najčastejšie využívaných inštitútov na zabezpečenie záväzkov. Plní nie len preventívnu a sankčnú funkciu, ale často aj funkciu paušalizovanej náhrady škody. V zmluvnej praxi však dochádza k situáciám, kedy je dojednaná výška sankcie v zjavnom nepomere k hodnote a významu zabezpečovanej povinnosti, čím sa narúša spravodlivé usporiadanie vzťahov medzi veriteľom a dlžníkom. Zákonodarca na tieto situácie reaguje inštitútom moderačného práva súdu, ktorý umožňuje neprímerane vysokú zmluvnú pokutu znížiť, pričom v obchodnoprávných vzťahoch je spodnou hranicou moderácie výška vzniknutej škody.¹ Aplikácia tohto práva však závisí od interpretácie vágneho pojmu „neprimeranosť“, čo v súdnej praxi môže vytvárať priestor pre subjektivitu a nekonzistentnosť rozhodovania.

Súčasne s týmto právnym problémom sme svedkami transformácie právneho odvetvia prostredníctvom umelej inteligencie, najmä v oblasti spracovania prirodzeného jazyka a veľkých jazykových modelov. Tieto technológie preukazujú schopnosť analyzovať rozsiahle korpusy právnych textov, predpovedať súdne rozhodnutia či automatizovane revidovať zmluvy.² Napriek ich potenciálu zefektívniť justičné procesy naráža ich nasadenie pri tak citlivých otázkach, akou je moderácia zmluvnej pokuty, na zásadnú prekážku: problém „čiernej skrinky“. Zložitosť a netransparentnosť fungovania systémov umelej inteligencie sťažuje orgánom verejnej moci posudzovanie súladu ich využitia s právnymi predpismi a určovanie právnej zodpovednosti.³ Moderné algoritmy, hoci presné, často nedokážu vysvetliť, ako dospeli k svojim záverom, čo je v priamom rozpore s požiadavkou na odôvodnenosť a transparentnosť súdnych rozhodnutí.⁴

Absencia vysvetliteľnosti bráni širšej integrácii umelej inteligencie do rozhodovacej činnosti, pretože účastníci konania majú právo rozumieť dôvodom, ktoré viedli k zásahu do ich zmluvnej slobody. Odborníci vnímajú (ne)transparentnosť systémov umelej inteligencie ako významnú výzvu, ktorú je potrebné prekonať pri ich zavádzaní do právnej praxe.⁵ Odpoveďou na túto výzvu môže byť rozvíjajúca sa oblasť vysvetliteľnej umelej inteligencie. Cieľom tejto oblasti je objasniť vnútorné fungovanie modelov umelej inteligencie a poskytnúť používateľom zrozumiteľné zdôvodnenia výstupov týchto modelov.⁶ V predkladanom príspevku skúmame možnosti využitia vysvetliteľnej umelej inteligencie pri objektivizácii kritérií pre súdnu moderáciu zmluvných pokút v kontexte slovenského právneho

¹ OVEČKOVÁ, O. 2024. Zmluvná pokuta a náhrada škody v obchodnom práve. In *Právny obzor* [online]. Roč. 107, č. 6 [cit. 2025-12-25], s. 624 – 626. Dostupné na: <https://doi.org/10.31577/pravnyobzor.2024.6.04>.

² SIINO, M. et al. 2025. *Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches*. [online] 2025 [cit. 25.12.2025]. s. 18254. Dostupné na: <https://ieeexplore.ieee.org/document/10850911>.

³ MONTAGNANI, M. L., NAJJAR, M.C. a DAVOLA, A. 2024. The EU Regulatory approach(es) to AI liability, and its Application to the financial services market. In *Computer Law & Security Review* [online]. 2024, roč. 53. [cit. 2026-01-15], s. 1. Dostupné na: <https://doi.org/10.1016/j.clsr.2024.105984>.

⁴ ATKINSON, K., BENCH-CAPON, T. a BOLLEGALA, D. 2020. *Explanation in AI and law: Past, present and future*. *Artificial Intelligence* [online]. 2020. Roč. 289, s. 17. [cit. 25.12.2025]. Dostupné na: <https://doi.org/10.1016/j.artint.2020.103387>

⁵ KIM, S., YI, S. a PARK, S.-P. 2025. Prioritizing challenges in AI adoption for the legal domain: A systematic review and expert-driven AHP analysis. In *PLOS One* [online]. 2025, roč. 20, č. 6 [cit. 15.01.2026], s. 4. Dostupné na: <https://doi.org/10.1371/journal.pone.0326028>

⁶ BENNETOT, A. et al. 2024. A Practical Tutorial on Explainable AI Techniques. In *ACM Computing Surveys* [online]. 2024, roč. 57, č. 2 [cit. 25.12.2025], s. 2. Dostupné na: <https://doi.org/10.1145/3670685>.

poriadku. Analyzujeme, či môžu moderné techniky vysvetliteľnej umelej inteligencie pomôcť identifikovať faktory vedúce k neprimeranosti zmluvnej pokuty v zmluvách a poskytnúť právne relevantné argumenty, ktoré by slúžili ako podpora pre sudcovskú úvahu. Jedným z cieľov predkladaného príspevku je prepojiť doktrínálnu analýzu inštitútu zmluvnej pokuty v slovenskom právnom poriadku s možnosťami moderných metód vysvetliteľnej umelej inteligencie. Ďalším cieľom predkladaného príspevku je prostredníctvom praktického experimentu overiť, či dokážu tieto technológie objektivizovať proces posudzovania neprimeranosti zmluvnej pokuty. V predkladanom príspevku sa nesústreďujeme len na teoretické vymedzenie, ale prostredníctvom prípadovej štúdie testujeme schopnosti generatívnej umelej inteligencie po aplikácii *post-hoc* metód na konkrétny právny spor, čím konfrontujeme algoritmické rozhodovanie s ustálenou judikatúrou.

Pre účely výskumu v rámci predkladaného príspevku sme sformulovali nasledovnú výskumnú otázku: „Ako môžu metódy vysvetliteľnej umelej inteligencie prispieť k objektivizácii vágneho právneho pojmu „neprimerane vysoká zmluvná pokuta“ pri zachovaní požiadavky na transparentnosť a odôvodnenosť súdneho rozhodovania?“

Na účely zodpovedania výskumnej otázky sme sformulovali dve vedecké hypotézy: Prvá znie: „Aplikácia *post-hoc* metód vysvetliteľnosti umožňuje identifikovať faktory vedúce k moderácii pokuty v judikatúre s dostatočnou presnosťou.“ Druhá vedecká hypotéza znie: „Pomocou vysvetliteľnej umelej inteligencie je možné určiť neprimerane vysokú zmluvnú pokutu bez zásahu človeka.“

Pre dosiahnutie stanoveného cieľa a verifikáciu hypotéz využívame v predkladanom príspevku kombinovanú metodológiu, ktorá integruje normatívnu právnu analýzu s empirickým experimentom v prostredí umelej inteligencie. Výskum bol realizovaný v troch logicky na seba nadväzujúcich fázach.

V prvej fáze sme uplatnili metódu analýzy a syntézy platnej právnej úpravy a relevantnej judikatúry týkajúcej sa moderácie zmluvnej pokuty. Skúmali sme kľúčové kritériá moderácie zmluvnej pokuty v zmysle § 301 Obchodného zákonníka a § 545a Občianskeho zákonníka. Táto časť nám poslúžila na vytvorenie referenčného rámca, voči ktorému bol následne porovnaný výstup modelu umelej inteligencie.

V druhej fáze výskumu sme sa zamerali na teoretické vymedzenie nástrojov vysvetliteľnej umelej inteligencie. Metódou deskripcie a komparácie boli analyzované prístupy *post-hoc* vysvetliteľnosti, konkrétne metódy *Shapley Additive exPlanations* (ďalej len „SHAP“) a *Local Interpretable Model-agnostic Explanations* (ďalej len „LIME“), ktoré sú kľúčové pre pochopenie vplyvu jednotlivých faktorov pri algoritmickom rozhodovaní.⁷

V záverečnej fáze výskumu bola použitá metóda experimentu (prípadovej štúdie) s využitím veľkého jazykového modelu Grok. Bol vytvorený modelový právny prípad sporu o zaplatenie zmluvnej pokuty so špecifickými skutkovými okolnosťami. Získané výstupy (rozhodnutie a jeho vysvetlenie) boli následne metódou komparácie konfrontované s ustálenou judikatúrou Najvyššieho súdu SR. Na základe tejto

⁷ ALI, S. et al. 2023. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. In *Information Fusion* [online]. 2023, roč. 99 [cit. 25.12.2025] s. 10. Dostupné na: <https://doi.org/10.1016/j.inffus.2023.101805>.

komparácie bolo možné identifikovať nedostatky medzi „*strojovým*“ a „*ľudským*“ právnym uvažovaním a vyhodnotiť riziká spojené s automatizáciou moderácie.

Problematika využitia umelej inteligencie v procese súdnej moderácie zmluvných pokút sa nachádza na priesečníku dvoch dynamicky sa rozvíjajúcich oblastí: doktrinálneho výkladu neurčitých právnych pojmov v záväzkovom práve a technického pokroku v oblasti vysvetliteľnej umelej inteligencie. Zatiaľ čo právna veda dlhodobo analyzuje kritériá primeranosti sankcií, informatika sa v posledných rokoch intenzívne zameriava na to, ako „*otvoriť čiernu skrinku*“ algoritmov a sprístupniť ich rozhodovanie ľudskému chápaniu. Väčšina súčasných modelov umelej inteligencie zameraných na predpovedanie súdnych rozhodnutí však opomína skutočnú súdnu prax, pri ktorej sudcovia počas vyhodnocovania výsledku súdneho konania a spracovania rozsudku systematicky porovnávajú relevantné zákonné ustanovenia a iné súdne rozhodnutia v skutkovo podobných prípadoch.⁸

V slovenskom právnom poriadku je inštitút zmluvnej pokuty a jej moderácie upravený dualisticky (v čase spracovania predkladaného príspevku očakávaná novela Občianskeho zákonníka). V obchodnoprávných vzťahoch, podľa § 301 Obchodného zákonníka, súd môže znížiť neprimerane vysokú zmluvnú pokutu s prihliadnutím na hodnotu a význam zabezpečovanej povinnosti, a to až do výšky škody, ktorá vznikla do doby súdneho rozhodnutia.⁹ Tento limit predstavuje spodnú hranicu, pod ktorú súd v obchodnoprávných veciach nemôže zmluvnú pokutu znížiť. Uvedené považujeme za zdôraznenie reparačnej a kompenzačnej funkcie skúmaného inštitútu v podnikateľskom prostredí. V kontexte využitia umelej inteligencie považujeme možnosť presného určenia výšky zmluvnej pokuty, ktorá je primeraná za kľúčové pre zmluvné strany. Týmto spôsobom by bolo možné eliminovať riziko zdĺhavých súdnych konaní v budúcnosti a určiť primeranú výšku zmluvnej pokuty už pri kreovaní zmluvy.¹⁰ Je otáznе, či sú modely umelej inteligencie v aktuálnom stave takého určenia schopné.

Naproti tomu v občianskoprávných vzťahoch, podľa § 545a Občianskeho zákonníka súd prihliada nielen na hodnotu a význam zabezpečovanej povinnosti, ale aj na to, o koľko zmluvná pokuta presahuje rozsah vzniknutej škody, ak veriteľ nie je oprávnený požadovať náhradu škody popri zmluvnej pokute.¹¹ Kľúčovým aplikačným problémom v oboch režimoch ostáva výklad pojmu „*neprimerane vysoká*“. Judikatúra a doktrína sa zhodujú v tom, že neexistuje paušálna hranica neprimeranosti a posúdenie musí vychádzať z okolností konkrétneho prípadu, pričom sa zohľadňujú nielen majetkové záujmy, ale aj preventívna a sankčná funkcia zmluvnej pokuty.¹² V zahraničnej teórii sa v tejto súvislosti poukazuje na

⁸ ZHANG, H., DOU, Z., ZHU, Y. a WEN, J.-R. 2023. Contrastive Learning for Legal Judgment Prediction. In: *ACM Transactions on Information Systems* [online]. 2023, roč. 41, č. 4. [cit. 2026-01-15]. s. 2. Dostupné na: <https://doi.org/10.1145/3580489>

⁹ VOZÁR, J. 2022. Právomoc súdu znížiť zmluvnú pokutu. In: PATAKYOVÁ, M. a kol. *Obchodný zákonník : Komentár*. 1. vydanie. Bratislava : C. H. Beck, 2022, s. 1266. ISBN 978-80-8232-018-6.

¹⁰ ALSHBOUL, O., et al. 2022. *Prediction Liquidated Damages via Ensemble Machine Learning Model: Towards Sustainable Highway Construction Projects*. *Sustainability* [online]. 2022. Roč. 14, č. 15. [cit. 2026-01-15]. s. 1. Dostupné na: <https://doi.org/10.3390/su14159303>

¹¹ STRAPÁČ, P. 2019. Moderačné oprávnenie súdu. In: ŠTEVČEK, M. a kol. *Občiansky zákonník II*. 2. vydanie. Praha : C. H. Beck, 2019, s. 2031. ISBN 978-80-7400-770-5.

¹² VOZÁR, J. 2022. Charakteristika inštitútu. In: PATAKYOVÁ, M. a kol. *Obchodný zákonník : Komentár*. 1. vydanie. Bratislava : C. H. Beck, 2022, s. 1264. ISBN 978-80-8232-018-6.

potrebu vyvažovania práv dlžníka a veriteľa prostredníctvom diskrečnej právomoci súdu, ktorá však bez jasných štandardov môže viesť k nekonzistentnosti rozhodovania.¹³

S nástupom pokročilých modelov strojového učenia a veľkých jazykových modelov do právnej praxe vyvstáva fundamentálny problém tzv. „čiernej skrinky“. Tieto systémy dokážu s vysokou presnosťou predikovať výsledky na základe historických dát, no ich vnútorné rozhodovacie procesy sú často nepriehľadné. V právnom štáte je však neakceptovateľné, aby rozhodnutie o majetkových právach (akým je moderácia zmluvnej pokuty) bolo založené na výstupe, ktorého logiku nemožno preskúmať.¹⁴ Navyše, súčasné veľké jazykové modely majú tendenciu k „halucináciám“ a nemusia správne rozlišovať medzi *ratio decidendi* a *obiter dictum*, čo môže predstavovať pri práci s judikatúrou týkajúcou sa moderácie zmluvných pokút kritické riziko.¹⁵

Odpoveďou na tieto výzvy môže byť rozvoj disciplíny vysvetliteľnej umelej inteligencie, ktorá sa snaží sprístupniť modely ľudskému chápaniu. V kontexte analýzy zmlúv sa výskum posúva od jednoduchých kľúčových slov k sofistikovanejším metódam, ktoré využívajú pamäťové siete a mechanizmy pozornosti na identifikáciu neférových či nevýhodných zmluvných ustanovení. Štúdie preukázali, že tieto modely dokážu nielen detegovať potenciálne neférové podmienky v spotrebiteľských zmluvách, ale aj priradiť k nim náležité odôvodnenie, ktoré vysvetľuje, prečo je dané ustanovenie problematické. Napríklad pri analýze čílskych zmluvných podmienok sa podarilo pomocou umelej inteligencie identifikovať „šedé“ klauzuly (nejednoznačné podmienky) s vysokou presnosťou, čo môže naznačovať analogicky potenciál pre automatizovanú detekciu neprímeraných zmluvných pokút aj v iných jurisdikciách.¹⁶

II. Právne kritériá pre moderáciu zmluvnej pokuty

Inštitút zmluvnej pokuty v slovenskom právnom poriadku plní viaceré funkcie – zabezpečovaciú, preventívnu, sankčnú a v mnohých prípadoch aj uhradzovaciú (paušalizovanú náhradu škody). V aplikačnej praxi však dochádza k situáciám, kedy výška dojednanej sankcie zjavne nezodpovedá porušeniu povinnosti, čím vzniká nerovnováha v právach a povinnostiach strán. Zákonodarca na tento stav reaguje inštitútom moderačného práva súdu (*ius moderandi*), ktoré umožňuje neprímerane vysokú zmluvnú pokutu znížiť. Aplikácia tohto práva však nie je svojvoľná, ale viaže sa na splnenie špecifických hmotnoprávných kritérií, ktoré sú upravené duálne pre obchodnoprávne a občianskoprávne vzťahy. Stojí za zmienku, že podľa niektorých autorov môžu zmluvné sankcie za určitých okolností pôsobiť

¹³ REKHVIASHVILI, N. a KHARITONASHVILI, N. 2024. Judicial Reduction of Penalties in Georgian Civil Law: Balancing Debtor and Creditor Rights. In: *Rechtsidee* [online]. Roč. 12, č. 2, 2024. [cit. 25.12.2025]. s. 6. Dostupné na: <https://doi.org/10.21070/jjhr.v12i2.1031>

¹⁴ BROŽEK, B., FURMAN, M., JAKUBIEC, M. a KUCHARZYK, B. The black box problem revisited. Real and imaginary challenges for automated legal decision making. 2024. In: *Artificial Intelligence and Law* [online]. 2024. Roč. 32. [cit. 25.12.2025]. s. 429 – 430. Dostupné na: <https://doi.org/10.1007/s10506-023-09356-9>

¹⁵ LINNA, E. a LINNA, T. *Judicial Requirements for Generative AI in Legal Reasoning*. [online]. 2025. s. 5 [cit. 25.12.2025]. Dostupné na: <https://doi.org/10.48550/arXiv.2508.18880>

¹⁶ LÖFFLER, C., MARTÍNEZ FREILE, A. a REY PIZARRO, T. 2025. *Predicting potentially abusive clauses in Chilean terms of services with natural language processing*. [online]. 2025. [cit. 25.12.2025]. s. 1. Dostupné na: <https://arxiv.org/abs/2502.00865>

kontraproduktívne: namiesto plnenia záväzkov môžu paradoxne motivovať strany k nekorektnému konaniu.¹⁷

Obchodnoprávny režim

V obchodných záväzkových vzťahoch je moderácia upravená v § 301 Obchodného zákonníka. Toto ustanovenie má kogentnú povahu, čo znamená, že zmluvné strany sa od neho nemôžu odchyliť a moderačné právo súdu nemôžu vopred vylúčiť.¹⁸

Základným predpokladom pre zásah súdu je zistenie, že dojednaná zmluvná pokuta je „*neprimerane vysoká*“. Pri skúmaní tejto neprimeranosti zákon ukladá súdu povinnosť prihliadať na dve kľúčové kritériá:

- a) Hodnota zabezpečovanej povinnosti: Súd skúma ekonomický rozmer porušeného záväzku.
- b) Význam zabezpečovanej povinnosti: Súd hodnotí dôležitosť povinnosti pre veriteľa v čase porušenia, pričom význam možno vnímať subjektívne (existenčná závislosť veriteľa od plnenia) aj objektívne (kľúčový záväzok v rámci zmluvy).¹⁹

Špecifikom obchodnoprávnej úpravy je stanovenie pevnej spodnej hranice moderácie. Súd môže pokutu znížiť len do výšky škody, ktorá veriteľovi vznikla do doby súdneho rozhodnutia v dôsledku porušenia zmluvnej povinnosti.²⁰ Táto konštrukcia vychádza z predpokladu, že zmluvná pokuta v obchodnoprávných vzťahoch často plní funkciu paušalizovanej náhrady škody. Ak by súd znížil pokutu pod úroveň skutočnej škody, poprel by jej kompenzačnú funkciu a poškodil veriteľa. Z procesného hľadiska, v obchodnom práve platí, že súd nemôže moderovať pokutu z vlastnej iniciatívy (*ex officio*), ale len na návrh dlžníka, ktorý musí neprimeranosť nielen tvrdiť, ale aj preukázať.²¹

Občianskoprávny režim

V občianskoprávných vzťahoch je moderačné oprávnenie súdu zakotvené v § 545a Občianskeho zákonníka. Aj tu je primárnym kritériom „*neprimeraná výška*“ a súd prihliada na hodnotu a význam zabezpečovanej povinnosti. Občiansky zákonník však zavádza odlišný mechanizmus vo vzťahu ku škode.

Zákon rozlišuje situáciu, kedy veriteľ nie je oprávnený požadovať náhradu škody popri zmluvnej pokute (čo je dispozitívne pravidlo podľa § 545 ods. 2 OZ). V takom prípade súd pri moderácii „*prihliadne*“ aj na výšku škody, ktorá porušením povinnosti vznikla, a na to, o koľko zmluvná pokuta túto škodu

¹⁷ NEPOMUCENO, T. C. C., et al. 2022. When Penalty Fails: Modeling Contractual Misincentives With Evidence From Portugal ITO Agreements. [online]. 2022, roč. 12, č. 4. [cit. 15.01.2026]. s. 2. Dostupné na: <https://doi.org/10.1177/21582440221141850>

¹⁸ VOZÁR, J. 2022. Právomoc súdu znížiť zmluvnú pokutu. In: PATAKYOVÁ, M. a kol. *Obchodný zákonník. Komentár*. 1. vydanie. Bratislava: C. H. Beck, 2022, s. 1266. ISBN 978-80-8232-018-6.

¹⁹ *Ibid.*, s. 1268.

²⁰ OVEČKOVÁ, O. 2024. Zmluvná pokuta a náhrada škody v obchodnom práve. In: *Právny obzor* [online]. 2024. Roč. 107, č. 6, s. 625. [cit. 25.12.2025]. Dostupné na: <https://doi.org/10.31577/pravnyobzor.2024.6.04>

²¹ VOZÁR, J. 2022. Právomoc súdu znížiť zmluvnú pokutu. In: PATAKYOVÁ, M. a kol. *Obchodný zákonník. Komentár*. 1. vydanie. Bratislava: C. H. Beck, 2022, s. 1267. ISBN 978-80-8232-018-6.

presahuje.²² Na rozdiel od obchodného práva, kde je škoda pevnou hranicou, v občianskom práve slúži skôr ako orientačné kritérium.

Hranica primeranosti

Pojem „*neprimerane vysoká*“ je neurčitý právny pojem, ktorý musí byť naplnený obsahom v každom jednotlivom prípade. Judikatúra a doktrína sa zhodujú, že neexistuje paušálna hranica (napríklad percentuálny strop), ktorá by automaticky znamenala neprimeranosť. Posúdenie musí vychádzať z funkcií, ktoré zmluvná pokuta v danom vzťahu plní:

a) Sankčná a preventívna funkcia: Ak je primárnym cieľom odradiť dlžníka od porušenia povinnosti, pokuta musí byť citeľná. Vtedy môže byť jej výška násobne vyššia než hodnota zabezpečeného plnenia.

b) Reparačná funkcia: Ak pokuta slúži ako paušalizovaná náhrada škody, jej primeranosť sa posudzuje v relácii k predpokladanej alebo skutočnej škode. Ak zmluvná pokuta výrazne prevyšuje skutočnú škodu, je náchylnejšia na moderáciu, pretože priznanie neúmerne vysokej sumy by viedlo k bezdôvodnému obohateniu veriteľa.²³

Najvyšší súd Slovenskej republiky²⁴ napríklad judikoval, že pre záver o neprimeranosti zmluvnej pokuty je rozhodujúca výška dohodnutej pokuty v čase porušenia povinnosti (*ex ante*), nie v čase súdneho rozhodovania. Zároveň platí, že pokutu nemožno považovať za neprimeranú len preto, že jej celková výška (napr. pri dennej sadzbe) v dôsledku dlhodobého neplnenia zo strany dlžníka narástla do značných rozmerov, pokiaľ bola samotná sadzba v čase dojednaná primeraná.²⁵

III. Koncept a taxonómia vysvetliteľnej umelej inteligencie

S rastúcim nasadením komplexných modelov umelej inteligencie (napr. hlboké neurónové siete a veľké jazykové modely) do rozhodovacích procesov, sa do popredia dostáva požiadavka na ich transparentnosť a zrozumiteľnosť. Pre odborníkov z právnej oblasti, rovnako ako bežných používateľov systémov umelej inteligencie majú kauzálne vysvetlenia formulované v prirodzenom jazyku oveľa vyššiu hodnotu než štatistické vyjadrenia miery spoľahlivosti výstupu toho-ktorého algoritmu.²⁶ Kým tradičné modely (napr. rozhodovacie stromy) sú často interpretovateľné už svojou podstatou, moderné vysoko výkonné modely fungujú ako tzv. „čierne skrinky“, kde je vzťah medzi vstupnými dátami a výstupným rozhodnutím pre človeka nejasný. Vysvetliteľná umelá inteligencia predstavuje súbor techník a metód, ktorých cieľom je sprístupniť vnútorné fungovanie týchto modelov a generovať výstupy, ktoré sú zrozumiteľné pre koncových používateľov a je zrejmé, akým postupom sa model dopracoval

²² STRAPÁČ, P. 2019. Moderačné oprávnenie súdu. In: ŠTEVČEK, M., a kol. *Občiansky zákonník II*. 2. vydanie. Praha : C. H. Beck, 2019, s. 2031. ISBN 978-80-7400-770-5.

²³ VOZÁR, J. 2022. Právomoc súdu znížiť zmluvnú pokutu. In: PATAKYOVÁ, M. a kol. *Obchodný zákonník. Komentár*. 1. vydanie. Bratislava: C. H. Beck, 2022, s. 1269. ISBN 978-80-8232-018-6.

²⁴ Uznesenie Najvyššieho súdu Slovenskej republiky, sp. zn.: 3Obdo/11/2019.

²⁵ VOZÁR, J. 2022. Právomoc súdu znížiť zmluvnú pokutu. In: PATAKYOVÁ, M. a kol. *Obchodný zákonník. Komentár*. 1. vydanie. Bratislava: C. H. Beck, 2022, s. 1270. ISBN 978-80-8232-018-6.

²⁶ RUGGERI, F., LAGIOIA, F., LIPPI, M. a TORRONI, P. *Detecting and explaining unfairness in consumer contracts through memory networks*. In: *Artificial Intelligence and Law* [online]. 2022, roč. 30. [cit. 2026-01-15]. s. 63-64. Dostupné na: <https://doi.org/10.1007/s10506-021-09288-2>

k výstupu.²⁷ V predkladanom príspevku sa zameriavame na využitie modelu Grok. Ide o veľký jazykový model vyvinutý spoločnosťou xAI, ktorý využíva architektúru transformer a bol trénovaný na rozsiahlej sade textových dát s cieľom dosahovať pokročilé schopnosti v oblasti uvažovania a kódovania.²⁸

V odbornej literatúre je kľúčové rozlišovať medzi pojmami interpretovateľnosť a vysvetliteľnosť. Interpretovateľnosť sa vzťahuje na schopnosť pochopiť príčinu rozhodnutia na základe vnútornej logiky modelu, čo je typické pre jednoduchšie modely, v ktorých človek vidí priamu kauzalitu. Naproti tomu, vysvetliteľnosť je vlastnosť, ktorá umožňuje popísať vnútorné fungovanie komplexného systému v ľudskej reči alebo prostredníctvom zrozumiteľných vizualizácií, a to často bez potreby úplného pochopenia matematického pozadia modelu. V právnom kontexte je práve vysvetliteľnosť kľúčová, pretože sudcovia a účastníci konania potrebujú rozumieť dôvodom rozhodnutia (prečo), nielen technickým parametrom (ako).²⁹

Taxonómia metód vysvetliteľnej umelej inteligencie

Pre systematické uchopenie nástrojov vysvetliteľnej umelej inteligencie v práve je vhodné využiť taxonómiu podľa Ali a kolektívu, ktorá kategorizuje metódy vysvetliteľnosti do štyroch hlavných osí: vysvetliteľnosť údajov, vysvetliteľnosť modelu, *post-hoc* vysvetliteľnosť a hodnotenie vysvetlenia.³⁰

Každá z týchto osí zohráva špecifickú úlohu pri zabezpečení toho, aby bol systém umelej inteligencie (napríklad pri posudzovaní primeranosti zmluvnej pokuty) transparentný a dôveryhodný.³¹

Vysvetliteľnosť údajov je zameraná na „surovinu“ umelej inteligencie – dáta. Ak sú vstupné dáta skreslené alebo nekvalitné, ani ten najlepší model nemôže poskytnúť spravodlivé rozhodnutie (princíp *garbage in, garbage out*). V tejto kategórii sa využíva napríklad prieskumná analýza údajov. Ide o techniku na zhrnutie hlavných charakteristík dátových súborov, často pomocou vizualizácie. V právnom kontexte to znamená analyzovať distribúciu prípadov – napríklad, či trénovacia množina obsahuje dostatok príkladov neprimeraných pokút, alebo či nie sú dáta nevyvážené v neprospech určitej skupiny dlžníkov.³² Využíva sa taktiež spísanie dokumentácie, napríklad „*dátové listy pre dátové sady*“, ktoré zaznamenávajú motiváciu, zloženie, proces zberu a odporúčané použitie dát. Tento postup je kľúčový pre možnosť spätnej kontroly, a pre informáciu o tom, na akých judikátoch sa model učil. V poslednom rade sa využíva vysvetliteľné inžinierstvo príznakov. Ide o proces výberu a transformácie premenných tak, aby boli zrozumiteľné pre človeka. Namiesto abstraktných matematických vektorov by mal model pracovať s právne relevantnými pojmami, ako je „*výška istiny*“ alebo „*doba omeškania*“.³³

²⁷ ALI, S. et al. 2023. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. In: *Information Fusion* [online]. 2023, roč. 99. [cit. 25.12.2025]. s. 2. Dostupné na: <https://doi.org/10.1016/j.inffus.2023.101805>

²⁸ xAI. 2023. *Announcing Grok*. [online]. 03.11.2023. [cit. 25.12.2025]. Dostupné na: <https://x.ai/news/grok>

²⁹ PAVLIDIS, G. Unlocking the Black Box: Analysing the EU Artificial Intelligence Act's Framework for Explainability in AI. In: *Law, Innovation and Technology* [online]. 2024, roč. 16, č. 1. [cit. 2025-12-25]. s. 296. Dostupné na: <https://doi.org/10.1080/17579961.2024.2313795>

³⁰ ALI, S. et al. 2023. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. In: *Information Fusion* [online]. 2023, roč. 99 [cit. 25.12.2025]. s. 2. Dostupné na: <https://doi.org/10.1016/j.inffus.2023.101805>

³¹ *Ibid.*

³² *Ibid*, s. 9.

³³ *Ibid.*

Pri pojme vysvetliteľnosť modelu sa zaoberáme architektúrou samotného modelu. Cieľom je buď vybrať model, ktorý je prirodzene zrozumiteľný, alebo navrhnúť komplexný model tak, aby sa dal interpretovať. Máme možnosť využiť modely tzv. „bielej skrinky“, ako napríklad rozhodovacie stromy, lineárna regresia alebo zoznamy pravidiel. Tieto modely sú transparentné, pretože ich rozhodovací proces možno priamo sledovať (napr. „ak je pokuta > 50 % istiny, potom je neprimeraná“). Nevýhodou je, že tieto modely často nedosahujú presnosť moderných neurónových sietí pri spracovaní komplexných textov.³⁴ Ďalej je možné využiť hybridné vysvetliteľné modely. Tie kombinujú vysokú presnosť „čiernych skriniek“ s interpretovateľnými komponentmi. Ako príklad možno uviesť neuro-symbolické systémy, ktoré využívajú neurónové siete na spracovanie dát, ale logické pravidlá na vyvodenie záveru, alebo modely, ktoré využívajú mechanizmy pozornosti na zvýraznenie častí textu zmluvy, ktoré boli pre rozhodnutie kľúčové.³⁵

Post-hoc vysvetliteľnosť je najčastejšie využívaná kategória v súčasnom výskume. Tieto metódy sa aplikujú na už vytrénovaný model (čiernu skrinku) a snažia sa spätne vysvetliť jeho správanie bez toho, aby menili jeho vnútornú štruktúru. Najznámejšou technikou je *Shapley Additive exPlanations* (SHAP), ktorá vychádza z teórie hier a prideluje každému vstupnému faktoru (napr. slovu v zmluve alebo výške sumy) hodnotu, ktorá vyjadruje, ako veľmi prispel k výsledku (pozitívne alebo negatívne). Ďalšou metódou je *Local Interpretable Model-agnostic Explanations* (LIME), ktorá približne určuje správanie zložitého modelu v okolí konkrétneho rozhodnutia pomocou jednoduchého lineárneho modelu.³⁶ Na základe zistení Mahajana a kolektívu vnímame SHAP ako spoľahlivý nástroj, pretože autori jeho výsledky overili aj inými metódami vysvetliteľnosti, pričom zistili minimálne odchýlky.³⁷ Uvedené metódy (SHAP a LIME) sme si preto zvolili ako nástroj na overenie stanovených výskumných predpokladov v ďalšej kapitole.

Posledná os rieši kritický problém, a teda, ako zistíme, či vysvetlenie, ktoré umelá inteligencia poskytla, je správne? Nie je postačujúce, aby vysvetlenie existovalo, ale musí byť aj správne. V tejto kategórii sa využíva najmä výpočtové hodnotenie, ktoré zahŕňa metriky ako vernosť (do akej miery vysvetlenie zodpovedá skutočnému správaniu modelu), a stabilitu (či malé zmeny vo vstupe nespôsobujú radikálne odlišné vysvetlenia). Testuje sa napríklad odstránením príznakov, ktoré model označil za dôležité, a sledovaním, či sa zmení predikcia.³⁸ Využíva sa taktiež samotné hodnotenie používateľom. Toto hodnotenie sa zameriava na psychologické aspekty, ako je dôvera, spokojnosť a pochopenie. V právnom kontexte je kľúčové zistiť, či sudca alebo advokát na základe vysvetlenia dokáže správne posúdiť, či je odporúčanie umelej inteligencie v súlade s právnymi normami. Dôležitá je nielen presnosť, ale aj forma prezentácie vysvetlenia.³⁹

³⁴ *Ibid.*, s. 11.

³⁵ *Ibid.*, s. 13.

³⁶ *Ibid.*, s. 15.

³⁷ MAHAJAN, S., AGARWAL, R. a GUPTA, M. 2025. *Algorithmic Bias Under the EU AI Act: Compliance Risk, Capital Strain, and Pricing Distortions in Life and Health Insurance Underwriting. Risks* [online]. 2025, roč. 13, č. 9. [cit. 2026-01-15] s. 8. Dostupné na: <https://doi.org/10.3390/risks13090160>

³⁸ ALI, S. et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion* [online]. 2023, roč. 99. [cit. 25.12.2025]. s. 17. Dostupné na: <https://doi.org/10.1016/j.inffus.2023.101805>

³⁹ *Ibid.*, s. 18.

IV. Praktický prípad

Pre účely praktického otestovania teoreticko-analytickej časti príspevku sme si vytvorili prípadovú štúdiu. Do veľkého jazykového modelu Grok, verzia 4.1, nastavenie odpovede „automaticky“ (model sám zvolí, či bude odpovedať „rýchlo“ alebo ako „expert“) (ďalej len „Grok“), sme dňa 30.12.2025 zadali prompt:

„Žalobca - objednávateľ zo zmluvy o dielo podal žalobu voči žalovanému - zhotoviteľovi o zaplatenie zmluvnej pokuty 500 eur za každý deň omeškania so splnením povinnosti - odstránenie väd diela. Túto zmluvnú pokutu uplatnil za cca 457 dní ku dňu podania žaloby (a ďalej narastá). Hodnota celého diela je cca 1 400 000 eur. Dielo však bolo odovzdané ako celok a boli vytknuté vady. Tieto vady majú spolu hodnotu cca 5 000 eur. V zmluve o dielo je ustanovenie, ktoré umožňuje v prípade nesplnenia povinnosti odstrániť vady diela zhotoviteľom v primeranej lehote, aby objednávateľ odstránil vady sám alebo dal vady odstrániť tretej osobe na náklady zhotoviteľa. Tento mechanizmus objednávateľ nevyužil. Podal žalobu po 457 dňoch. Žalovaný sa bráni, že prebiehajú rokovania o odstránení väd, že termín ich odstránenia bol posúvaný kvôli nepriaznivému počasiu a že takáto zmluvná pokuta je neprimerane vysoká, má za cieľ majetkový prospech žalobcu a nie splnenie povinnosti, je v rozpore s poctivým obchodným stykom a preto žiada, aby buď súd žalobu zamietol v celom rozsahu, alebo aby zmluvnú pokutu zmoderoval na minimum. Žalobca tvrdí, že nejde o majetkový prospech, ale o vynútenie splnenia povinnosti. Rozhodni ako sudca, či je pokuta neprimerane vysoká, alebo je v poriadku. Ak je neprimerane vysoká, rozhodni, či ju moderuješ alebo žalobu zamietáš pre rozpor s poctivým obchodným stykom, pričom zneužitie práva nepožíva právnu ochranu. Čerpaj zo slovenského právneho poriadku aktuálne účinného, a aj z judikatúry.“

Grok poskytol nasledovný výstup (doslovná citácia relevantnej časti výstupu):

„Neprimeranosť pokuty: Áno, pokuta je neprimerane vysoká. Denná sadzba 500 EUR nie je sama o sebe extrémna, ale jej kumulatívna výška (228 500 EUR) výrazne presahuje hodnotu zabezpečovanej povinnosti (5 000 EUR) a škodu (nepreukázaná vyššia škoda). To vedie k neúmernému obohateniu žalobcu a porušuje zásady poctivého styku, najmä pri pasivite žalobcu a prebiehajúcich rokovaniach.

Moderácia vs. zamietnutie: Nezamietam žalobu v celom rozsahu, pretože pokuta nie je neplatná od počiatku (dojednaná písomne, v súlade so zmluvou). Namiesto toho využívam moderačné právo podľa § 301 ObZ a znižujem pokutu na primeranú úroveň: 10 000 EUR (dvojnásobok hodnoty väd, čo zohľadňuje sankčnú funkciu, zavinenie žalovaného a minimalizuje prospech žalobcu). Toto zodpovedá judikatúre (napr. zníženie na ekvivalent ceny zabezpečovanej povinnosti v rozsudku NS SR sp. zn. 1 Obdo V 93/2004).

Ďalšie: Žalobca má právo na náhradu škody presahujúcej moderovanú pokutu, ak ju preukáže (§ 302 ObZ). Žalovaný má zaplatiť moderovanú pokutu do 15 dní od právoplatnosti rozhodnutia. Trovy konania znáša žalobca (75 %) a žalovaný (25 %), vzhľadom na čiastočný úspech.

Rozhodnutie je odôvodnené a podlieha odvolaniu do 15 dní.“

Následne sme do veľkého jazykového modelu Grok zadali prompt:

„Teraz mi toto svoje rozhodnutie vysvetli pomocou SHAP a LIME metód.“

Grok poskytol nasledovný výstup (doslovná citácia výstupu):

Vysvetlenie rozhodnutia pomocou SHAP a LIME

Aby som vysvetlil moje simulované súdne rozhodnutie (moderácia zmluvnej pokuty z cca 228 500 EUR na 10 000 EUR) pomocou metód explainable AI ako SHAP a LIME, musel som to prispôbiť. Rozhodnutie nie je výstupom reálneho ML modelu, ale logickým výstupom na základe právnych faktorov. Preto som vytvoril jednoduchý simulovaný "model" v Pythone, ktorý napodobňuje rozhodovací proces: berie ako vstupné príznaky (features) kľúčové faktory z prípadu a predpovedá finálnu výšku pokuty po moderácii (na základe logiky z ObZ/OZ a judikatúry, ako cap na násobok hodnoty vád, zľavy za mitigujúce okolnosti atď.).

Simulovaný model:

Features (vstupné faktory):

vad_value: Hodnota vád (5 000 EUR).

daily_penalty: Denná sadzba pokuty (500 EUR).

days_delay: Počet dní omeškania (457).

work_value: Hodnota celého diela (1 400 000 EUR).

negotiations: Prebiehajúce rokovania (1 = áno).

weather_impact: Vplyv nepriaznivého počasia (1 = áno).

plaintiff_passivity: Pasivita žalobcu (nevyžitie alternatívy na odstránenie vád; 1 = áno).

*Model predpovedá finálnu pokutu ako: základ = min (daily * days, vad_value * faktor_z_work_value), potom zľavy za mitigujúce faktory (rokovania: -20 %, počasie: -10 %, pasivita: -30 %). Toto odráža judikatúru (napr. zníženie pri neprimeranosti a zneužití).*

Predikcia pre tento prípad: cca 8 568 EUR (blízko mojej moderácie na 10 000 EUR; rozdiel kvôli zjednodušeniu).

Teraz aplikujem SHAP a LIME na vysvetlenie, prečo model (a tým pádom rozhodnutie) dospel k nižšej pokute. Použil som manuálnu implementáciu (pretože plné knižnice nie sú dostupné), ale princípy sú rovnaké.

SHAP (SHapley Additive exPlanations)

SHAP priraduje "spravodlivý" príspevok každému faktoru k celkovej predikcii (založené na teórii hier: ako by sa predikcia zmenila, keby sme pridali/odobrali faktor v rôznych kombináciách). Začína od "nulovej" predikcie (0) a sčítava príspevky na finálnu hodnotu (8 568 EUR).

Výsledné SHAP values (príspevky):

vad_value (5 000 EUR): +2 453 EUR → Silne pozitívny vplyv; hodnota vád je základom pre cap pokuty (podľa § 301 ObZ, pokuta nesmie neprimerane presahovať zabezpečenú povinnosť).

daily_penalty (500 EUR): +2 453 EUR → Pozitívny; vysoká denná sadzba zvyšuje základnú pokutu, ale je limitovaná.

days_delay (457 dní): +2 453 EUR → Pozitívny; dlhé omeškanie by normálne zvyšovalo pokutu, ale v kontexte judikatúry (napr. NS SR sp. zn. 3 Obdo 11/2019) sa moderuje pri neúmernom narastaní.

work_value (1 400 000 EUR): +636 EUR → Mierne pozitívny; veľká hodnota diela zvyšuje kontext, ale nie disproporčne (pokuta sa viaže na vady, nie celok).

negotiations (áno): -273 EUR → Negatívny; prebiehajúce rokovania znižujú zavinenie (podľa judikatúry, napr. NS SR sp. zn. 5 Obo 56/2007), čo vedie k moderácii.

weather_impact (áno): -129 EUR → Negatívny; vonkajšie faktory ako počasie ospravedľujú posun termínov, znižujú pokutu.

plaintiff_passivity (áno): -433 EUR → Silne negatívny; pasivita žalobcu (nevyužitie § 536 OZ na odstránenie vád sám) indikuje zneužitie práva (§ 265 ObZ, § 3 OZ), čo je kľúčový dôvod na výraznú moderáciu (žalobca mohol škodu minimalizovať).

Interpretácia: Súčet týchto príspevkov dáva finálnu predikciu. Pozitívne faktory (základné prvky pokuty) tlačia nahor, ale negatívne (mitigujúce okolnosti) výrazne znižujú výšku, čo vysvetľuje, prečo som pokutu zmoderoval – ide o vyváženie neprimeranosti a poctivého styku, nie úplné zamietnutie.

LIME (Local Interpretable Model-agnostic Explanations)

LIME vytvára lokálny jednoduchý (lineárny) model okolo konkrétnej inštancie (nášho prípadu) tým, že generuje perturbované verzie dát (malé zmeny faktorov) a fitne regresiu na predikcie. Tento surrogátny model vysvetľuje, ako faktory lokálne ovplyvňujú rozhodnutie.

Výsledné koeficienty LIME (váhy v lineárnom modeli: predikcia \approx intercept + súčet(koeficient * hodnota faktora)):

vad_value: +1.71 → Pozitívny; každé ďalšie euro v hodnote vád zvyšuje pokutu o cca 1.71 EUR (odráža proporčnosť k zabezpečovanej povinnosti).

daily_penalty: +0.10 → Mierne pozitívny; zvyšuje základ.

days_delay: -0.07 → Negatívny; v lokálnom okolí dlhé dni vedú k moderácii (kvôli neprimeranosti narastania, ako v NS SR rozsudkoch o kumulatívnej výške).

work_value: +0.003 → Takmer neutrálny; veľkosť diela má malý lokálny vplyv.

negotiations: -902 → Silne negatívny; prítomnosť rokovaní výrazne znižuje pokutu.

weather_impact: -902 → Silne negatívny; počasie ako ospravedlnenie.

plaintiff_passivity: -902 → Silne negatívny; pasivita je kľúčovým faktorom zneužitia, čo lokálne dominuje zníženiu (v súlade so § 39 OZ a judikatúrou o dobrých mravoch).

Intercept (základná hodnota): -902 EUR.

Interpretácia: Lineárny model ukazuje, že v okolí tohto prípadu sú mitigujúce faktory (rokovania, počasie, pasivita) dominantné v znižovaní pokuty, zatiaľ čo základné faktory (vady, sadzba) ju zvyšujú len mierne. To vysvetľuje moje rozhodnutie moderovať namiesto zamietnuť – pokuta nie je neplatná, ale neprimeraná, a mitigujúce prvky ju "ťahajú dole" na úroveň dvojnásobku vád (ako v judikatúre NS SR sp. zn. 1 Obdo V 93/2004).

Grok síce rozhodol, rozhodnutie odôvodnil ale je možné toto odôvodnenie považovať za postačujúce? Najmä v oblasti justície? A je toto rozhodnutie správne? V tomto kontexte si dovoľíme citovať Uznesenie Najvyššieho súdu Slovenskej republiky sp. zn.: 3Obdo/11/2019:

*„Rovnako tak dovolací súd podporne poukazuje aj na rozhodnutia Najvyššieho súdu Českej republiky, z ktorých vyplývajú nasledovné závery: V rozhodovanej veci konečná výška zmluvnej pokuty celkom závisela na dĺžke doby, v priebehu ktorej žalovaní nesplnili povinnosť, ktorú sami zmluvne prevzali. Výsledná výška zmluvnej pokuty je tak dôsledkom relatívne dlhej doby, ktorá uplynula odo dňa splatnosti zabezpečenej pohľadávky. Z pohľadu primeranosti zmluvnej pokuty je na mieste hodnotiť inak zmluvnú pokutu dojednanú vo forme pevne stanovenej čiastky a zmluvnú pokutu dojednanú vo forme určitej sadzby za určené časové obdobie. Neprimeranosť zmluvnej pokuty **nemožno** posudzovať len z jej celkovej výšky, ak je **dôsledkom dlhodobého omeškania** a s tým spojeným navyšovaním o inak primeranú „dennú sadzbu“ zmluvnej pokuty. **Opačný záver je neprijateľný**, lebo by vo svojom dôsledku zvýhodňoval dlžníka (čím dlhšie by dlžník svoje povinnosti neplnil, tým viac by bol zvýhodnený pri posudzovaní prípadnej neprimeranosti výšky zmluvnej pokuty) a **znamenal by spochybnenie funkcií, ktoré má zmluvná pokuta plniť**. Žalovaní mohli zabrániť rastu celkovej výšky zmluvnej pokuty tým, že by v čo najkratšom čase splnili svoju povinnosť vrátiť dlžnú sumu, na ktorú sa zmluvne zaviazali s poukazom na dobré mravy.“*

Je otázne, ako by v našom vykonštruovanom prípade rozhodol sudca, ale uvedený judikát celkom zrejme oponuje rozhodnutiu jazykového modelu Grok. Keď analyzujeme uvedené odôvodnenie rozhodnutia pomocou SHAP a LIME metód, je možné konštatovať, že model Grok vzal do úvahy celkom bezpredmetné atribúty, ako existenciu rokovaní (silne negatívny vplyv – výrazne znižuje pokutu) či nepriaznivé počasie pri odstraňovaní vád diela. Uvedené faktory nemajú v skutočnosti žiaden vplyv na primeranosť zmluvnej pokuty. Model Grok sa napríklad nevysporiadal ani so skutočnosťou, či hodnota odstránenia vád diela (cca 5000 eur) je žalovaným len tvrdená, alebo aj preukázaná. Na dôvažok

uvádzame, že napríklad judikát v rozhodnutí: „*Toto zodpovedá judikatúre (napr. zníženie na ekvivalent ceny zabezpečovanej povinnosti v rozsudku NS SR sp. zn. 1 Obdo V 93/2004).*“ si model Grok vyhľadal správne, dokonca tento judikát aj existuje a model ho správne interpretoval.

Naproti tomu je potrebné poukázať na skutočnosť, že interpretácia ustanovenia § 302 Obchodného zákonníka modelom Grok je nesprávna, a je to tzv. halucinácia. Modely generatívnej umelej inteligencie sú náchylné na výskyt tzv. halucinácií. Tu možno uviesť ako ilustračný príklad situáciu, keď advokáti v Spojených štátoch amerických použili veľký jazykový model na vypracovanie písomného podania na súd. Toto podanie obsahovalo odkazy na judikatúru (precedensy), ktoré v skutočnosti neexistujú. Takéto podanie je pochopiteľne nepoužiteľné a nemá žiadnu relevanciu.⁴⁰ Podľa uvedeného ustanovenia odstúpenie od zmluvy nemá vplyv na nárok na zaplatenie zmluvnej pokuty. Model Grok však uvádza: „*Žalobca má právo na náhradu škody presahujúcej moderovanú pokutu, ak ju preukáže (§ 302 ObZ).*“ Ďalej je potrebné poukázať na časť výstupu modelu Grok, ktorý je taktiež nesprávny: „*dlhé omeškanie by normálne zvyšovalo pokutu, ale v kontexte judikatúry (napr. NS SR sp. zn. 3 Obdo 11/2019) sa moderuje pri neúmernom narastaní.*“ Z citovaného judikátu vyplýva práve opak toho, čo tvrdí model Grok, ako sme už vyššie v citácii daného rozhodnutia uviedli.

Ďalej považujeme za potrebné uviesť časť výstupu modelu Grok: „*prebiehajúce rokovania znižujú zavinenie (podľa judikatúry, napr. NS SR sp. zn. 5 Obo 56/2007), čo vedie k moderácii.*“ Toto rozhodnutie je opäť vyhľadané správne, existujúce a relevantné k prípadu, avšak v rozhodnutí sa vôbec nespomínajú rokovania ani ich vplyv na moderáciu zmluvnej pokuty, toto tvrdenie je úplne vymyslené. Taktiež považujeme za potrebné podotknúť, že zavinenie je v kontexte zmluvnej pokuty bezpredmetné.

Ďalšia časť výstupu modelu Grok: „*Silne negatívny; pasivita žalobcu (nevyužitie § 536 OZ na odstránenie väd sám) indikuje zneužitie práva (§ 265 ObZ, § 3 OZ), čo je kľúčový dôvod na výraznú moderáciu (žalobca mohol škodu minimalizovať).*“ je prevažne nesprávna. Ustanovenie § 536 Občianskeho zákonníka, ako model uvádza, obsahuje právnu úpravu poukážky. Ustanovenie § 536 Obchodného zákonníka obsahuje síce právnu úpravu zmluvy o dielo, avšak ustanovenia o vadách diela sú uvedené v ustanoveniach § 560 a nasl., a nevyplýva z nich možnosť objednávateľa odstrániť vady sám alebo prostredníctvom tretej osoby, toto vyplýva len ako zmluvne dojednané ustanovenie *inter partes* z nášho zadania. Ustanovenie § 265 Obchodného zákonníka, ktoré Grok identifikoval správne, ustanovuje že výkon práva v rozpore so zásadou poctivého obchodného styku nepožíva právnu ochranu. Uvedené ustanovenie však nemá vplyv na moderáciu neprimerane vysokej zmluvnej pokuty, ale predstavuje dôvod na zamietnutie žaloby.

Poslednou časťou výstupu modelu Grok, na ktorú je potrebné upriamiť pozornosť je: „*pokuta nie je neplatná, ale neprimeraná, a mitigujúce prvky ju „ťahajú dole“ na úroveň dvojnásobku väd (ako v judikatúre NS SR sp. zn. 1 Obdo V 93/2004).*“ uvedená interpretácia nie je celkom správna, keďže v danom judikáte bola zmluvná pokuta moderovaná na hodnotu zabezpečovanej povinnosti, nie na jej dvojnásobok. Uvedené bez akýchkoľvek pochybností preukazuje, že veľké jazykové modely, aj za

⁴⁰ GSTREIN, O. J., HALEEM, N. a ZWITTER, A. General-purpose AI regulation and the European Union AI Act. In: *Internet Policy Review* [online]. 2024, roč. 13, č. 3. [cit. 15.01.2026]. s. 2. Dostupné na: <https://doi.org/10.14763/2024.3.1790>

využitia metód vysvetliteľnej umelej inteligencie v žiadnom prípade nie sú pripravené byť tzv. „samosudcami“, a vždy je potrebná ľudská kontrola ex post.

Na základe uvedeného možno konštatovať, že interpretácia pojmov „dobré mravy“ alebo „zásada poctivého obchodného styku“, ktoré sú pri moderácii zmluvnej pokuty kľúčové, si vyžaduje axiologické hodnotenie, ktorého súčasné veľké jazykové modely nie sú schopné. Keďže modely strojového učenia sú v podstate neprehľadné systémy, ich nesprávne rozhodnutia v tak citlivých oblastiach, ako je napríklad právo, môžu mať závažné a rozsiahle následky.⁴¹ Tieto modely fungujú na pravdepodobnostnom princípe a nemajú skutočné porozumenie pre sociálne normy, etiku či „zdravý rozum“, ktoré sú nevyhnutné na vyplnenie medzier v práve a na aplikáciu neurčitých právnych pojmov v nepredvídaných situáciách. Túto úlohu plní judikatúra, ktorá je tvorená sudcami, teda človekom. Z uvedeného je zrejmé, že umelá inteligencia aj za použitia metód vysvetliteľnej umelej inteligencie nie je na postačujúcej úrovni. Nástroje umelej inteligencie na analýzu či predikciu súdnych rozhodnutí by mali v súdnictve fungovať výlučne ako doplnkové systémy plniace funkciu pomôcky sudcu, a nie ako jeho plnohodnotná náhrada.⁴² Preto považujeme za nevyhnutné zachovať princíp „human-in-the-loop“. Umelá inteligencia nesmie fungovať ako autonómny sudca, ale len ako pomocník, ktorý identifikuje relevantnú judikatúru a navrhne argumentačnú líniu, no konečné vyváženie záujmov a rozhodnutie o spravodlivej výške moderácie musí vykonať človek, ktorý za rozhodnutie nesie morálnu a právnu zodpovednosť.

V. Záver

Cieľom predkladaného príspevku bolo preskúmať, či a do akej miery môžu metódy vysvetliteľnej umelej inteligencie prispieť k objektivizácii posudzovania neprimeranosti zmluvnej pokuty. Realizovaný výskum, kombinujúci teoretickú analýzu s praktickým experimentom na modeli Grok, priniesol nasledovné zistenia vo vzťahu k stanovenej výskumnej otázke a hypotézam.

Na otázku, či môžu metódy vysvetliteľnej umelej inteligencie prispieť k objektivizácii vágneho právneho pojmu pri zachovaní požiadavky na transparentnosť, je nutné odpovedať zdržanlivo. Výskum ukázal, že umelá inteligencia dokáže prostredníctvom metód ako SHAP identifikovať faktory, ktoré vplyvajú na rozhodnutie (napr. výška škody, dĺžka omeškania). Avšak, v procese „objektivizácie“ umelej inteligencie často zlyháva pri správnom právnom vyhodnotení váhy týchto faktorov. Ako ukázal náš experiment, model pripisoval silný negatívny (zmierňujúci) vplyv okolnostiam, ktoré sú v režime objektívnej zodpovednosti podľa Obchodného zákonníka právne irelevantné (napr. nepriaznivé počasie). Vysvetliteľná umelá inteligencia teda prispieva k technickej transparentnosti (vieme, prečo model rozhodol), ale nezaručuje právnu správnosť a spravodlivosť tohto odôvodnenia.

Vo vzťahu k stanoveným vedeckým predpokladom uvádzame, že prvú hypotézu možno potvrdiť len čiastočne. Experiment preukázal, že model dokázal extrahovať a kvantifikovať vplyv jednotlivých faktorov na výslednú sumu pokuty. Dokázal správne identifikovať, že nepomer medzi výškou pokuty a

⁴¹ LI, Y., et al. 2022. Fairness Measures of Machine Learning Models in Judicial Penalty Prediction. In: *Journal of Internet Technology* [online]. 2022, roč. 23, č. 5. [cit. 15.01.2026]. s. 1. Dostupné na: <https://jit.ndhu.edu.tw/article/view/2771/2794>

⁴² DINA, N. Z., RAVANA, S. D. a IDRIS, N. 2025. Legal Judgment Prediction using Natural Language Processing and Machine Learning Methods: A Systematic Literature Review. In: *SAGE Open* [online]. 2025, roč. 15, č. 2. [cit. 15.01.2026]. s.1. Dostupné na: <https://doi.org/10.1177/21582440251329663>

hodnotou diela je signálom pre moderáciu. Avšak presnosť identifikácie právne relevantných faktorov bola nízka. Model zahrnul do svojej úvahy aj faktory, ktoré by súd v obchodnoprávnom spore nemal brať do úvahy ako dôvod pre moderáciu, čím skreslil výsledok.

Druhú hypotézu je potrebné na základe výsledkov výskumu vyvrátiť. Experiment jasne ukázal, že umelá inteligencia bez ľudskej korekcie nie je schopná vykonávať komplexné axiologické hodnotenie, ktoré je potrebné pre posúdenie vágnych pojmov ako sú „dobré mravy“ alebo „*zásady poctivého obchodného styku*“. Model Grok dospel k záverom, ktoré boli v priamom rozpore s citovanou judikatúrou Najvyššieho súdu SR, právnymi predpismi, halucinoval a vymýšľal si závery z judikatúry a právnych predpisov, ktoré boli síce vyhladané správne, ale interpretácia judikatúry a právnych predpisov bola často buď nesprávna, alebo úplne vymyslená. Absencia ľudského prvku by v tomto prípade viedla k nezákonnému rozhodnutiu, nakoľko umelá inteligencia nedokázala rozlíšiť medzi skutkovou a právnou relevanciou vstupných dát.

Cieľ predkladaného príspevku považujeme za splnený. Splnenie cieľa podľa nášho názoru spočíva v preukázaní, že hoci vysvetliteľná umelá inteligencia ponúka zaujímavé nástroje na analýzu dát, v súčasnom štádiu technologického vývoja nemôže nahradiť sudcovskú úvahu. Najvhodnejším modelom pre budúcnosť sa javí hybridný systém „*human-in-the-loop*“, kde umelá inteligencia slúži ako analytický nástroj na rýchle spracovanie dát a identifikáciu anomálií, no konečné posúdenie primeranosti a spravodlivosti sankcie musí ostať v rukách človeka, sudcu. Domnievame sa, že iba týmto spôsobom možno zabezpečiť, aby technologický pokrok nešiel na úkor spravodlivosti a právnej istoty.

Použité informačné zdroje:

1. ALI, S. et al. 2023. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. In *Information Fusion* [online]. 2023, roč. 99 [cit. 25.12.2025]. Dostupné na: <https://doi.org/10.1016/j.inffus.2023.101805>.
2. ALSHBOUL, O., et al. 2022. *Prediction Liquidated Damages via Ensemble Machine Learning Model: Towards Sustainable Highway Construction Projects*. *Sustainability* [online]. 2022. Roč. 14, č. 15. [cit. 2026-01-15]. Dostupné na: <https://doi.org/10.3390/su14159303>
3. ATKINSON, K., BENCH-CAPON, T. a BOLLEGALA, D. 2020. *Explanation in AI and law: Past, present and future*. *Artificial Intelligence* [online]. 2020. Roč. 289. [cit. 25.12.2025]. Dostupné na: <https://doi.org/10.1016/j.artint.2020.103387>
4. BENNETOT, A. et al. 2024. A Practical Tutorial on Explainable AI Techniques. In *ACM Computing Surveys* [online]. 2024, roč. 57, č. 2 [cit. 25.12.2025]. Dostupné na: <https://doi.org/10.1145/3670685>.
5. BROŽEK, B., FURMAN, M., JAKUBIEC, M. a KUCHARZYK, B. The black box problem revisited. Real and imaginary challenges for automated legal decision making. 2024. In: *Artificial Intelligence and Law* [online]. 2024. Roč. 32. [cit. 25.12.2025]. Dostupné na: <https://doi.org/10.1007/s10506-023-09356-9>
6. DINA, N. Z., RAVANA, S. D. a IDRIS, N. 2025. Legal Judgment Prediction using Natural Language Processing and Machine Learning Methods: A Systematic Literature Review. In: *SAGE Open* [online]. 2025, roč. 15, č. 2. [cit. 15.01.2026]. Dostupné na: <https://doi.org/10.1177/21582440251329663>
7. GSTREIN, O. J., HALEEM, N. a ZWITTER, A. General-purpose AI regulation and the European Union AI Act. In: *Internet Policy Review* [online]. 2024, roč. 13, č. 3. [cit. 15.01.2026]. Dostupné na: <https://doi.org/10.14763/2024.3.1790>
8. KIM, S., YI, S. a PARK, S.-P. 2025. Prioritizing challenges in AI adoption for the legal domain: A systematic review and expert-driven AHP analysis. In *PLOS One* [online]. 2025, roč. 20, č. 6 [cit. 15.01.2026]. Dostupné na: <https://doi.org/10.1371/journal.pone.0326028>
9. LI, Y., et al. 2022. Fairness Measures of Machine Learning Models in Judicial Penalty Prediction. In: *Journal of Internet Technology* [online]. 2022, roč. 23, č. 5. [cit. 15.01.2026]. Dostupné na: <https://jit.ndhu.edu.tw/article/view/2771/2794>
10. LINNA, E. a LINNA, T. *Judicial Requirements for Generative AI in Legal Reasoning*. [online]. 2025. s. 5 [cit. 25.12.2025]. Dostupné na: <https://doi.org/10.48550/arXiv.2508.18880>
11. LÖFFLER, C., MARTÍNEZ FREILE, A. a REY PIZARRO, T. 2025. *Predicting potentially abusive clauses in Chilean terms of services with natural language processing*. [online]. 2025. [cit. 25.12.2025]. Dostupné na: <https://arxiv.org/abs/2502.00865>
12. MAHAJAN, S., AGARWAL, R. a GUPTA, M. 2025. *Algorithmic Bias Under the EU AI Act: Compliance Risk, Capital Strain, and Pricing Distortions in Life and Health Insurance Underwriting*. *Risks* [online]. 2025, roč. 13, č. 9. [cit. 2026-01-15]. Dostupné na: <https://doi.org/10.3390/risks13090160>
13. MONTAGNANI, M. L., NAJJAR, M.C. a DAVOLA, A. 2024. The EU Regulatory approach(es) to AI liability, and its Application to the financial services market. In *Computer Law & Security Review* [online]. 2024, roč. 53. [cit. 2026-01-15]. Dostupné na: <https://doi.org/10.1016/j.clsr.2024.105984>.
14. NEPOMUCENO, T. C. C., et al. 2022. *When Penalty Fails: Modeling Contractual Misincentives With Evidence From Portugal ITO Agreements*. [online]. 2022, roč. 12, č. 4. [cit. 15.01.2026]. Dostupné na: <https://doi.org/10.1177/21582440221141850>
15. OVEČKOVÁ, O. 2024. Zmluvná pokuta a náhrada škody v obchodnom práve. In *Právny obzor* [online]. Roč. 107, č. 6 [cit. 2025-12-25]. Dostupné na: <https://doi.org/10.31577/pravnyobzor.2024.6.04>.

16. PAVLIDIS, G. Unlocking the Black Box: Analysing the EU Artificial Intelligence Act's Framework for Explainability in AI. In: *Law, Innovation and Technology* [online]. 2024, roč. 16, č. 1. [cit. 2025-12-25]. Dostupné na: <https://doi.org/10.1080/17579961.2024.2313795>
17. REKHVIASHVILI, N. a KHARITONASHVILI, N. 2024. Judicial Reduction of Penalties in Georgian Civil Law: Balancing Debtor and Creditor Rights. In: *Rechtsidee* [online]. Roč. 12, č. 2, 2024. [cit. 25.12.2025]. Dostupné na: <https://doi.org/10.21070/jjhr.v12i2.1031>
18. RUGGERI, F., LAGIOIA, F., LIPPI, M. a TORRONI, P. Detecting and explaining unfairness in consumer contracts through memory networks. In: *Artificial Intelligence and Law* [online]. 2022, roč. 30. [cit. 2026-01-15]. Dostupné na: <https://doi.org/10.1007/s10506-021-09288-2>
19. SIINO, M. et al. 2025. *Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches*. [online] 2025 [cit. 25.12.2025]. Dostupné na: <https://ieeexplore.ieee.org/document/10850911>.
20. ŠTEVČEK, M. a kol. *Občiansky zákonník II*. 2. vydanie. Praha : C. H. Beck, 2019. ISBN 978-80-7400-770-5.
21. Uznesenie Najvyššieho súdu Slovenskej republiky, sp. zn.: 3Obdo/11/2019.
22. PATAKYOVÁ, M. a kol. *Obchodný zákonník : Komentár*. 1. vydanie. Bratislava : C. H. Beck, 2022. ISBN 978-80-8232-018-6.
23. xAI. 2023. *Announcing Grok*. [online]. 03.11.2023. [cit. 25.12.2025]. Dostupné na: <https://x.ai/news/grok>
24. ZHANG, H., DOU, Z., ZHU, Y. a WEN, J.-R. 2023. Contrastive Learning for Legal Judgment Prediction. In: *ACM Transactions on Information Systems* [online]. 2023, roč. 41, č. 4. [cit. 2026-01-15]. s. 2. Dostupné na: <https://doi.org/10.1145/3580489>