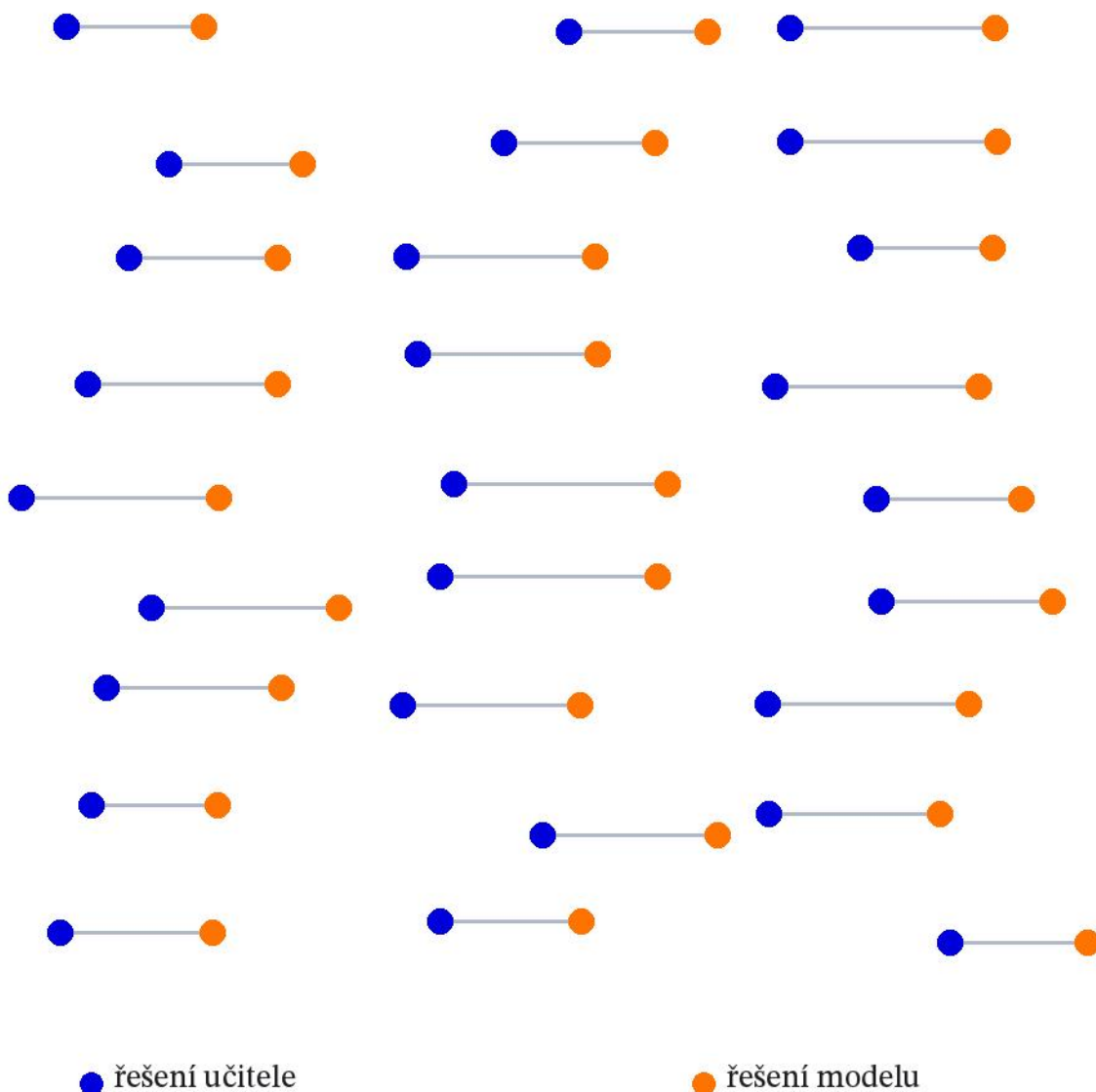


Náročné chování žáků v době AI

Kazuistiky z praxe a jazykové modely
v pedagogickém výzkumu



Habilitační práce
(monografie)

Brno 2026 Jan Nehyba

MASARYKOVA UNIVERZITA

Pedagogická fakulta

Náročné chování žáků v době AI

Kazuistiky z praxe a jazykové modely v pedagogickém
výzkumu

Habilitační práce

Obor řízení: Pedagogika

Brno 2026

Jan Nehyba

Obsah

Prohlášení	1
Prohlášení o využití umělé inteligence	3
Abstrakt	5
Abstract	7
Poděkování	9
Poznámka k publikacím	11
1. Úvod	13
1.1 Problém	13
1.2 Projekt Edustories a data knihy	14
1.3 Cíl a výzkumné otázky	15
1.4 Struktura knihy	15
1.5 Dřívější výstupy a dostupnost dat	16
Východiska	17
2. Teoretická východiska	19
2.1 Náročné chování žáků: vymezení a český kontext	19
Vymezení pojmu	19
Kázeň v české pedagogické tradici	20
Náročné chování jako profesní problém	21
Řízení třídy jako výzkumný rámec	21
Kazuistika jako pramen	22
2.2 Přístupy k řešení náročného chování	23
Pozitivní behaviorální intervence a podpora (PBIS)	24
Sociálně-emoční učení (SEL)	25
Nenásilná komunikace (NVC)	26
Restorativní praxe	28
Srovnání a vztah ke knize	29
2.3 Repertoáry učitelů a učňovství pozorováním	30
Repertoár jako jednotka popisu	30

Učňovství pozorováním	30
Repertoár nejmenšího odporu	31
Osy popisu řešení	32
2.4 Velké jazykové modely v pedagogickém výzkumu	33
Od tématu k nástroji	33
Model v roli anotátora	33
Model v roli posuzovatele a generátora	34
Jazyková a doménová mezera přenosu	35
Český kontext	35
2.5 Průsečík obou linií a výzkumná mezera	36
3. Projekt Edustories: kontext vzniku dat	39
3.1 Východisko: proč sbírat kazuistiky náročného chování	39
3.2 Projekt Edustories	40
3.3 Sběr kazuistik: zrcadlová šablona	41
3.4 Od textů ke korpusu: extrakce, anonymizace, publikace	42
3.5 Tři anotační etapy	44
4. Metodologie	45
4.1 Korpus jako časová řada řezů	45
4.2 Tři anotační etapy a jejich zaučování	46
Třetí etapa	46
4.3 Od textu k datům	47
4.4 Kódovací schémata a řezy dat	48
4.5 Multiplicita anotací	49
4.6 Známa omezení dat	49
4.7 Etika	50
4.8 Reprodukovatelnost	50
Empirické studie	53
5. Studie 1: Korpus a jeho obsah	55
5.1 Rozsah a růst korpusu	55
5.2 Základní charakteristiky kazuistik	57
5.3 Co kazuistiky zachycují: chování, řešení, dopady	58
5.4 Jak se řešení pojí s chováním	62
5.5 Dopad podle typu řešení	64
5.6 Shrnutí	66
6. Studie 2: Jazykový model jako anotátor	67
6.1 Východiska	67
6.2 Metoda	68
Kódovací schéma a lidské kódování	68

Modelové kódování a srovnávací vzorek	68
Míry shody	69
6.3 Shoda mezi lidskými anotátorkami	69
6.4 Shoda jazykového modelu s lidmi	71
Překryv s příklady v instrukci	71
Shoda po kategoriích	72
Záměny mezi kategoriemi	72
6.5 Kvalitativní pohled na vybranou neshodu	73
6.6 Shrnutí	74
7. Studie 3: Řešení generovaná umělou inteligencí versus řešení učitelů	77
7.1 Východiska	77
7.2 Metoda	78
Materiál	78
Hodnocení	78
Analýza	79
7.3 Spolehlivost hodnocení	80
7.4 Výsledky	81
Deskriptivní obraz	81
Primární modely	83
Rozpoznávání rámců	84
Citlivostní analýza	85
7.5 Shrnutí	87
8. Studie 4: Kvalita kazuistik, spolehlivost jejího posouzení a predikce úspěchu řešení	89
8.1 Východiska	89
8.2 Metoda	90
Materiál a anotátorky	90
Nástroje	91
Průběh etapy a rekalibrace	91
Uzavření etapy	92
Analýza	92
8.3 Rozdělení hodnocení	93
8.4 Spolehlivost posouzení	95
Kvalita	96
Dopad	97
Vhodnost	97
Rozdíly mezi částmi etapy kolem rekalibrace	98
Shoda napříč etapami a opakované posouzení dopadu	99
8.5 Typ řešení, hodnocený dopad a vhodnost	100
8.6 Predikce dopadu řešení jazykovým modelem	102
8.7 Shrnutí	106

Integrace	107
9. Diskuse	109
9.1 Oblouk čtyř studií	109
9.2 Repertoár nejmenšího odporu	112
9.3 Co lze jazykovému modelu svěřit	113
9.4 Spolehlivost lidského posouzení v průběhu etap	115
9.5 Omezení	117
9.6 Důsledky	118
Pro přípravu učitelů a školní praxi	118
Pro platformu Edustories	118
Pro jazykové modely v pedagogickém výzkumu	119
9.7 Další výzkum	119
9.8 Shrnutí	120
10. Závěr	121
10.1 Odpovědi na výzkumné otázky	121
10.2 Přínos knihy	123
10.3 Směrování	123
Literatura	125
Přílohy	135
Příloha A. Kódovací manuály	135
A.1 První etapa: kódovací kniha obsahového kódování	135
A.2 Druhá etapa: manuál hodnocení kvality řešení	144
A.3 Třetí etapa: operacionalizace kvality kazuistik	147
Příloha B. Prompty jazykových modelů	151
B.1 Extrakce a strukturace kazuistik	152
B.2 Anonymizace v promptu	154
B.3 Klasifikace řešení jazykovým modelem (model v roli anotátora)	154
B.4 Generace srovnávacího řešení pro druhou etapu	157
B.5 Rámcovaná řešení platformy Edustories	159
B.6 Predikce dopadu řešení jazykovým modelem	159
Příloha C. Doplnkové tabulky	161
C.1 Registr anotačních událostí	161
C.2 Přehled velikostí datových řezů	162
C.3 Přehled měř shody	163
Příloha D. Etické náležitosti	167
D.1 Informované prohlášení přispěvatele	167

D.2 Anonymizační protokol	170
Schválení etickou komisí	170
Výběrové ověření anonymizace	173

Prohlášení

Prohlašuji, že jsem habilitační práci vypracoval samostatně s využitím pouze citované literatury, informací a zdrojů a že tato práce nebyla předložena k udělení žádného předchozího titulu.

V Brně dne

.....

vlastnoruční podpis autora

Prohlášení o využití umělé inteligence

Při přípravě této práce autor využíval nástroje umělé inteligence (Claude, Anthropic) k organizaci pracovních podkladů, ke kontrole pravopisu, typografie a formátování textu a příloh, k práci s verzovaným repozitářem (Git/GitHub), k propojování a kontrole konzistence částí rukopisu s analytickými výstupy. Ve výzkumné části jsou jazykové modely zároveň předmětem a nástrojem výzkumu: klasifikovaly typy řešení (Studie 2), generovaly srovnávací řešení (Studie 3), předpovídaly dopad řešení (Studie 4) a podílely se na anonymizaci textů korpusu; postupy dokumentují kapitoly 6 až 8 a příloha B. Po použití těchto nástrojů autor veškerý obsah zkontroloval a upravil a nese za něj plnou odpovědnost.

Abstrakt

Zvládání náročného chování žáků patří k obtížným stránkám učitelské profese a příprava učitelů v něm naráží na mezeru mezi obecnými principy a jednáním v konkrétní situaci. Monografie se tématu věnuje ve dvou rovinách. Pedagogicky: korpus projektu Edustories čítá přes tři tisíce autentických kazuistik psaných převážně studujícími učitelství na pedagogických praxích; na lidsky kódovaném řezu korpusu kniha popisuje, jaké typy náročného chování, řešení a dopadů pisatelé zaznamenávají. V řešeních převažuje rozhovor a další postupy reagující na incident. Metodologicky: kniha kriticky prověřuje velké jazykové modely jako výzkumný nástroj pro analýzu takového korpusu. Na srovnávacím vzorku se jazykový model shodl s lidským kódem typu řešení na úrovni Cohena $\kappa = 0,64$; protože vzorek nebyl oddělen od vývoje instrukce pro model, nejde o nezávislé ověření. Po dodatečném vyloučení kazuistik, jejichž řešení se objevila mezi příklady v instrukci, činí $\kappa = 0,61$. V zaslepeném experimentu na 324 dvojicích řešení z kazuistik agresivního chování hodnotily anotátorky řešení, která podle jediné instrukce vygeneroval model gpt-4o, jako vhodnější a proaktivnější než autentická řešení zaznamenaná v kazuistikách; ta byla reaktivnější a behaviorálnější. Hodnocen byl text řešení, ne jednání ve třídě. Čtvrtá studie na dvojité posouzeném překryvu 693 kazuistik ukazuje po vyloučení případů projednaných při rekalibraci vysokou vzájemnou shodu v posouzení použitelnosti kazuistiky ($\kappa = 0,77$) při velmi nevyrovnaném rozdělení kategorií, dobrou shodu v posouzení dopadu řešení ($\kappa = 0,69$) a jen střední shodu v posouzení vhodnosti. Opakované posouzení dopadu týchž textů jednou pozdější anotátorkou se s dřívějšími kódy shoduje středně ($\kappa = 0,59$). Vztahová a kolektivní řešení se v lidsky kódovaných kazuistikách pojí s častěji popsaným dlouhodobým úspěchem než řešení represivní. Kategorie dopadu z první anotační etapy však částečně zahrnují normativní

Abstrakt

úsudek o přijatelnosti řešení a nejsou nezávislým měřením účinnosti. Kódovaný dopad řešení dokáže jazykový model z textu kazuistiky bez popisu výsledku předvídat jen omezeně: vyvážená přesnost 60 % je nad úrovní hádání nejčastější kategorie, ale pro doporučování postupů ve třídě nestačí. Přínosem knihy je empirický popis repertoárů řešení náročného chování v kazuistikách, postup pro nasazení jazykových modelů v pedagogickém výzkumu i s vymezením jeho hranic a doprovodný repozitář, z něhož lze přepočítat hlavní statistické výsledky.

Klíčová slova: náročné chování žáků, kazuistiky, řízení třídy, velké jazykové modely, model v roli anotátora, spolehlivost anotací, reprodukovatelnost, učitelské vzdělávání.

Abstract

Managing challenging student behaviour is among the difficult aspects of the teaching profession, and teacher education faces a gap between general principles and action in concrete situations. The monograph addresses the topic on two levels. Pedagogically, the Edustories corpus comprises more than three thousand authentic case studies written mostly by pre-service teachers during their school placements; on a human-coded snapshot of the corpus, the book describes the types of challenging behaviour, responses and outcomes that the writers record. Conversation and other incident-driven responses prevail. Methodologically, the book critically examines large language models as research instruments for analysing such a corpus. On a comparison sample, a language model agreed with the human code of response type at Cohen's $\kappa = 0.64$; because the sample was not kept separate from the development of the model's prompt, this is not an independent validation. After retrospectively excluding cases whose solutions appeared among the examples in the prompt, κ is 0.61. In a blinded experiment on 324 solution pairs from case studies of aggressive behaviour, annotators rated solutions generated by gpt-4o from a single prompt as more appropriate and more proactive than the authentic solutions recorded in the case studies, which were more reactive and behaviourally oriented. The ratings concerned written solutions, not classroom action. On a doubly rated overlap of 693 case studies, after excluding cases discussed during recalibration, the fourth study finds high inter-rater agreement on case usability ($\kappa = 0.77$) with a highly skewed category distribution, good agreement on the outcome of the solution ($\kappa = 0.69$) and only moderate agreement on its appropriateness. When one later annotator re-rated the outcome in the same texts, her codes agreed moderately with the earlier ones ($\kappa = 0.59$). In the human-coded case studies, relational and class-level solutions are associated with more

Abstract

frequently reported long-term success than punitive ones. The outcome categories of the first annotation campaign, however, partly incorporate normative judgments about the acceptability of a solution and do not independently measure effectiveness. A language model predicts the coded outcome of a solution from the case text without the outcome description only to a limited extent: a balanced accuracy of 60% is above guessing the most frequent category but insufficient for recommending classroom strategies. The book contributes an empirical description of response repertoires for challenging behaviour in the case studies, a procedure for deploying language models in educational research together with a delineation of its limits, and a companion repository from which the main statistical results can be recomputed.

Keywords: challenging student behaviour, case studies, classroom management, large language models, LLM-as-annotator, annotation reliability, reproducibility, teacher education.

Poděkování

Kniha vznikla v rámci projektu TA ČR TQ01000030. Děkuji anotátorkám výzkumného týmu Edustories, jejichž pečlivá práce tvoří páteř všech tří anotačních etap (v knize vystupují pod pseudonymy, protože anonymizaci uplatňujeme důsledně i tam, kde by poděkování bylo zaslouženě jmenovité), Michalu Štefánikovi za dlouholetou spolupráci na výpočetní straně projektu a všem studujícím učitelství, kteří své zkušenosti proměnili v kazuistiky.

Poznámka k publikacím

Tato monografie je původní vědeckou prací napsanou pro toto habilitační řízení. Není totožná s žádnou disertační ani jinou kvalifikační prací, kterou autor dříve předložil k získání akademického titulu. Navazuje na publikované výstupy projektu Edustories, zejména na příspěvek ve sborníku konference ČAPV (Nehyba, Karasová, Škarková, et al., 2024) a na datovou sadu Edustories-en (Nehyba, Karasová, & Štefánik, 2024). Data a analýzy Studií 1 až 3 se překrývají s časopiseckými rukopisy autorského týmu, které v době odevzdání nebyly publikovány; Studie 3 přepočítává analýzu jednoho z nich na týchž datech. Analýzy Studií 1 až 3 byly pro knihu provedeny znovu z verzovaných dat. Studie 4 nebyla publikována jinde.

1. Úvod

1.1 Problém

Zvládání náročného chování žáků patří k obtížným stránkám učitelské profese. Pro začínající učitele bylo v klasickém přehledu nejčastěji uváděným problémem a spoluutvářelo známý „šok z reality“ při vstupu do praxe (Veenman, 1984); současná data mezinárodního šetření TALIS shrnuje oddíl 3.1. Novější česká studie popisuje, jaké situace začínající učitelé řeší, jak na ně reagují a s jakým dopadem na chování žáků (Karasová & Nehyba, 2025). Kniha vychází z předpokladu, rozvinutého v kapitole 2, že příprava učitelů zprostředkuje obecné principy snáze než to, co zkušení praktici získávají léty: konkrétní, situačně ukotvený repertoár řešení. Obecná znalost přístupů k náročnému chování sama nestačí ve chvíli, kdy se dvě děti ve třetí lavici začnou prát. Tato mezera mezi obecným poznáním a jednáním v konkrétní situaci je jádrem problému, jemuž se kniha věnuje.

Problém není specificky český, má však v českém prostředí konkrétní podobu. Náročné chování se v domácím odborném jazyce ustálilo jako zastřešující pojem teprve nedávno, především zásluhou metodického doporučení školní inspekce (Felcmanová et al., 2021). Domácí výzkum se dosud opírá hlavně o pozorování praxí studentů učitelství, o analýzy výukové komunikace a o malé kvalitativní vzorky začínajících učitelů (podrobně kapitola 2). Popisných údajů velkého rozsahu o tom, jaké situace čeští učitelé a studenti učitelství za náročné považují, jak je řeší a jak jejich řešení dopadají, je proto málo. Na takový popis se kniha zaměřuje.

1. Úvod

Osvědčeným prostředkem, jak mezeru mezi principem a situací překlenout, jsou kazuistiky, tedy autentické případové popisy z praxe. Aby mohly sloužit výzkumu i přípravě učitelů v širším měřítku, je jich potřeba mnoho a je třeba je systematicky sbírat a umět analyzovat. Tím se otevírá druhá rovina problému. Rozsáhlý soubor nestrukturovaných pedagogických textů přesahuje kapacitu tradiční kvalitativní analýzy: ruční kódování tisíců kazuistik vyžaduje měsíce práce vyškoleného týmu a ani shoda lidských anotátorů není úplná. Velké jazykové modely nabízejí možnost tuto kapacitu rozšířit. Do českých škol mezitím rychle pronikly: s nástroji umělé inteligence (AI) experimentovala už krátce po nástupu ChatGPT značná část učitelů (Kopecký et al., 2023) a používá je nadpoloviční většina žáků (Kopecký & Voráč, 2024). Role modelů ve výzkumu však není samozřejmá. Je třeba doložit, nakolik se shodují s lidským úsudkem, a vymezit, k čemu je lze použít a k čemu už ne. Kniha proto s jazykovým modelem zachází jako s nástrojem, který je třeba nejprve kriticky prověřit, a teprve potom mu svěřit výzkumnou práci.

1.2 Projekt Edustories a data knihy

Empirickým základem knihy je projekt Edustories, který sbírá a zpřístupňuje kazuistiky náročního chování psané studujícími učitelství a učiteli. Z projektu pochází rostoucí korpus, jehož zveřejněný řez čítá přes tři tisíce kazuistik, a tři anotační etapy, které nad korpusem proběhly. První přinesla lidské kódování obsahu, druhá zaslepené hodnocení učitelských a generovaných řešení a třetí posouzení kvality kazuistik, dopadu a vhodnosti řešení (kapitola 8). V projektu je třeba rozlišovat tři roviny: veřejnou platformu, jež slouží praxi a přípravě učitelů; databázi kazuistik, jejíž část zůstává neveřejná a z níž pocházejí datované řezy analyzované v knize; a kurátorovaný anotovaný výběr v anglickém překladu (datová sada Edustories-en), který slouží výzkumné komunitě. Projekt, jeho institucionální zázemí i vznik dat podrobně představuje kapitola 3. Kniha stojí výhradně na těchto datech; jak jsou analýzy dokumentovány a jak si je lze ověřit, popisují kapitola 4 a oddíl 1.5.

1.3 Cíl a výzkumné otázky

Cílem knihy je na datech projektu Edustories propojit obě roviny problému: popsat, co kazuistiky vypovídají o zvládání náročného chování, a zároveň prověřit velké jazykové modely jako nástroj jejich výzkumu. Tomuto dvojímu cíli odpovídají čtyři výzkumné otázky, z nichž každá je rozpracována v jedné empirické studii:

- **VO1:** Jak lze vybudovat a popsat rozsáhlý korpus autentických kazuistik a jaké typy náročného chování, řešení a dopadů zachycuje? (Studie 1, kap. 5)
- **VO2:** Nakolik se velký jazykový model shodne s vyškolenými lidskými anotátory při deduktivní klasifikaci obsahu kazuistik? (Studie 2, kap. 6)
- **VO3:** Jak se liší řešení generovaná jazykovým modelem od autentických řešení učitelů v jejich vhodnosti a orientaci? (Studie 3, kap. 7)
- **VO4:** Jakou kvalitu mají kazuistiky, jak spolehlivě ji lze posoudit a nakolik dokáže jazykový model předvídat úspěšnost řešení? (Studie 4, kap. 8)

Otázky jsou seřazeny záměrně. Studie 1 popisuje materiál a jeho pedagogický obsah. Studie 2 srovnává model v roli anotátora s lidským kódováním téhož materiálu; shodu ověřuje u typu řešení. Ve Studii 3 se role obrací: lidé zaslepeně hodnotí řešení vygenerovaná modelem spolu s autentickými řešeními z kazuistik, které platforma v době výběru vedla v kategoriích fyzické a verbální agrese. Studie 4 zjišťuje, jak spolehlivé je lidské posouzení kazuistik a nakolik model předvídá kódovaný dopad řešení. Pedagogická a metodologická linie se tak střídají a obě vycházejí z jednoho korpusu.

1.4 Struktura knihy

Knihy postupuje od teorie přes metodologii ke čtyřem empirickým studiím a jejich syntéze. Kapitola 2 klade teoretické základy obou linií: náročného chování, přístupů k jeho řešení a repertoárů učitelů na straně jedné a využití jazykových modelů v pedagogickém výzkumu na straně druhé. Kapitola 3 představuje projekt Edustories a vznik dat. Kapitola 4 popisuje

1. Úvod

metodologii celé knihy: časovou osu korpusu, tři anotační etapy, zpracování textů, kódovací schémata, etiku a zásady reprodukovatelnosti. Kapitoly 5 až 8 přinášejí čtyři empirické studie odpovídající výzkumným otázkám. Kapitola 9 obě linie propojuje v diskusi a pojmenovává limity; kapitola 10 shrnuje přínos a naznačuje další směřování. Přílohy obsahují kódovací manuály, instrukce pro jazykové modely (prompty), doplňkové tabulky a etické náležitosti.

1.5 Dřívější výstupy a dostupnost dat

Z projektu Edustories byly dosud publikovány zejména stručná zpráva pro českou pedagogickou výzkumnou obec ve sborníku konference ČAPV (Nehyba, Karasová, Škarková, et al., 2024) a kurátorovaný anotovaný výběr korpusu zveřejněný jako datová sada Edustories-en (Nehyba, Karasová, & Štefánek, 2024). Metodologicky knize předchází také starší práce autorského týmu s hlubokými jazykovými modely nad reflektivními texty (Nehyba & Štefánek, 2023), která stojí na jiných datech. Všechny analýzy byly pro knihu provedeny nově z verzovaných dat. Popis korpusu a Studie 4 pracují s nejnovějším řezem. Obsahové kódování ve Studiích 1 a 2 i Studie 3 stojí na datech starších řezů a anotačních etap, z nichž vycházejí i připravované časopisecké rukopisy (viz poznámka k publikacím na začátku knihy).

Skripty, analytické notebooky, hotové reporty a odvozená data bez textů kazuistik jsou dostupné v [doprovodném repozitáři](#) (Nehyba, 2026) spolu s návodem k přepočtu; postup popisuje oddíl 4.8. Hlavní statistické výsledky empirických studií lze z odvozených dat, v nichž anotátorky vystupují pod pseudonymy, přepočítat bez textů kazuistik a bez opětovného spuštění jazykového modelu. Původní anotaci ani generování odpovědí modelu však bez neveřejných textů zopakovat nelze. Úplná data Studie 3 včetně textů jsou v balíčku na platformě OSF (odkaz bude zveřejněn s publikací souvisejícího článku, do té doby na vyžádání). Zveřejněné kazuistiky jsou dostupné na platformě projektu; některé kazuistiky analyzované ve starších řezech mezi nimi nejsou.

Východiska

2. Teoretická východiska

Kniha stojí na průsečíku dvou linií. První je pedagogická: náročné chování žáků, přístupy k jeho řešení a postupy, jimiž na ně učitelé reagují. Druhá je metodologická: velké jazykové modely jako nástroj pedagogického výzkumu. U obou linií kapitola vymezuje pojmy, s nimiž kniha dále pracuje, shrnuje stav poznání včetně českého kontextu a formuluje otázky, na které odpovídají empirické kapitoly. Postupuje od vymezení náročného chování (2.1) přes čtyři přístupy k jeho řešení (2.2) a pojem učitelského repertoáru (2.3) k jazykovým modelům ve výzkumu (2.4); závěrečný oddíl (2.5) obě linie spojuje.

2.1 Náročné chování žáků: vymezení a český kontext

Vymezení pojmu

Kniha pracuje s pojmem náročné chování (challenging behaviour) jako se zastřešujícím označením pro chování žáků, které se výrazně odlišuje od chování očekávaného v daném věku a které negativně ovlivňuje učení, vztahy nebo bezpečí ve škole. Opírá se přitom o české metodické doporučení, které pojem do domácího prostředí systematicky zavedlo (Felcmanová et al., 2021). Pojem je záměrně širší než tradiční „nekázeň“: zahrnuje projevy od nepozornosti přes verbální a fyzickou agresi až po záškoláctví a šikanu, aniž by předjímal jejich příčinu. Vymezení přitom výslovně počítá s tím, že záleží na úhlu pohledu: co je náročné, určuje pohled pedagoga, který s chováním pracuje, a stejné jednání může být v různých kontextech chápáno různě (Felcmanová et al., 2021). Pro knihu je to podstatné. Jejím materiálem jsou kazuistiky, které psali převážně studující učitelství o situacích z praxe, jež

2. Teoretická východiska

sami zažili nebo jež řešil jejich provázející učitel. Korpus tedy nezachycuje chování žáků „o sobě“, ale tak, jak je konkrétní aktéři viděli, rámovali a řešili.

Kázeň v české pedagogické tradici

V české pedagogice se o tomtéž jevu dlouho mluvilo jazykem kázně. Kázeň patřila ke klasickým, normativně pojatým tématům oboru a nekázeň se vymezovala jako porušování stanovených školních norem; v polistopadové literatuře toto pojetí soustavně rozpracovaly Bendlový monografie o školní kázni a kázeňských problémech (Bendl, 2001, 2011). Takové vymezení nabízí jasné kritérium, má však i omezení. Normativní ukotvení činí z chování především přestupek a odvádí pozornost od jeho funkce. Soustředění na porušení pravidel navíc z definice vylučuje projevy, které žádnou výslovnou normu neporušují, a přesto narušují výuku nebo ukazují na žákovu nouzi. Slovník oboru se proto v posledních dvou desetiletích posouval od nekázně k problémovému a rizikovému chování, s nímž pracuje školní prevence (Ministerstvo školství, mládeže a tělovýchovy [MŠMT], 2010). Inkluzivní novela školského zákona (zákon č. 82/2015 Sb.) spolu s navazující vyhláškou (vyhláška č. 27/2016 Sb.) uložila školám poskytovat podpůrná opatření pro žáky se speciálními vzdělávacími potřebami. Pojem náročné chování následně používá doporučení školní inspekce (Felcmanová et al., 2021).

Novější domácí literatura chápe náročné chování funkcionálně: jako reakci na stres, jako důsledek dosud nerozvinutých sociálních a emočních dovedností, případně jako prostředek naplňování potřeb, které žák neumí naplnit jinak (Kubíčková & Felcmanová, 2021). Tento výklad, opřený mimo jiné o výzkum nepříznivých zkušeností v dětství, má přímý důsledek pro hodnocení reakcí učitele: represivní zásah může incident krátkodobě ukončit, ale mívá jeho funkci. Normativní rámec kázně neodmítáme; jako interpretační pozadí však používáme funkcionální perspektivu, protože lépe odpovídá tomu, jak jsou kazuistiky v korpusu napsány, včetně částí věnovaných anamnéze a dlouhodobému výsledku řešení. Starší terminologie se v knize přesto objevuje: kódovací kniha první anotační etapy (příloha A) používá

ještě pojem problémové chování a některé její kategorie vycházejí z normativní logiky porušení pravidel.

Náročné chování jako profesní problém

Zvládání chování žáků patří k nejčastěji uváděným problémům začínajících učitelů (oddíl 3.1 k tomu uvádí současná česká data ze šetření TALIS 2024). Klasický přehled starších studií o vnímaných problémech začínajících učitelů řadil kázeň ve třídě na první místo, před motivování žáků a práci s individuálními rozdíly, a přechod z přípravy do praxe popsal jako „šok z reality“ (Veenman, 1984). Na tuto linii navazuje soudobý výzkum reakcí učitele zaměřených na žáka (Karasová & Nehyba, 2023) i studie z českého prostředí, která popisuje, jakým kázeňským situacím začínající učitelé čelí, jak na ně reagují a jaký má jejich reakce bezprostřední dopad na chování žáků (Karasová & Nehyba, 2025). Význam tématu nespočívá jen v subjektivní zátěži: řízení třídy ovlivňuje i výsledky učení žáků (viz následující pododdíl). Pro studující učitelství, z jejichž praxí pochází převážná část korpusu (kapitola 3), a pro začínající učitele je opora v této oblasti zvlášť potřebná.

Řízení třídy jako výzkumný rámec

Zastřešující oblastí, v níž se reakce na náročné chování zkoumají, je řízení třídy (classroom management), v části české literatury označované také jako vedení třídy. Přehledová studie, která pojem do české literatury uvedla, člení tuto oblast podle tří dichotomií: behavioristické versus humanistické přístupy, behaviorální versus instrukční management a proaktivní versus reaktivní strategie učitele (Lukas & Lojďová, 2018). Dvě z těchto dichotomií kniha ve Studii 3 používá jako osy popisu řešení (reaktivní–proaktivní a humanistická–behaviorální, oddíl 2.3). Metaanalýza studií z primárního vzdělávání ukazuje malé, ale statisticky významné účinky: strategie a programy řízení třídy zlepšují v průměru akademické, behaviorální i sociálně-emoční výsledky žáků, přičemž nejúčinnější byly programy zaměřené na sociálně-emoční rozvoj (Korpershoek et al., 2016). Systematický přehled zároveň vymezuje okruh

2. Teoretická východiska

konkrétních praktik s dostatečnou empirickou oporou, od jasné struktury a nacvičených rutin přes aktivní výuku po posilování vhodného chování a odstupňované reakce na chování nevhodné (Simonsen et al., 2008). Empirická opora pro řadu postupů tedy existuje. Ve Studii 1 se kniha vrací k otázce, nakolik se tyto poznatky projevují v reakcích, které kazuistiky popisují.

V domácím pedagogickém výzkumu lze k tématu rozlišit tři proudy. První zkoumá řízení třídy přímo: rozsáhlá empirická monografie brněnského týmu popisuje na základě pozorování, rozhovorů a dotazníků reaktivní i proaktivní strategie studentů učitelství na dlouhodobé praxi a jejich provázejících učitelů (Vlčková et al., 2020). Druhý proud se věnuje výukové komunikaci: autoři na základě dlouhodobých pozorování a analýz videozáznamů popsali, jak učitelské otázky, zpětná vazba a mocenské vztahy spoluutvářejí vztah mezi učitelem a žáky, a formulovali ideál dialogického vyučování (Šedřová et al., 2012). Třetí proud sleduje začínající učitele: kvalitativní studie popisuje, jak se jejich vnímání nekázně posouvá od jevu „na straně žáka“ k projevu žákovy diskomfortu (Ševčíková et al., 2023). K tématu se vyjádřila i školská správa: metodické doporučení České školní inspekce shrnulo přístupy k náročnému chování pro školy a školská zařízení (Felcmanová et al., 2021). Pokud je nám známo, systematických popisů toho, jak učitelé a studující učitelství náročné chování řeší a jak o tom píší, opřených o velký soubor kazuistik místo dotazníků nebo malých kvalitativních vzorků, je v české literatuře zatím málo. Takový popis umožňuje korpus Edustories.

Kazuistika jako pramen

Volbu kazuistiky jako datového pramene je třeba zdůvodnit, protože má specifické silné i slabé stránky. Silnou stránkou je situovanost: pisatel popisuje konkrétní situaci, kterou zažil sám nebo kterou řešil jeho provázející učitel, včetně kontextu, průběhu řešení a jeho výsledku. Text tak zachycuje jednání v celém situačním oblouku, a ne jen izolovanou odpověď na položku dotazníku. Strukturovaná šablona sběru (popis situace, anamnéza, řešení, výsledek; kapitola 3) tento oblouk dále zvýrazňuje. Slabou stránkou je dvojí zprostředkování: kazuistika je zpětná rekonstrukce, v níž pisatel jednání už jednou interpretoval, a je psána z

jedné perspektivy, takže podléhá výběrové paměti i sebe prezentaci. Korpus proto chápeme jako záznam toho, jak učitelé a studující učitelství náročné situace vnímají a rámuji a jaká řešení považují za hodná zaznamenání, ne jako záznam toho, co se ve třídách „objektivně“ děje. Pro studium repertoárů je právě tato vrstva podstatná, protože repertoár se reprodukuje prostřednictvím toho, co si aktéři ze situací odnášejí a co si o nich vyprávějí.

2.2 Přístupy k řešení náročného chování

V mezinárodní literatuře se ustálilo několik pojmenovaných přístupů k náročnému chování, které nabízejí ucelený rámec: teorii toho, proč náročné chování vzniká, principy práce s ním a zpravidla i strukturovaný program zavedení. Kniha pracuje se čtyřmi z nich: s pozitivní behaviorální intervencí a podporou (PBIS), se sociálně-emočním učením (SEL), s nenásilnou komunikací (NVC) a s restorativní praxí. Důvod výběru je pragmatický: právě v těchto čtyřech rámcích generuje platforma Edustories ke kazuistikám alternativní řešení (spolu s perspektivou proaktivní a reaktivní, která odpovídá první ose popisu řešení z oddílu 2.3). Ve Studii 3 tytéž rámce slouží jako štítky, jimiž anotátorky posuzovaly, zda hodnocené řešení některý z nich uplatňuje. Do platformy byly zařazeny jako mezinárodně známé, vzájemně odlišné přístupy, které lze formulovat jako instrukce pro generování řešení. Čtveřice není vyčerpávající: stojí mimo ni například celá tradice individuálních funkcionálních analýz chování nebo trauma-informovaný přístup, který počítá s možnými dopady traumatických zkušeností žáka a který domácí literatura doporučuje kombinovat s celoškolní pozitivní podporou (Kubíčková & Felcmanová, 2021).

Pro interpretaci výsledků je podstatné, že žádný ze čtyř přístupů není, pokud je nám známo, v českých školách zavedeným systémem; český kontext každého z nich shrnují následující pododdíly. Čtveřice tedy v knize vystupuje jako vnější normativní horizont.

Čtyři přístupy se liší tím, na jaké úrovni zasahují, i tím, jak silně jsou doloženy. PBIS zasahuje na úrovni školy jako systému, SEL na úrovni kurikula a kompetencí žáka, NVC na úrovni komunikace jednotlivého učitele a restorativní praxe na úrovni vztahů a společenství

2. Teoretická východiska

třídy. PBIS a SEL mají oporu v metaanalýzách; nenásilná komunikace a restorativní praxe vyšly z praktické zkušenosti a normativní teorie a experimentálně se ověřují až dodatečně. Z obou rozdílů vychází otázka, k níž se empirické kapitoly vrací: zda se učitelské repertoáry přiklání k přístupům s nejsilnější empirickou oporou, nebo spíše k postupům, které jsou jednotlivci v situaci nejnáze dostupné. Studie 1 k ní přispívá nepřímo, skladbou kategorií řešení v korpusu. Ve Studii 3 anotátorky rozpoznávaly čtyři přístupy v hodnocených řešeních kazuistik z kategorií fyzické a verbální agrese: v učitelských řešeních méně často než v generovaných, ale s nízkou vzájemnou shodou, takže výsledek vypovídá spíše o celkové tendenci (oddíly 7.2, 7.3 a 7.4). Tento rozdíl může zčásti odrážet slabou přítomnost rámců v českých školách, a ne jen volby jednotlivých učitelů. Následující pododdíly u každého přístupu shrnují původ a principy, stav evidence, hlavní kritiky a český kontext; oddíl uzavírá srovnávací tabulka.

Pozitivní behaviorální intervence a podpora (PBIS)

Pozitivní behaviorální intervence a podpora (Positive Behavioral Interventions and Supports, PBIS; v projektu zkráceně „pozitivní behaviorální podpora“) vychází z aplikované behaviorální analýzy a je koncipována jako celoškolský systém, ne jako technika jednotlivého učitele. Její zakladatelé ji popisují jako příklad uplatnění behaviorální vědy ve společensky významném rozsahu. Zapojené školy výslovně definují očekávané chování, systematicky je nacvičují a oceňují, pracují s odstupňovaným kontinuem důsledků a svá rozhodnutí opírají o průběžně sbíraná data o chování. Ve Spojených státech byl tento rámec zaveden ve více než 21 000 školách (Horner & Sugai, 2015). Jádrem rámce je vícevrstvý model podpory: univerzální prevence pro všechny žáky, cílená podpora pro žáky se zvýšeným rizikem a intenzivní individuální intervence pro žáky s nejzávažnějšími obtížemi.

PBIS patří k empiricky dobře prověřeným celoškolským rámcům. Aktualizovaná metaanalýza skupinových studií, zahrnující randomizované i kvaziexperimentální studie z USA a Evropy, doložila významné snížení kázeňských postihů a dočasných vyloučení z vyučování a zlepšení akademických výsledků i vnímaného bezpečí a organizačního zdraví školy (Lee &

Gage, 2020). Síla rámce spočívá v propracovaném zavádění: PBIS je definován měřitelnými prvky, což umožňuje zkoumat nejen, zda funguje, ale i za jakých podmínek.

Behaviorálnímu jádru PBIS lze vytknout, že systémy výslovného oceňování mohou posilovat vnější motivaci na úkor vnitřní a zužovat morální rozvoj na nácvik poslušnosti. Z funkcionálního pohledu navíc důraz na viditelné chování může odvádět pozornost od jeho příčin. Praktickým limitem je náročnost zavedení: účinky stojí a padají s věrností, s jakou škola rámec uplatňuje, což vyžaduje podporu vedení školy, čas a průběžný sběr dat. V Česku PBIS nemá institucionální zázemí srovnatelné s americkým. Domácí literatura jej doporučuje jako slibný rámec celoškolské podpory duševního zdraví v kombinaci s trauma-informovaným přístupem (Kubíčková & Felcmanová, 2021) a jeho prvky se objevují v metodickém doporučení inspekce (Felcmanová et al., 2021). Školám jej v posledních letech zprostředkovává organizace Society for All, která vydala českou verzi metodiky třídních strategií PBIS (Felcmanová et al., 2023).

Sociálně-emoční učení (SEL)

Sociálně-emoční učení (Social and Emotional Learning, SEL) vychází z předpokladu, že sociální a emoční dovednosti, jako jsou sebeuvědomění, seberegulace, sociální vnímavost, vztahové dovednosti a odpovědné rozhodování, lze systematicky vyučovat a že jejich rozvoj předchází náročnému chování lépe než následné kázeňské zásahy. Na rozdíl od PBIS, který upravuje prostředí a důsledky, se SEL zaměřuje na kompetence žáka; oba rámce se proto v praxi často kombinují.

Evidence SEL je rozsáhlá. Často citovaná metaanalýza univerzálních školních programů doložila významné zlepšení dovedností, postojů, chování i akademického výkonu, který se zlepšil přibližně o 11 percentilových bodů (Durlak et al., 2011). Následný přehled dvanácti metaanalýz účinky potvrdil a zároveň upozornil, že poznatky o moderátorech, tedy o tom, pro koho a za jakých podmínek programy fungují nejlépe, zůstávají nejednotné (Durlak et al., 2022). Novější metaanalýza experimentálních studií z desítek zemí ukazuje, že ve skupinách s intervencí SEL měli žáci oproti kontrolním skupinám lepší dovednosti, chování i

2. Teoretická východiska

akademické výsledky a lepší bylo i školní klima (Cipriano et al., 2023). Souhrnně jde o rámec s opakovaně potvrzenou metaanalytickou oporou.

Slabinou výzkumu SEL je, že účinky se často měří krátce po intervenci a pomocí sebevýpovědí; přenos do reálného chování v konfliktní situaci je proto doložen slaběji. Z kulturního hlediska lze namítnout, že katalogy „správných“ emočních dovedností odrážejí nevyslovené normy středostavovské kultury a mohou jiné styly projevu označovat za patologické. Stejně jako u PBIS klesají účinky při nedůsledném zavedení, a právě věrnost zavedení se v běžném provozu škol udržuje nejhůře.

V Česku SEL nemá podobu programu, má však blízkého příbuzného. Osobnostní a sociální výchova (OSV) byla od zavedení rámcových vzdělávacích programů průřezovým tématem základního vzdělávání (MŠMT, 2023) a revize programu z roku 2025 z ní činí samostatný vzdělávací obor (MŠMT, 2025). Sleduje obdobné cíle jako SEL: sebepoznání, seberegulaci, vztahy a komunikaci ve skupině. Její česká didaktika se přitom od počátku hlásí k britské tradici personal and social education, z níž převzala i název a důraz na zkušenostní, prožitkové učení (Valenta, 2006). Na rozdíl od SEL však OSV nemá standardizovaný program ani metaanalytickou evidenci účinků a její podoba závisí na jednotlivé škole a učiteli. Pro knihu má SEL dvojitý význam: je jedním ze čtyř rámců platformy a pomáhá interpretovat korpus, v němž řada popsanych řešení nevysloveně pracuje právě s regulací emocí a se vztahy. Štítek SEL ve Studii 3 přitom mezi převzatým zahraničním programem a domácí OSV nerozlišuje.

Nenásilná komunikace (NVC)

Nenásilná komunikace (Nonviolent Communication, NVC) je komunikační model, který rozvinul Marshall Rosenberg: mluvčí odlišuje pozorování od hodnocení, pojmenovává pocity, spojuje je s potřebami a formuluje konkrétní prosby a ve stejné struktuře přijímá sdělení druhých (Rosenberg, 2015). V pedagogickém kontextu NVC nabízí jazyk pro deeskalaci konfliktu a pro vztahovou práci se žákem. Na rozdíl od PBIS a SEL nejde o program s kuri-

kulem, ale o praxi jednotlivce, kterou lze uplatnit okamžitě a bez systémové podpory. Tato snadná dostupnost může vysvětlovat oblibu NVC v kurzech dalšího vzdělávání.

Evidence NVC je výrazně slabší než u předchozích dvou přístupů. Mapující přehled (scoping review) ze zdravotnické práce shrnul jen sedm empirických studií nevelkého rozsahu s různorodou metodikou. Jeho autoři hodnotí NVC jako přínosný nástroj (Adriani et al., 2024); rozsah a kvalita zahrnutých studií však podle našeho soudu takový závěr podpirají jen slabě. Jedna z mála kontrolovaných studií, randomizovaný experiment s mediky, doložila po krátkém tréninku NVC mírný nárůst sebesposuzované empatie, ne však v implicitních kognitivních testech (Epinat-Duclos et al., 2021). Recenzovanou metaanalýzu účinků NVC ve školním prostředí se nám nepodařilo najít. S NVC proto zacházíme opatrně: jako s vlivným a v praxi rozšířeným jazykem vztahové práce, jehož účinnost je doložena spíše kazuisticky než experimentálně. Tento rozdíl v síle evidence tvoří pozadí Studie 3, v níž anotátorky rozpoznávaly znaky NVC a dalších přístupů; instrukce pro generování řešení přitom žádný rámec nepředepisovala.

Vedle slabé evidence hrozí u NVC mechanické užití: čtyřkroková struktura se snadno změní ve formuli, kterou mluvčí odříkává bez postoje, z něhož model vychází. Takto užitá formule může působit manipulativně nebo odtažitě. Další námitka se týká asymetrie školní situace. Model předpokládá dialog rovnocenných, zatímco učitel vůči žákovi vystupuje i jako nositel moci a garant pravidel, takže „čisté“ nenásilné jednání naráží na institucionální roli. V Česku se NVC šíří především překlady Rosenbergových knih (Rosenberg, 2022), kurzy dalšího vzdělávání a popularizační literaturou, tedy cestami, které recenzovanou empirickou oporu nevytvářejí. O to důležitější je otázka, zda a v jaké podobě se vztahový, na potřeby zaměřený jazyk objevuje v řešeních popsanych v kazuistikách. Studie 1 k ní poskytuje jen nepřímé údaje, například četnost rozhovoru a nerespektující komunikace mezi kategoriemi řešení (oddíly 5.3 až 5.5).

Restorativní praxe

Restorativní praxe přenáší do škol principy restorativní justice. Místo otázky, jaké pravidlo bylo porušeno a jaký trest následuje, klade otázky, komu vznikla újma, jaké potřeby z ní vyplývají a kdo má povinnost újmu napravit. Typickými nástroji jsou restorativní rozhovory a kruhy, v nichž ten, kdo újmu způsobil, poškození i třída jako společenství hledají nápravu. Přístup je výslovně koncipován jako alternativa k vylučujícím trestům (dočasné či trvalé vyloučení ze školy); jejich nerovnoměrné dopady na znevýhodněné skupiny žáků patří k hlavním důvodům jeho zavádění.

Evidence restorativní praxe je smíšená a metodologicky poučná. Narativní přehled výzkumu dokumentuje rychlé šíření přístupu a převážně příznivé výsledky terénních studií (Fronius et al., 2019). Navazující přehled kvantitativních studií dospívá k závěru, že programy mohou zlepšit klima a snížit počet kázeňských postihů, zatímco výsledky pro šikanu, absence a akademický výkon jsou neprůkazné (Darling-Hammond et al., 2020). Nejprísněji navržená studie dopadla střizlivěji. Ve dvouletém klastrovém randomizovaném experimentu na školách nižšího sekundárního stupně se intervenční a kontrolní školy na úrovni randomizace významně nelišily; zkušenost s restorativními praktikami, kterou žáci sami uváděli, však souvisela s lepším vztahem ke škole, lepším klimatem i nižší viktimizací (Acosta et al., 2019). Rozdíl mezi nevýznamným účinkem programu a příznivou souvislostí uváděné zkušenosti s výsledky lze interpretovat jako náznak, že záleží na kvalitě zavedení; tato souvislost však nemá experimentální oporu. Restorativní praxe tedy patrně nefunguje jako „balíček“, který stačí formálně zavést.

Slabinami restorativní praxe jsou náročnost autentického provedení (kruh vedený bez důvěry a času se snadno stává rituálem), riziko sekundární viktimizace poškozených a napětí mezi restorativní logikou a formálními kázeňskými řády škol. V Česku se restorativní principy objevují v metodických materiálech a v nabídce neziskových organizací; systematický domácí výzkum jejich účinků jsme v recenzované literatuře nenašli. Kódovací kniha první etapy restorativní praxe jako samostatnou kategorii nerozlišuje; Studie 1 proto její výskyt v korpusu nekvantifikuje (příloha A, tabulka A.2).

Srovnání a vztah ke knize

Tabulka 2.1 čtveřici shrnuje.

Tabulka 2.1. Čtyři přístupy k řešení náročného chování, o něž se opírá platforma Edustories i Studie 3. Hodnocení síly evidence se opírá o přehledové a experimentální zdroje uvedené v posledním sloupci, podrobně rozebrané v oddílu 2.2.

Přístup	Primární zaměření			
	a mechanismus	Síla evidence	Hlavní limity	Klíčové zdroje
PBIS	prostředí a důsledky; celoškolský systém očekávání, nácviku, oceňování a dat	silná (metaanalýza randomizovaných i kvaziexperimentálních studií)	behaviorální jádro; náročná věrnost zavedení; závislost na vedení školy	Horner & Sugai, 2015; Lee & Gage, 2020
SEL	kompetence žáka; kurikulární rozvoj sociálních a emočních dovedností	silná (opakované velké metaanalýzy)	krátkodobá měření a sebevypovědi; kulturní normativita; věrnost zavedení	Durlak et al., 2011; Durlak et al., 2022; Cipriano et al., 2023
NVC	komunikace jednotlivce; pozorování–pocity–potřeby–prosby	slabá (několik malých studií; jeden randomizovaný experiment mimo školní prostředí)	chybí evidence ze školního prostředí; riziko mechanického užití „formule“	Rosenberg, 2015; Adriani et al., 2024; Epinat-Duclos et al., 2021
Restorativní praxe	vztahy a náprava újmy; rozhovory a kruhy místo vylučujících trestů	smíšená (randomizovaný experiment bez efektu programu; příznivá souvislost uváděné zkušenosti s výsledky)	stojí na autentickém provedení; napětí s formálními řády; časová náročnost	Fronius et al., 2019; Darling-Hammond et al., 2020; Acosta et al., 2019

2. Teoretická východiska

Pro knihu mají tyto čtyři rámce dvojí roli. Tvoří normativní horizont: popisují, jak by řešení náročného chování mohlo vypadat, kdyby se opíralo o poznatky oboru, a další kapitoly sledují, jak daleko od tohoto horizontu leží řešení popsaná v kazuistikách. Zároveň tvoří část nabídky platformy a ve Studii 3 slouží anotátorkám jako štítky. Přiřadit reálné řešení k přístupu je přitom věcí interpretace: učitelka, která vede restorativně laděný rozhovor, se na žádný rámec neodvolává, a tentýž postup může obsahovat prvky několika přístupů zároveň. Taková přiřazení proto chápeme jako úsudky poučených posuzovatelek, u nichž je třeba vykázat shodu; ve Studii 3 je tato shoda nízká (oddíly 7.3 a 7.4).

2.3 Repertoáry učitelů a učňovství pozorováním

Repertoár jako jednotka popisu

Způsob, jímž učitelé na náročné chování reagují, kniha uchopuje pojmem repertoár: soubor osvojených postupů, které má učitel v situaci k dispozici a z nichž pod časovým tlakem vybírá. Pojem přebírá ze dvou zdrojů. V sociologii kultury označuje zásobu návyků, dovedností a stylů, z níž aktéři skládají strategie jednání (Swidler, 1986); etnometodologická studie vyučování jím popsala udržování pořádku ve třídě jako soubor rutinních, často nepovšimnutých metodických dovedností učitele (Payne & Hustler, 1980). Kniha jej užívá v úzkém, popisném smyslu: pojem se netýká toho, co učitel zná nebo čemu věří, ale toho, jak v situaci jedná, a počítá s tím, že deklarovaná přesvědčení a jednání pod emočním tlakem se mohou rozcházet. Repertoár je zároveň jednotkou, kterou lze empiricky sledovat na velkém souboru situací: četnosti a kombinace kategorií řešení v korpusu, popsané ve Studii 1, jsou jeho přibližnou operacionalizací, protože zachycují jednání tak, jak je pisatelé kazuistik zaznamenali.

Učňovství pozorováním

Podstatná část repertoáru nevzniká ve formální přípravě. Lortie (1975) ve své klasické sociologické studii učitelství popsal, jak budoucí učitelé procházejí dlouhým „učňovstvím pozo-

rováním“ (apprenticeship of observation). Během školní docházky tráví tisíce hodin sledováním učitelů z pozice žáka a odnášejí si z ní intuitivní, biograficky ukotvené představy o tom, jak se učí a jak se zjednává pořádek. Tato tichá průprava je neúplná, protože žák vidí jen vnější stránku vyučování, a ne rozhodování, přípravu a záměry učitele. Je také konzervativní, protože reprodukuje postupy, které byly ve třídách vidět před lety, bez ohledu na jejich doloženou účinnost. Pregraduální příprava ji podle Lortieho spíše doplňuje, než aby ji přepisovala.

Učňovství pozorováním nabízí vysvětlení, proč může být podstatná část repertoáru tacitní: osvojila se nápodobou dřív, než ji bylo možné pojmově uchopit, a pod tlakem situace se může uplatnit rychleji než rozvaha. Z toho vyvozujeme dva předpoklady, z nichž vychází uspořádání knihy. Prvním je, že repertoár se patrně špatně měří přímou otázkou: lze očekávat, že učitelé v dotaznících vypovídají spíše o svých přesvědčeních a o tom, co považují za správné, než o tom, co v afektem nabitě situaci skutečně udělali. Popisy konkrétních situací, jaké shromažďuje korpus, mají k jednání blíž. Obdobně postupuje i domácí výzkum, který reaktivní a proaktivní strategie zachycoval přímým pozorováním praxe (Vlčková et al., 2020) nebo analýzou konkrétních kázeňských situací začínajících učitelů (Karasová & Nehyba, 2025). Druhým předpokladem je, že má-li se repertoár rozšiřovat, nestačí předkládat poznatky; je třeba nabízet alternativní pohledy na konkrétní situace v okamžiku, kdy o nich učitel přemýšlí. Na tom staví platforma Edustories, která ke vložené kazuistice nabízí řešení v šesti přístupech (kapitola 3). Čtyři přístupy z oddílu 2.2 tvoří jen část této nabídky.

Repertoár nejmenšího odporu

Ze spojení učňovství pozorováním s časovým tlakem reálné situace vyplývá interpretační hypotéza: repertoáry učitelů budou vychýleny k postupům, které jsou ve školní zkušenosti nejviditelnější a v situaci nejdostupnější, tedy k reaktivním, individuálně zaměřeným zásahům typu domluvy, napomenutí nebo trestu. Proaktivní, systémové a na pojmenovaných rámcích postavené postupy budou naopak vzácné. S touto hypotézou kniha porovnává četnosti zaznamenaných postupů; samotný mechanismus vzniku repertoáru empiricky netes-

2. Teoretická východiska

tuje. Poznatky oboru přitom ukazují opačným směrem: metaanalýza programů a strategií řízení třídy dokládá jejich přínos (Korpershoek et al., 2016) a přehled praktik s empirickou oporou klade důraz na prevenci a strukturu (Simonsen et al., 2008). Napětí mezi tím, co obor doporučuje, a tím, co repertoáry obsahují, je jedním z ústředních motivů knihy. Repertoár vychýlený k nejsnáze dostupným postupům označujeme jako repertoár nejmenšího odporu; pojem zde zavádíme, Studie 1 mu dává empirický obsah a kapitola 9 jej rozvíjí.

Osy popisu řešení

Aby bylo možné srovnávat učitelská a generovaná řešení, popisuje je kniha ve Studii 3 na třech osách, z nichž dvě vycházejí z literatury o řízení třídy. Čtvrtou, hodnoticí dimenzí je celková vhodnost řešení pro danou situaci. Osa reaktivní–proaktivní odlišuje zásah po incidentu od utváření podmínek, které incidentům předcházejí; není hodnotově zcela neutrální, protože prevence bývá v přehledové literatuře spojována s příznivějšími výsledky (Korpershoek et al., 2016; Simonsen et al., 2008). Osa humanistická–behaviorální odlišuje důraz na vztah, potřeby a dialog od důrazu na pravidla, důsledky a posilování; navazuje na dichotomii popsanou v české literatuře o řízení třídy (Lukas & Lojdová, 2018). Osa systémová–situační odlišuje zásah na úrovni třídy či školy od zásahu mířícího na konkrétní incident. Osa systémová–situační je popisná. Osa humanistická–behaviorální je v kódovacím manuálu vymezena zčásti i hodnotově: behaviorální pól tam vedle důrazu na pravidla a důsledky zahrnuje také direktivnost a neohlížení se na potřeby žáka (příloha A.2). Hodnoticí povahu má vedle osy reaktivní–proaktivní především celková vhodnost. Všechny čtyři dimenze posuzovaly anotátorky při zaslepeném srovnání učitelských a generovaných řešení.

2.4 Velké jazykové modely v pedagogickém výzkumu

Od tématu k nástroji

Ve vzdělávacím výzkumu se o velkých jazykových modelech nejprve uvažovalo jako o tématu: jako o technologii, jejíž příchod do škol je třeba promyslet. Vlivná stať shrnula příležitosti (personalizace, podpora učitelů, tvorba materiálů) i rizika (halucinace, zkreslení, ochrana dat, závislost) a zdůraznila potřebu kompetencí na straně učitelů i žáků (Kasneci et al., 2023). Systematický přehled využití jazykových modelů ve vzdělávání pak vedle zmapování automatizovaných úloh upozornil na slabou replikovatelnost a transparentnost publikovaných inovací a na nedostatečné ošetření etiky (Yan et al., 2024). Tato kniha navazuje na jiný, mladší proud výzkumu: modely v ní slouží jako výzkumný nástroj pro analýzu rozsáhlého korpusu pedagogického textu a kniha jejich spolehlivost zároveň prověřuje. Předpokladem takového užití je ověřit, nakolik se kódování modelu shoduje s lidským kódováním, a znát hranice jeho spolehlivosti; shoda přitom sama nedokládá správnost.

Model v roli anotátora

Diskusi o jazykových modelech jako anotátorech otevřely v roce 2023 studie z výpočetní společenské vědy. Na vzorcích tweetů a novinových článků se ukázalo, že model bez příkladů v instrukci (zero-shot) překonává v anotačních úlohách placené anotátory najímané přes internetové platformy (crowdsourcing): přesnost byla v průměru výrazně vyšší, konzistencí model převyšoval i vyškolené anotátory a náklady byly řádově nižší (Gilardi et al., 2023). Navazující práce s daty z jedenácti zemí ukázala, že model při určování politické příslušnosti autorů zpráv překonává nejen anotátory z platforem, ale i expertní kodéry a klasifikátory trénované na ručně anotovaných datech, a to napříč jazyky (Törnberg, 2025). Tyto výsledky vysvětlují, proč se model v roli anotátora (LLM-as-annotator) rychle začal běžně zvažovat všude tam, kde lidské kódování omezuje rozsah obsahové analýzy.

Další výzkum počáteční optimismus zpřesnil. Systematické hodnocení řady modelů na

2. Teoretická východiska

pětidvaceti úlohách výpočetní společenské vědy ukázalo, že v klasifikačních úlohách modely zpravidla nepřekonávají specializované klasifikátory dodatečně trénované na konkrétní úlohu; autoři je doporučují chápat jako partnery lidského výzkumníka, ne jako jeho náhradu (Ziems et al., 2024). Metodologická stať, zatím publikovaná jako nerecenzovaný preprint, upozornila, že výstupy modelu jsou nedeterministické a citlivé na drobné změny instrukce, takže konzistence může klesnout pod prahy vědecké reliability; její autor proto varuje před nasazením modelu bez ověření na lidsky anotovaných datech (Reiss, 2023). Kritika z hlediska otevřené vědy připomněla, že spoléhání na proprietární modely ohrožuje reprodukovatelnost výzkumu, protože poskytovatelé modely průběžně mění bez verzování a přístup k nim je uzavřený a zpoplatněný (Ollion et al., 2024).

Z této literatury kniha přejímá tři zásady: vykládat shodu modelu s lidmi vždy spolu s reliability lidského kódování; ověřovat model na vlastních datech, a ne spoléhat na cizí srovnávací sady úloh; a dokumentovat verze modelů i instrukcí (kapitola 4, přílohy B a C). Lidskou reliability přitom lze považovat za strop shody jen tehdy, když obě míry pocházejí ze stejného uspořádání a vzorku, což ve Studii 2 neplatí (oddíl 6.4).

Model v roli posuzovatele a generátora

Vedle anotace se ustálila role modelu jako posuzovatele kvality textových výstupů (LLM-as-judge). Zakládající studie tohoto přístupu ukázala, že výkonný model se při posuzování odpovědí shoduje s lidskými preferencemi podobně, jako se lidé shodují mezi sebou. Zároveň popsala zkreslení takového posuzování, zejména citlivost na pořadí posuzovaných odpovědí a preferenci delších textů; náznak, že model upřednostňuje vlastní výstupy, autoři na svých datech nepotvrdili (Zheng et al., 2023). Třetí rolí je model jako generátor obsahu, v pedagogickém kontextu typicky návrhů řešení či zpětné vazby. Kniha empiricky prověřuje jinou trojici: anotátora ve Studii 2, generátora ve Studii 3 a prediktora dopadu ve Studii 4. Generovaná řešení ve Studii 3 hodnotily zaslepeně lidské anotátorky; role modelu jako posuzovatele proto zůstává v knize jen teoretickým kontextem. Kniha navazuje i na dřívější

společnou práci svého autora s hlubokými jazykovými modely nad reflektivními texty studujících učitelství (Nehyba & Štefánek, 2023).

Jazyková a doménová mezera přenosu

Zakládající studie, o které se oddíl opírá, mají společný rys: téměř všechny pracují s anglickými texty z politické komunikace a sociálních médií. Tweety a novinové články, na nichž se ukázala převaha modelu nad anotátory (Gilardi et al., 2023), jsou texty krátké, veřejné a žánrově ustálené; i rozšíření na více zemí zůstalo u politických zpráv (Törnberg, 2025) a úlohy, na nichž se modely hodnotily ve výpočetní společenské vědě, pokrývají příbuzné žánry (Ziems et al., 2024). Kazuistiky v tomto korpusu jsou jiné: dlouhé, narativní, emočně nasycené a psané česky, tedy v jazyce, který je v trénovacích datech modelů zastoupen řádově slaběji než angličtina. Kódovací kategorie navíc vycházejí z pedagogické teorie. Přenositelnost zakládajících výsledků na takový materiál je proto otevřená empirická otázka. Totéž platí pro roli posuzovatele a generátora: pedagogická kvalita navrženého řešení je jemnější kritérium než preference mezi dvěma odpověďmi chatbota. Kniha proto role anotátora, generátora a prediktora ověřuje přímo na českém pedagogickém materiálu: anotaci a predikci porovnává s lidským kódováním, generovaná řešení podrobuje zaslepenému lidskému hodnocení a u každé úlohy uvádí i shodu lidských anotátek. Roli posuzovatele samostatně neověřuje.

Český kontext

Česká recenzovaná literatura o umělé inteligenci v pedagogice je mladá; významnou část tvoří monotematické číslo časopisu *Pedagogika* věnované umělé inteligenci ve vzdělávání. Srovnávací studie národních politik dokládá rozdílnou připravenost zemí na začlenění generativní umělé inteligence do škol a formuluje podmínky, za nichž má toto začlenění smysl (Černochová et al., 2025). Text z rubriky příkladů inspirativní praxe představuje využití generativních nástrojů podle úrovně Bloomovy taxonomie (Kopecký et al., 2025). Přehledová

2. Teoretická východiska

studie popisuje selhávání modelů na klasických logických úlohách a vykládá je jako principiální mez algoritmického porozumění (Hončík et al., 2025). Vymezením gramotnosti v oblasti umělé inteligence se zabývá domácí přehledová studie (Černý, 2024). Pro rané období českého diskurzu jsou příznačné nerecenzované texty o ChatGPT ve školní praxi, které vycházely jako série na metodickém portálu (Černý, 2023).

O rychlosti, s jakou generativní nástroje pronikly do českých škol, vypovídají především rozsáhlé průzkumy. Podle dotazníkového šetření mezi tisíci učiteli z jara 2023 měla už tehdy s nástroji umělé inteligence zkušenost zhruba třetina z nich a většina očekávala proměnu výukových metod i nároků na dovednosti (Kopecký et al., 2023). Navazující šetření mezi desítkami tisíc žáků základních a středních škol ukázalo, že ChatGPT používá nadpoloviční většina žáků, výrazně častěji mimo školu než ve škole (Kopecký & Voráč, 2024). Ministerstvo reagovalo sadou doporučení pro ředitele, učitele, žáky a rodiče, která z jeho podnětu vyšla před školním rokem 2023/2024 (Národní pedagogický institut ČR, 2023). České školy se tedy s generativní umělou inteligencí setkávají ve velkém a česká pedagogika ji začala zkoumat jako téma. Jazykový model jako kriticky prověřený nástroj pedagogického výzkumu, použitý na velký korpus autentického českého textu a porovnaný s lidským kódováním, se v domácí literatuře podle našich znalostí zatím neobjevuje; výjimkou je dřívější autorova práce zmíněná výše. Na tuto roli modelu se zaměřuje tato kniha.

2.5 Průsečík obou linií a výzkumná mezera

Obě linie kapitoly se potkávají v jednom materiálu: v rozsáhlém, průběžně rostoucím korpusu autentických kazuistik náročného chování, psaných česky převážně studujícími učiteli a také učiteli. V pedagogické linii jde o otázku, co tento materiál vypovídá o repertoárech učitelů: jaké kategorie chování a řešení v něm převažují, jak se řešení kombinují a s jakými dopady, a jak daleko leží od normativního horizontu čtyř pojmenovaných přístupů. V metodologické linii jde o otázku, nakolik lze při práci s takovým materiálem spoléhat na jazykový model: zda se jeho kódování obsahu shoduje s lidským, zda jím generovaná řešení obstojí v

zaslepeném lidském hodnocení a kde jeho spolehlivost končí. Odpovědi na obě otázky na sobě závisejí. Pedagogické závěry z velkého korpusu jsou dosažitelné jen s nástrojem použitelným ve velkém rozsahu a nástroj lze prověřit jen na materiálu, jehož pedagogický obsah je nezávisle popsán.

Kniha se proto zaměřuje na tři okruhy. Prvním je popis učitelských repertoárů řešení náročného chování opřený o velký soubor autentických situací; takových popisů je v české pedagogice podle našich znalostí zatím málo. Druhým je ověření jazykového modelu v roli anotátora, generátora a prediktora na pedagogickém materiálu v jiném jazyce než angličtině, a to vůči lidskému hodnocení s vykázanou reliabilitou. Zakládající studie o modelu v roli anotátora, citované v oddílu 2.4, pracují převážně s anglickými texty z politické komunikace a sociálních médií. Třetím okruhem je propojení obou otázek v jedné práci, které plyne z jejich vzájemné závislosti popsané výše. Kapitola 3 popisuje projekt a platformu, kapitola 4 metodologii a kapitoly 5 až 8 čtyři empirické studie.

3. Projekt Edustories: kontext vzniku dat

Tato kapitola popisuje prostředí, ve kterém vznikla veškerá data této monografie: projekt Edustories. Vysvětluje, proč a jak se sbírají kazuistiky náročného chování žáků, jak vypadá platforma, která je zpřístupňuje učitelské veřejnosti, a proč jsou tytéž kazuistiky zároveň vzdělávacím obsahem pro učitele a výzkumným korpusem pro studium možností velkých jazykových modelů v pedagogickém výzkumu.

3.1 Východisko: proč sbírat kazuistiky náročného chování

Náročné chování žáků patří k nejčastěji uváděným problémům začínajících učitelů a k oblastem, na které se cítí nejméně připraveni. Tak to popsal už klasický přehled studií o začínajících učitelích (Veenman, 1984) a potvrzuje to i současný domácí výzkum (Karasová & Nehyba, 2025). Podle mezinárodního šetření TALIS 2024 platí tento obraz i pro Českou republiku: jen 31 % českých učitelů souhlasí, že jim počáteční vzdělávání dalo nápady, jak úspěšně zvládat chování ve třídě, což je nejnižší podíl ze všech zúčastněných zemí, a zvládnutí chování žáků je nejčastěji uváděnou oblastí velké potřeby dalšího vzdělávání, zvláště u začínajících učitelů (Havlíčková et al., 2025; OECD, 2025). Kniha přitom vychází z předpokladu (oddíl 1.1), že příprava učitelů zprostředkuje obecné principy snáze než repertoár konkrétních, situačně ukotvených řešení. Kazuistiky (případové studie) se v přípravě pomáhajících profesí používají jako jeden ze způsobů, jak tento rozdíl zmenšovat: popis situace, jejího pozadí, zvoleného řešení a jeho výsledku může čtenáři přiblížit rozhodovací situaci dřív, než v ní bude stát sám.

3. Projekt Edustories: kontext vzniku dat

Projekt Edustories průběžně a ve velkém sbírá kazuistiky, které píší především studující učitelství na praxi, a zpřístupňuje je jako rostoucí databázi. Rozsah sbírky otevírá druhou, metodologickou otázku: jak s tak velkým korpusem nestrukturovaného pedagogického textu pracovat ve výzkumu, když ruční analýza naráží na kapacitní hranice (oddíl 1.1). Obě linie, pedagogická i metodologická, se v knize prolínají.

3.2 Projekt Edustories

Edustories je aplikovaný výzkumný projekt řešený na Pedagogické fakultě Masarykovy univerzity (MU) a podpořený Technologickou agenturou ČR v programu SIGMA (projekt TQ01000030, září 2023 až říjen 2026). Plný název, „Využití umělé inteligence a virtuální reality v rozvoji kompetencí k řešení náročného chování ve třídě“, vystihuje jeho aplikační zaměření: cílem je podpořit studující učitelství a začínající učitele v kompetencích ke zvládnutí náročného chování, a to dvěma vzájemně se doplňujícími, volně dostupnými cestami. První je webová platforma psaných kazuistik s podporou umělé inteligence; druhou je větev virtuální reality se simulovanými náročnými situacemi, která stojí mimo záběr této knihy. Projekt je založen multidisciplinárně na pomezí pedagogiky, psychologie a počítačové lingvistiky.

Na projektu je zároveň vidět, jak rychle se v posledních letech měnily technologické podmínky pedagogického výzkumu. Návrh vznikl na podzim 2022 a počítal s tehdejšími stavem jazykových technologií: se sémantickým vyhledáváním podobných kazuistik, s algoritmy strojového učení pro predikci na základě anotací a s tréninkem respektující komunikace prostřednictvím konverzačního chatbota na specializované platformě. Jen několik týdnů po dokončení návrhu však byl zveřejněn ChatGPT a nástup velkých jazykových modelů proměnil technické řešení projektu dříve, než začal. Generování alternativních řešení v pojmenovaných rámcích, jazyková podpora anonymizace i strojová klasifikace obsahu dnes stojí na velkých jazykových modelech. Záměr projektu, tedy rozvoj učitelských kompetencí skrze kazuistiky, se nezměnil; vyměnily se nástroje. Tato proměna je jedním z důvodů, proč

knihy zkoumá jazykové modely jako výzkumný nástroj: projekt je využívá a Studie 2 až 4 porovnávají jejich výstupy s lidským úsudkem.

Hlavním výstupem projektu je veřejná webová platforma edustories.cz. Ta zpřístupňuje zveřejněné kazuistiky s fulltextovým a významovým vyhledáváním. Jejím datovým základem je anonymizovaná databáze, která je rovněž výstupem projektu TA ČR. Platforma ke kazuistikám nabízí vedle autentického učitelského řešení také řešení generovaná umělou inteligencí v šesti pojmenovaných rámcích (proaktivní a reaktivní přístup, sociálně-emoční učení, pozitivní behaviorální podpora, restorativní praxe a nenásilná komunikace) a doplňuje je o gamifikované tréninkové moduly. Vedle platformy projekt vydává výzkumnou datovou sadu Edustories-en (oddíl 3.4) a e-learningovou oporu v univerzitním informačním systému. Ani platforma, ani datová sada Edustories-en nezpřístupňují celou databázi ani její dřívější řezy. Pro tuto monografii je podstatné, že databáze platformy slouží i jako úložiště sběru: kazuistiky odevzdané v Informačním systému MU se do ní průběžně importují, databáze roste a každý její časový řez představuje jiný stav korpusu. Důsledky tohoto růstu pro analýzu rozebírá kapitola 4; přesné počty kazuistik v jednotlivých řezech uvádí kapitola 5.

3.3 Sběr kazuistik: zrcadlová šablona

Jádrem sběru jsou povinné pedagogické praxe studujících učitelství na MU. Psaní kazuistiky si studující mohli zvolit jako úkol praxe a její poskytnutí k výzkumu bylo dobrovolné. Řez z května 2026 obsahuje v menší míře i záznamy z jiných zdrojů, například ručně vložené na platformě nebo ze zahraničí. Studující podle strukturované šablony zaznamenávají situaci náročného chování, kterou na praxi zažili sami nebo kterou řešil jejich provázející učitel, a popíše, v jaké roli v ní vystupovali a jak byla řešena. Šablona má dvě zrcadlové části: kazuistiku úspěšného řešení a kazuistiku řešení neúspěšného. Každá kazuistika má předepsanou čtyřdílnou strukturu: popis vzniku situace na úrovni chování, anamnézu žáka či třídy, popis zvoleného řešení a popis výsledku. K tomu šablona sbírá deskriptivní údaje o vyučujícím v dané situaci (aprobace, délka praxe, absolvované kurzy zvládání náročného chování) a o

3. Projekt Edustories: kontext vzniku dat

žákovi (věk, ročník, diagnózy, zájmy).

Zápis kazuistiky má i didaktický smysl: pisatel situaci zpětně rozkládá na pozorovatelné chování, jeho pozadí, zvolené řešení a výsledek. Pro studující tak může být psaní kazuistiky učební příležitostí, která navazuje na reflektivní semináře doprovázející praxe. Pro projekt je výhodou, že texty vznikají v prostředí, v němž lze pisatele u zdroje poučit o anonymizaci a ohlídat náležitosti informovaného prohlášení (oddíl 3.4). Z povahy sběru plyne i složení korpusu, s nímž musí analýzy počítat: převažuje perspektiva studujících na praxi; věkové a další charakteristiky popsaných situací popisuje kapitola 5.

Zrcadlové uspořádání šablony má vlastní zdůvodnění. Sbírky dobré praxe mohou podléhat chybě přeživších: neúspěchy se nezapisují a čtenář pak může nabýt dojmu, že popsaná řešení fungují. Šablona proto od téhož pisatele žádá úspěšně i neúspěšně řešenou situaci, takže neúspěch je v záznamu rovnocennou datovou kategorií. Pro výzkum repertoárů jsou neúspěšná řešení přinejmenším stejně informativní jako úspěšná, protože ukazují, které postupy učitelé volí, i když nefungují. Z jednoho odevzdání tak mohou v databázi vznikat dvě samostatné kazuistiky; kniha proto počítá jednotky korpusu jako kazuistiky, nikoli jako odevzdání (oddíl 4.3). Čtyřdílná struktura navíc usnadňuje strojové zpracování: kazuistika rozdělená na čtyři části je jednotkou všech anotačních etap i všech experimentů s jazykovými modely v této knize.

3.4 Od textů ke korpusu: extrakce, anonymizace, publikace

Mezi odevzdaným textem studujícího a kazuistikou v databázi leží postup zpracování, který podrobně popisují přílohy B a D. Zde jej shrnujeme ve třech krocích. Nejprve extrakce a strukturace: odevzdané soubory se programově rozdělují na jednotlivé kazuistiky a jejich čtyři části; tam, kde struktura není explicitní, asistuje jazykový model. Poté anonymizace, pojatá jako trojstupňová s následnou ruční kontrolou: první stupeň provádějí sami pisatelé (poučení, aby neuváděli identifikační údaje), druhý stupeň automatické rozpoznávání a náhrada jmen osob a míst (tzv. pojmenovaných entit) v českém textu, třetí stupeň jazykový

3.4 Od textů ke korpusu: extrakce, anonymizace, publikace

model. Od července 2026 anonymizují databázi po lokálním rozpoznávání pojmenovaných entit nástrojem NameTag (model CNEC) jazykové modely provozované v infrastruktuře e-INFRA CZ na serverech Masarykovy univerzity. Nakonec publikace: kazuistiky, které projdou kontrolou kvality, se zveřejňují na platformě; část korpusu zůstává neveřejná. Publikací status je proto samostatnou proměnnou korpusu a analýzy s ním musí výslovně pracovat. Pro výzkumnou komunitu je zveřejněn kurátorovaný, anotovaný výběr korpusu v češtině s anglickým překladem jako datová sada Edustories-en (Nehyba, Karasová, & Štefánek, 2024); platforma slouží především praxi (studujícím učitelství, učitelům a dalším zájemcům), datová sada slouží výzkumu.

Rozlišení mezi platformou a datovou sadou je podstatné, protože jde o dvě různá rozhodnutí. O publikaci na platformě rozhoduje redakce platformy: kazuistika prošla anonymizací a redakční kontrolou kvality a je užitečná čtenáři z praxe. Zařazení do výzkumné datové sady je kurátorské rozhodnutí: do výběru vstupovaly kazuistiky, které prošly lidským kódováním první etapy (většinou jednou anotátorkou), takže nesou anotace, s nimiž může výzkumná komunita dále pracovat. Datová sada proto není kopií databáze, ale jejím anotovaným výběrem k určitému datu; její rozsah a vztah k řezům korpusu uvádí kapitola 5. Pro reprodukovatelnost knihy z toho plyne praktický důsledek: analýzy neběží nad „živou“ platformou, ale nad datovanými řezy databáze; seznamy kazuistik, s nimiž pracují analýzy studií, jsou uloženy v doprovodném repozitáři (Nehyba, 2026), jak popisuje kapitola 4.

Sběr probíhá s informovaným prohlášením přispěvatelů; jejich identifikační údaje, které eviduje informační systém, do analyzovaných dat nevstupují. Výzkumný projekt schválila Etická komise pro výzkum MU pod jednacím číslem EKV-2022-118. Anonymizaci jsme ověřili výběrově na náhodném vzorku zveřejněných kazuistik (příloha D.2); anonymizační protokol a znění informovaného prohlášení přispěvatele uvádí příloha D.

3.5 Tři anotační etapy

Nad rostoucím korpusem proběhly tři anotační etapy, na nichž stojí empirické studie knihy. Každá etapa měla vlastní cíl, vlastní kódovací schéma a vlastní posloupnost zaučování anotátorek: od zkušebních anotací přes společné rozборы sporných případů až po měřenou shodu a teprve poté finální anotování. V první etapě (2023 až 2024) anotátorky kodovaly obsah kazuistik do kategorií typů chování, typů řešení a dopadů; na ni navázalo srovnání lidského kódování s kódováním jazykovým modelem (Studie 2). Ve druhé etapě (2025) anotátorky v zaslepeném uspořádání hodnotily kvalitu učitelských řešení a řešení generovaných umělou inteligencí (Studie 3). Ve třetí etapě (2026) posuzovaly kvalitu samotných kazuistik, dopad řešení a jeho vhodnost (Studie 4, kapitola 8). Deskriptivní obraz toho, co kazuistiky obsahují, podává na základě první etapy Studie 1.

Tři etapy spolu věcně souvisejí. První etapa vytvořila lidské kódování obsahu, které kniha využívá dvojím způsobem: popisuje jím korpus (Studie 1) a porovnává s ním strojovou klasifikaci (Studie 2). Shoda lidí mezi sebou, měřená v jiném uspořádání, přitom není stropem pro shodu modelu s lidským kódováním (kapitola 6). Ve druhé etapě se vztah obrátil: lidé nekontrolovali model v úloze, kterou sami umějí, ale hodnotili dvojice řešení, aniž jim instrukce sdělila, že jedno z nich vygeneroval model (krycí instrukce). Instrukce měla hodnocení chránit před zkreslením očekáváním; její meze i meze zaslepení rozebírá kapitola 7. Třetí etapa zjišťovala, jak spolehlivě lze posoudit kvalitu kazuistiky a úspěšnost řešení, tedy konstrukty, u nichž se obtížně shodují i lidští posuzovatelé. Etapy tak postupně zužují otázku, co lze jazykovému modelu v pedagogickém výzkumu svěřit a co nikoli.

Část kazuistik prošla více etapami, takže se studie částečně opírají o tytéž kazuistiky. Evidenci těchto překryvů, multiplicitu anotací a pravidla, jak s nimi analyticky nakládat, popisuje oddíl 4.5 a dokládá příloha C.

4. Metodologie

Kapitola popisuje metodologii knihy jako celku, aby jednotlivé studie mohly odkazovat na společný rámec: jak chápeme korpus v čase, jak proběhly tři anotační etapy, jak vznikala data od textu k analyzovatelné podobě, jaká kódovací schémata se použila, jak nakládáme s opakovanými anotacemi téže kazuistiky, jaké platí etické náležitosti a jak lze výpočty ověřit.

4.1 Korpus jako časová řada řezů

Data nepocházejí z jednorázového sběru, ale z databáze, která v čase roste (kapitola 3). Každá analýza se proto váže na pojmenovaný časový řez korpusu, nikoli na „korpus“ obecně. Kniha pracuje s řezy od roku 2023 po řez z května 2026 (přehled uvádí Tabulka 5.1 ve Studii 1). Z tohoto pojetí plynou dva důsledky, které platí napříč studiemi. Čísla z různých řezů nelze bez rozmyslu srovnávat ani slučovat; rozdíly ve velikosti korpusu odrážejí jeho růst a stahování kazuistik z publikace, nikoli chybu. Publikační status kazuistiky (zveřejněná, či nezveřejněná) je navíc samostatná proměnná: řez z května 2026 obsahuje pouze zveřejněné kazuistiky, přičemž část dříve zveřejněných byla mezitím stažena. Analýzy proto s publikačním statutem pracují výslovně a zveřejněné a nezveřejněné kazuistiky bez uvedení nesměšují.

4.2 Tři anotační etapy a jejich zaučování

Nad korpusem proběhly tři anotační etapy (kapitola 3). Každá měla stejnou vnitřní logiku: zaučování anotátorek na malém vzorku, jedno či více kalibračních kol se společným rozbořem sporných případů, změření shody a teprve poté finální anotování. Tato posloupnost je pro výklad spolehlivosti důležitá: shodu uvádíme z finálních kol, nikoli ze zácviků, a případný vývoj shody mezi kalibrací a finálním kolem popisujeme jako samostatné zjištění (kapitola 9). Úplný registr etap, jejich kroků a zúčastněných anotátorek (pod pseudonymy) uvádí příloha C.

Finální kolo první etapy (kvalitativní obsahové kódování) provádělo pět anotátorek formou dělby práce: každá kodovala jiné kazuistiky, aby se pokryl co největší rozsah korpusu, s vestavěným překryvem pro kontrolu shody. Druhá etapa (hodnocení kvality řešení) zapojila šest anotátorek, které měly v zaslepeném uspořádání nezávisle ohodnotit každou dvojici řešení (chybějící hodnocení rozebírá kapitola 7). Materiál etapy tvořilo 324 kazuistik, které platforma v době výběru vedla v kategoriích fyzické a verbální agrese, s úplným popisem situace i učitelského řešení. Ke každé z nich jazykový model vygeneroval alternativní řešení; instrukce požadovala přibližně stejnou délku a styl jako u učitelského řešení (podrobně Studie 3 a příloha B).

Třetí etapa

Třetí etapa (kvalita kazuistik, dopad a vhodnost řešení) proběhla se dvěma anotátorkami, v knize pod pseudonymy A2 a A12, nad téměř celým zveřejněným řezem z května 2026. Řez byl rozdělen na dvě poloviny s vestavěným překryvem: anotátorka A12 ohodnotila 1 947 kazuistik, anotátorka A2 jich ohodnotila 1 938; obě společně hodnotily překryv 693 kazuistik, z něhož se počítá shoda. Etapa měla stejnou posloupnost jako předchozí: zácvik na šesti kazuistikách, první pilot na 19 kazuistikách a druhý pilot na 49 kazuistikách se společným rozbořem neshod, poté finální anotování od března do září 2026. Uprostřed finálního kola, v

červenci 2026, se výklad škály vhodnosti mezi anotátorkami rozešel; proto proběhla rekaliibrace, tedy společný podklad s kotvícími kazuistikami pro jednotlivé stupně škály a schůzka nad spornými kazuistikami (u vhodnosti jich bylo 40, další byly sporné u dopadu nebo kvality). Rekaliibraci uprostřed sběru uvádíme jako změnu protokolu a z primárního výpočtu shody vylučujeme sporné kazuistiky všech tří dimenzí, nejen vhodnosti; jejich celkový počet uvádí kapitola 8. Sběr třetí etapy skončil 8. září 2026; analytický řez tvoří finální tabulky obou anotátorek k tomuto dni (dále řez ze září 2026). Změny známek po společné schůzce se promítly přímo do finálních tabulek, a tedy i do řezu. Průběh etapy, shodu i rozdíly mezi částmi etapy před rekaliibrací a po ní popisuje Studie 4 (kapitola 8).

4.3 Od textu k datům

Odevzdané texty procházejí postupem zpracování, jehož tři kroky (extrakce a strukturace, vícestupňová anonymizace, publikace) shrnuje kapitola 3. Z metodologického hlediska jsou podstatné tři body. Extrakce dělí odevzdaný soubor na jednotlivé kazuistiky a jejich čtyři části, přičemž u části dat asistuje jazykový model; strukturace se proto může u některých kazuistik lišit kvalitou a některé starší kazuistiky nesou stopy nedokonalé extrakce (viz oddíl 4.6). Anonymizace je trojstupňová s ruční kontrolou; její protokol uvádí příloha D a etické náležitosti shrnuje oddíl 4.7. Publikace je selektivní, což zakládá proměnnou publikačního statusu (oddíl 4.1).

Pro analýzy je klíčové, že tytéž kazuistiky se objevují v různých řezech a souborech pod odlišnými identifikátory. Napříč knihou proto pracujeme s jednotným identifikátorem kazuistiky, který propojuje záznamy z různých řezů; tam, kde chybí sdílený identifikátor, se kazuistiky párují přes otisk textu (kontrolní kód vypočtený ze začátku textu po sjednocení velkých a malých písmen a odstranění mezer a interpunkce; shoduje se u textů se shodným začátkem) a nejednoznačné případy se neslučují. Tento krok je předpokladem jakéhokoli srovnání napříč etapami.

Jazykové modely vstupují do dat a analýz knihy při extrakci a strukturaci textů a ve třetím

4. Metodologie

stupni anonymizace (příloha B.1 a B.2), jako klasifikátor ve Studii 2 (příloha B.3), jako generátor srovnávacích řešení ve Studii 3 (příloha B.4) a jako prediktor dopadu ve Studii 4 (modely nasazené na e-INFRA CZ, příloha B.6). Modely použité v jednotlivých krocích uvádí příloha B.

4.4 Kódovací schémata a řezy dat

Každá etapa má vlastní kódovací schéma; schémata se nesměšují. První etapa kodovala tři dimenze (typ chování, typ řešení, dopad) do kategorií, jejichž plné znění uvádí příloha A. Ve druhé etapě anotorky hodnotily řešení na čtyřech ordinálních škálách a označovaly, zda řešení rozpoznatelně odráží některý z pojmenovaných rámců řešení náročného chování. Třetí etapa posuzovala kvalitu kazuistiky, dopad a vhodnost řešení na vlastních škálách. Škály s podobným názvem v různých etapách (například vhodnost ve druhé a třetí etapě) mají odlišný rozsah i význam a nesmějí se slučovat. Jedinou výjimkou, kterou kniha používá, je dopad řešení: první etapa jej kodovala do čtyř kategorií a třetí do tří, a pro srovnání obou etap (kapitola 8) se částečný a krátkodobý úspěch první etapy slučují do jedné kategorie výslovným pravidlem, ne posuzovatelským úsudkem.

Toto sloučení kategorií však neodstraňuje rozdíl mezi popsáním výsledkem a hodnotovým posouzením. Manuál první etapy řadí mezi neúspěchy i eticky nepřijatelné prostředky; třetí etapa může takovou kazuistiku označit za plochou, tedy nepoužitelnou (kapitola 8, příloha A.3). Ze štítku neúspěchu tak nelze poznat, zda vyjadřuje popsáný výsledek, nebo nepřijatelnost prostředku. Asociace s dopadem proto vykládáme jako vztahy mezi takto definovanými kategoriemi, nikoli jako nezávislé měření pedagogické účinnosti.

Deskriptivní výsledky první etapy (Studie 1) uvádíme na pevně vymezeném řezu 1 359 ručně kódovaných kazuistik. Řez tvoří deduplikované kazuistiky z tabulky korpusu ze srpna 2024, jimž byla přiřazena lidská anotace; seznam jeho kazuistik je uložen v doprovodném repozitáři. Že anotační sloupce tabulky nesou lidské kódování, dokládá srovnání s anotačními listy etapy: množiny přiřazených kategorií se shodují u 97 % kazuistik v typu chování, u 95 % v

typu řešení a u 98 % v dopadu; anotační list se podařilo přiřadit k 1 350 kazuistik řezu. Obdobně je vymezen řez třetí etapy z 8. září 2026: finální tabulky obou anotátorek k tomuto dni, spárované s kazuistikami zveřejněného řezu přes jednotný identifikátor (oddíl 4.3); seznam kazuistik řezu je uložen v doprovodném repozitáři.

4.5 Multiplicita anotací

Protože kazuistiky prošly etapami nerovnoměrně, je počítání s opakovanými anotacemi samostatným metodologickým problémem. V první etapě mohlo být jedné kazuistice přiřazeno více kategorií zároveň a ve druhé etapě mohlo být jednomu řešení přiřazeno více rámců; frekvence proto vždy odlišujeme jako počet kazuistik (či řešení) s danou kategorií od počtu přiřazení. Na úrovni celku prošlo 973 kazuistik alespoň dvěma různými etapami a 236 všemi třemi; třetí etapa totiž pokryla téměř celý zveřejněný řez z května 2026, takže se překrývá s oběma předchozími. Počítáme-li jemněji jednotlivé anotační události (táž etapa mohla kazuistiku zasáhnout opakovaně, například v pilotním a finálním kole), bylo ve dvou událostech anotováno 790 kazuistik, ve třech 241, ve čtyřech 20 a v jednom případě v pěti; celkem tak více než jednou událostí prošlo 1 052 kazuistik. Přehled anotačních událostí uvádí příloha C; úplná matice pokrytí (kazuistika $\times$ událost) je součástí doprovodného repozitáře.

Opakované anotace téže kazuistiky proto nepovažujeme za nezávislá pozorování. Tam, kde je to relevantní, modelujeme kazuistiku jako náhodný efekt (smíšené modely ve Studii 3) a hodnocení téže kazuistiky na neslučitelných škálách nikde neslučujeme.

4.6 Znamá omezení dat

Data mají tři známá omezení. Za prvé, 52 kazuistik bylo dotčeno zkrácením textu při automatické extrakci; u menší části z nich byl text v aktuálním řezu opraven, u většiny nikoli, a z analýz závislých na délce textu proto vylučujeme všechny dotčené. Za druhé, kódovací schémata se mezi verzemi mírně vyvíjela (například počet kategorií řešení); verzi schématu

4. Metodologie

proto u každé analýzy výslovně uvádíme a kategorie napříč verzemi nemapujeme mlčky. Za třetí, nízký koeficient shody mezi anotátory může mít dvě odlišné příčiny: skutečnou nejednoznačnost úsudku danou jemností a překryvností kategorií, a silně nevyvážené rozdělení odpovědí, které srazí koeficient i při vysoké hrubé shodě (paradox shody). Protože z koeficientu samotného nelze poznat, která příčina převažuje, uvádíme koeficienty spolu s hrubou shodou a u silně nevyvážených škál třetí etapy doplňujeme i koeficient odolný vůči nevyváženosti (kapitola 8).

4.7 Etika

Kazuistiky v korpusu vznikají od roku 2021; k jejich výzkumnému využití pisatelé udělili informované prohlášení (příloha D.1). Výzkumný projekt schválila Etická komise pro výzkum Masarykovy univerzity pod jednacím číslem EKV-2022-118, s finálním schválením dne 27. 9. 2023. Autoři kazuistik přitom nejsou předmětem výzkumu; jejich postavení rozvádí příloha D. Identifikační údaje přispěvatelů, které eviduje informační systém, do analyzovaných dat nevstupují a kazuistiky jsou anonymizované; v textu knihy nevystupují přímé identifikátory žáků, škol ani pisatelů. Anotátorky uvádíme pouze pod projektovými pseudonymy A1 až A12. Texty nezveřejněných kazuistik se nedostávají do žádných veřejných výstupů knihy. Anonymizační protokol uvádí příloha D.

4.8 Reprodukovatelnost

Příprava dat probíhá v jazyce Python, inferenční statistika v jazyce R. Analýzy jsou zapsány jako skripty, hlavní analýzy čtyř studií jako spustitelné dokumenty, které z odvozených datových tabulek vytvoří výslednou zprávu a tabulku výsledných hodnot (název míry a její hodnota, spolu s údajem o skriptu, verzi vstupních dat a nastavení náhodného výběru). Výsledky analýz uvedené v textu pocházejí z těchto tabulek. Analýzy pro tuto knihu se řídily písemným analytickým plánem; nejde o veřejnou předregistraci. Studie 2 a 3 přepočítávají

analýzy týchž dat provedené dříve v projektu a v rukopisech článků; analýzy Studie 4 proběhly až po sepsání plánu.

Doprovodný repozitář (Nehyba, 2026) zpřístupňuje analýzy v R, tabulky výsledných hodnot, zprávy z analýz ve formátu HTML, skripty obrázků i veřejnou datovou podmnožinu bez textů kazuistik a bez osobních údajů; jeho součástí je český návod, jak si čísla knihy ověřit bez programátorských znalostí, a konfigurace prostředí pro opakování analýz.

Studie 1 a 2 jsou přepočitatelné z veřejných odvozených dat, která nenesou texty kazuistik ani jména: Studie 1 z tabulky počtů slov, věku a publikačního statusu na úrovni kazuistiky a z anotačních štítků, Studie 2 z tabulky kategorií přiřazených člověkem a modelem. Studie 3 se přepočítá z tabulky hodnocení dvojic řešení, v níž jsou anotátorky pod pseudonymy a která neobsahuje texty kazuistik; její úplná podoba (včetně textů) je uložena v balíčku na platformě OSF (odkaz bude zveřejněn s publikací souvisejícího článku, do té doby na vyžádání). Studie 4 se přepočítá z tabulky hodnocení třetí etapy, opět bez textů a pod pseudonymy. Odvozené tabulky umožňují přepočet statistik, nikoli opakování samotného kódování nebo generování odpovědí. Součástí reprodukce jsou i uložené predikce; nové volání modelu by bylo novým experimentem.

Empirické studie

5. Studie 1: Korpus a jeho obsah

První studie odpovídá na otázku, jak vypadá vybudovaný korpus kazuistik a co obsahuje (VO1). Popisuje rozsah a růst korpusu (5.1), základní charakteristiky kazuistik (5.2), typy chování, řešení a dopadů, které kazuistiky zachycují (5.3), souběhy řešení s typy chování (5.4) a vztah řešení k vnímanému výsledku (5.5). Všechny analýzy kapitoly jsou deskriptivní a kauzální závěry z nich nevyvozujeme.

5.1 Rozsah a růst korpusu

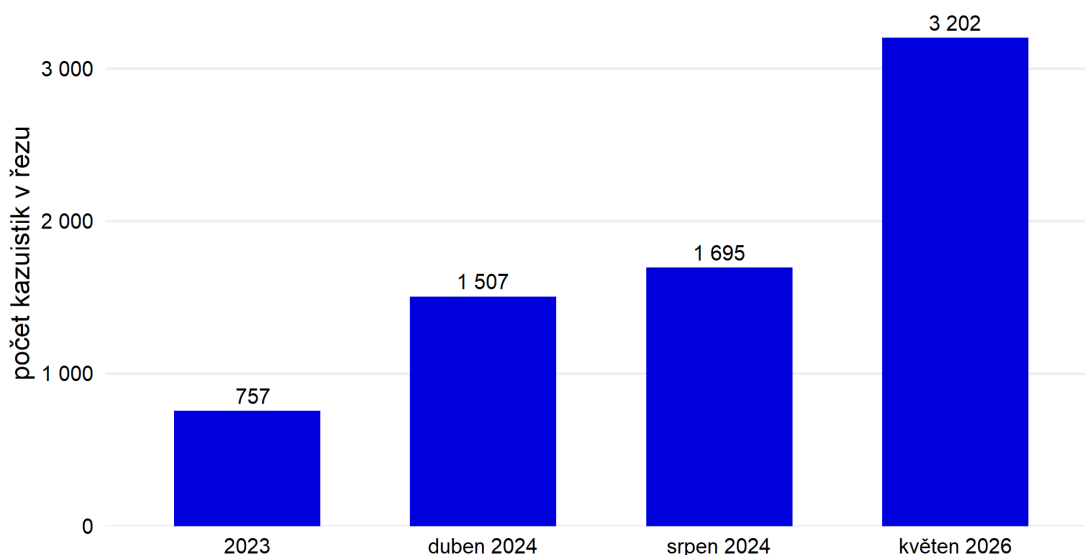
Korpus je průběžně rostoucí databáze (kapitola 3). Analýzy se proto vždy vážou na pojmenovaný časový řez. Tabulka 5.1 shrnuje velikost korpusu v jednotlivých řezech od nejstaršího sběru po aktuální stav.

Tabulka 5.1. Růst korpusu v pojmenovaných časových řezech.

Řez	Počet kazuistik (u veřejné datové sady počet záznamů)
2023	757
duben 2024	1 507
srpen 2024	1 695
veřejná datová sada*	1 492
květen 2026	3 202

5. Studie 1: Korpus a jeho obsah

Poznámka. Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [výstupní tabulka ↗](#)). Definice řezů uvádí kapitola 4. *Veřejná datová sada není chronologickým bodem růstu: jde o kurátorovaný výběr z řezu srpen 2024, zveřejněný jako samostatný datový produkt; uvedený počet jsou záznamy výběru, nikoli unikátní kazuistiky (příloha C.2).

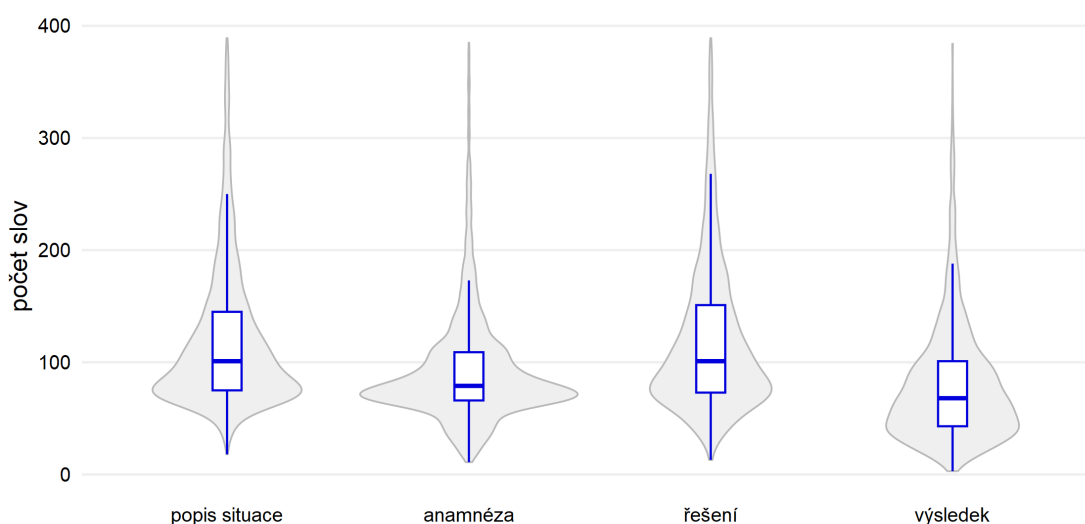


Obrázek 5.1. Růst korpusu v chronologických řezech (hodnoty viz Tabulka 5.1; veřejná datová sada jako kurátorovaný výběr není bodem růstu, proto ve figuře chybí). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Nejnovější řez z května 2026 obsahuje 3 202 zveřejněných kazuistik. Růst přitom není prostým přibýváním: srovnání se starším řezem z téhož období ukazuje, že 395 kazuistik, které v něm byly, se v aktuálním řezu již nevyskytuje. Mezitím byly staženy z publikace nebo z databáze odstraněny; takový pohyb patří k redakčnímu životu databáze. Pro výzkum z toho plyne, že „korpus“ bez data není definovaný objekt a replikace se musí vztahovat k datovému řezu, ne k živé databázi. Analýzy proto pracují s publikačním statusem jako se samostatnou proměnnou (kapitola 4). Veřejně dostupná datová sada (kapitola 3) je kurátorovaný výběr z řezu ze srpna 2024 čítající 1 492 záznamů, jež odpovídají 1 246 unikátním kazuistikám (příloha C.2). Od plného řezu se neliší jen počtem: do výběru vstupovaly kazuistiky s dokončeným lidským kódováním.

5.2 Základní charakteristiky kazuistik

Kazuistiky mají čtyřdílnou strukturu (kapitola 3) a jsou to souvislé texty střední délky. V aktuálním řezu má popis situace medián 102 slov (mezikvartilové rozpětí 76 až 148), popis řešení medián 102 slov (74 až 154), anamnéza medián 79 slov (66 až 109) a popis výsledku medián 68 slov (43 až 102). Rozdělení délek jsou pravostranně zešíkmená s dlouhým chvostem podrobných textů; desetina popisů situace přesahuje 221 slov (Obrázek 5.2). Ze statistik délky jsme vyřadili všech 52 kazuistik, jejichž text byl dříve zkrácen při automatické extrakci (kapitola 4).



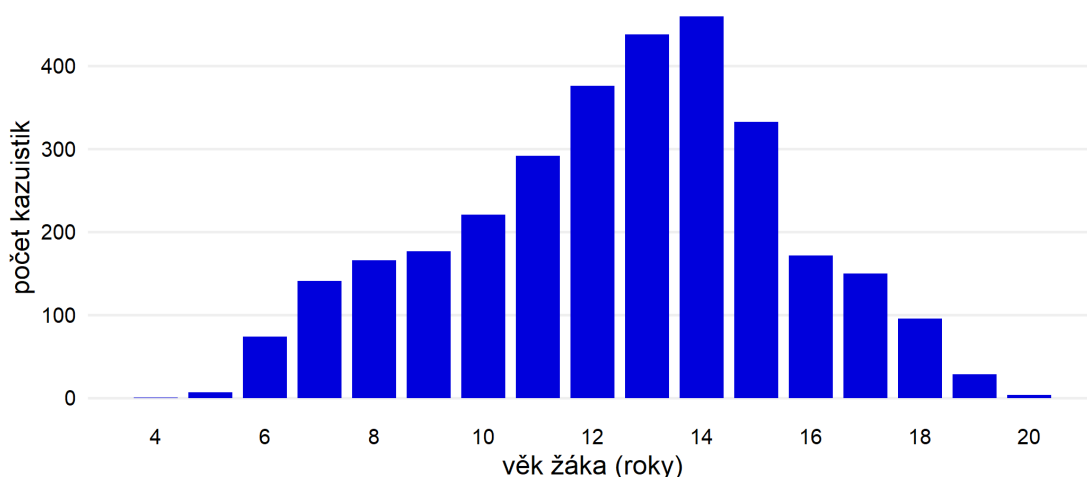
Obrázek 5.2. Rozdělení délek čtyř částí kazuistiky (počet slov; houslový graf s vyznačeným mezikvartilovým rozpětím a mediánem; pro čitelnost oříznuto na 99. percentil). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Pisatelé věnují nejvíce slov popisu situace a řešení; popis výsledku bývá stručnější a v mediánu je nejkratší částí kazuistiky. Pro pedagogické čtení to znamená, že korpus je bohatší na to, co se v situaci dělalo, než na to, jak to dopadlo. Pro Studii 4 z toho plyne, že lidské posouzení dopadu se opírá o nejkratší část kazuistiky, kterou jazykový model při predikci nevidí vůbec.

Medián věku žáka, jehož se kazuistika týká, je 13 let (mezikvartilové rozpětí 10 až 15 let;

5. Studie 1: Korpus a jeho obsah

údaj o věku je uveden u 3 142 kazuistik; 15 záznamů s nulovým věkem bylo vyřazeno jako chybějící údaj). Rozložení věku (Obrázek 5.3) se soustřeďuje do let typických pro druhý stupeň základní školy; zastoupen je ale i předškolní věk a věk typický pro střední školu. Typ školy kazuistiky samostatně nezaznamenávají. Šíře věkového rozpětí má význam pro obě linie knihy. Pedagogicky umožňuje číst tytéž kategorie chování napříč věkem; nepozornost prvňáka a nepozornost deváťáka jsou však různé jevy pod jedním štítkem, což je třeba mít na paměti při čtení frekvencí. Metodologicky to znamená, že texty, na nichž jsou jazykové modely v dalších studiích prověřovány, pocházejí z různých věkových skupin. Jednotlivé studie však pracují s vlastními výběry (Studie 3 například jen s kazuistikami, které platforma v době výběru vedla v kategoriích fyzické a verbální agrese), takže věkové rozpětí celého řezu se do nich automaticky nepřenáší.



Obrázek 5.3. Rozložení věku žáků v kazuistikách (počet kazuistik podle uvedeného věku; graf zobrazuje věk do dvaceti let, pět kazuistik se staršími žáky v něm není). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

5.3 Co kazuistiky zachycují: chování, řešení, dopady

Obsah kazuistik popisujeme podle první anotační etapy na reprodukovatelném řezu 1 359 ručně kódovaných kazuistik (vymezení řezu uvádí kapitola 4). Řez pochází z tabulky kor-

pusu ze srpna 2024 a obsahuje i kazuistiky, které ve zveřejněném řezu z května 2026 nejsou. Obsahový popis se proto netýká jen dnes zveřejněných kazuistik.

Slovem učitel v této kapitole označujeme toho, kdo situaci v kazuistice řešil. Pisateli jsou převážně studující učitelství, kteří zaznamenávají situaci, kterou na praxi zažili sami nebo kterou řešil jejich provázející učitel (kapitola 3).

Kódování je vícenásobné: jedné kazuistice mohla anotátorka přiřadit více kategorií. U typů chování připadá na jednu kazuistiku v průměru 1,84 kategorie, u typů řešení 2,29, zatímco dopad je téměř vždy jediný (poměr 1,01). Část kazuistik zachycuje jediný projev, mnoho jich však kombinuje více projevů (například verbální agrese s narušováním výuky); u řešení jsou kombinace ještě častější (například rozhovor se žákem spolu se spoluprací s odborníky). Frekvence proto uvádíme jako počet kazuistik, v nichž se kategorie objevila, a nelze je počítat do sta procent. Vícenásobné kódování musí zohledňovat i míry shody (Studie 2) a souběhy řešení s typy chování v oddílu 5.4 počítáme na úrovni kazuistik.

Většinu kazuistik řezu kódovala jediná anotátorka. Shoda mezi anotátorkami byla u typu chování pod nižším z konvenčních prahů a u typu řešení ještě nižší (kapitola 6). Frekvence kategorií a rozdíly mezi nimi proto čteme jako hrubý obraz, ne jako přesné měření.

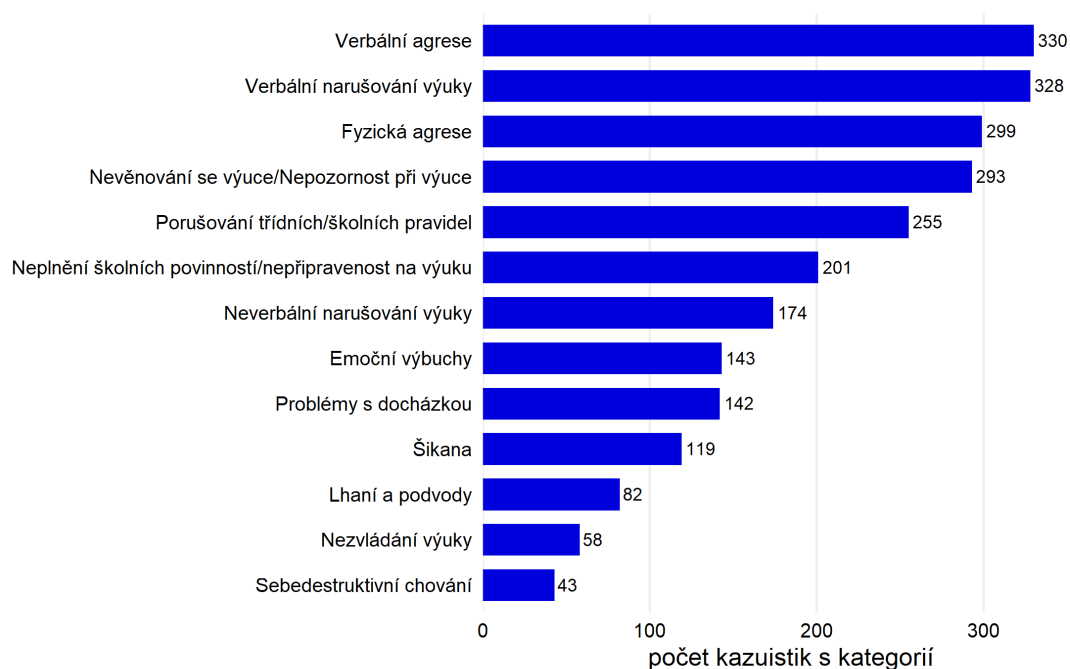
Nejčastěji zachycenými projevy jsou verbální agrese (v 330 kazuistikách), verbální narušování výuky (328), fyzická agrese (299), nepozornost a nevěnování se výuce (293) a porušování třídních či školních pravidel (255). Úplné rozdělení všech třinácti kategorií ukazuje Obrázek 5.4. Rozdělení klesá pozvolna a žádný projev v něm nedominuje; nezanedbatelné zastoupení mají i závažné a organizačně náročné kategorie jako šikana, emoční výbuchy nebo problémy s docházkou. Korpus tedy nezachycuje jen běžnou nekázeň, ale i závažnější situace, které kódovací schéma rozlišuje.

Kódovací strom první etapy nabízel čtrnáct kategorií. Čtrnáctou byla „Diagnóza“, definovaná jako „Žák/žákyně má diagnostikovanou některou z diagnóz SVP“ (tedy speciálních vzdělávacích potřeb). Definice nepopisuje chování, ale stav žáka, takže do dimenze typů náročného chování nepatří. Kategorii proto z deskriptivy vyřazujeme a diagnóza vstupuje do

5. Studie 1: Korpus a jeho obsah

analýzy jako kontextová proměnná (oddíl 5.4). Většinou se navíc objevovala jen jako doplněk k jinému kódu chování a anotátorky ji užívaly velmi nestejně často.

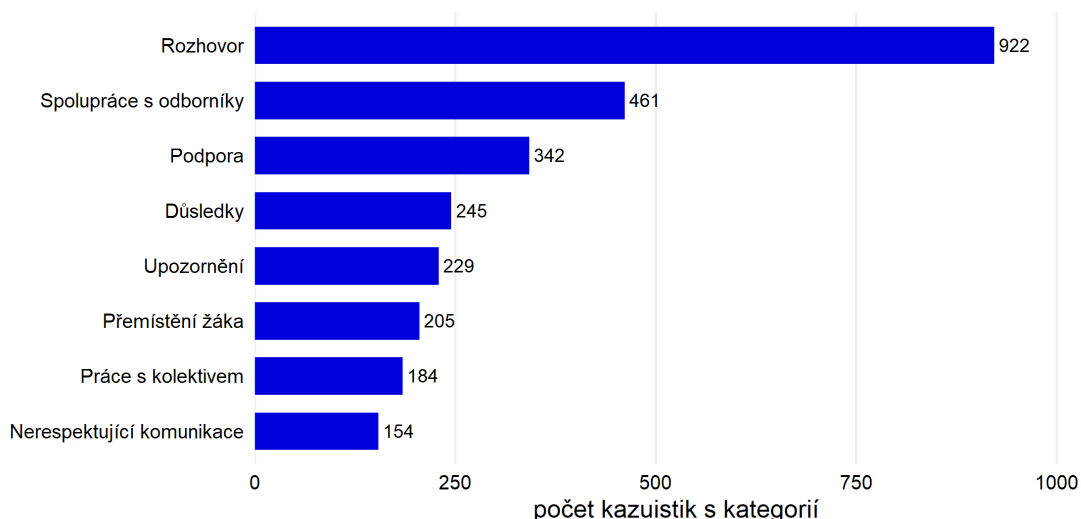
Kategorie neplnění školních povinností (201 kazuistik) ukazuje, že pisatelé za náročné nepovažují jen jednání namířené proti učiteli nebo spolužákům, ale i to, že žák neplní školní práci. Na samém konci rozdělení stojí sebedestruktivní chování. Objevuje se jen v desítkách kazuistik, ale právě u něj může mít chybné řešení nejzávažnější následky; proto je důležitým testovacím případem pro jakýkoli budoucí podpůrný nástroj. Frekvence přitom odrážejí to, co pisatelé považovali za hodné zaznamenání, ne výskyt náročného chování v českých školách; u dobrovolně psaných kazuistik lze čekat nadreprezentaci dramatických situací.



Obrázek 5.4. Typy náročného chování v první anotační etapě (počet kazuistik s kategorií; kódování je vícenásobné). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku](#) ↗).

V repertoáru řešení výrazně převažuje rozhovor, zaznamenaný v 922 kazuistikách. Následuje spolupráce s odborníky (461), podpora žáka (342), práce s důsledky (245) a upozornění (229).

5.3 Co kazuistiky zachycují: chování, řešení, dopady



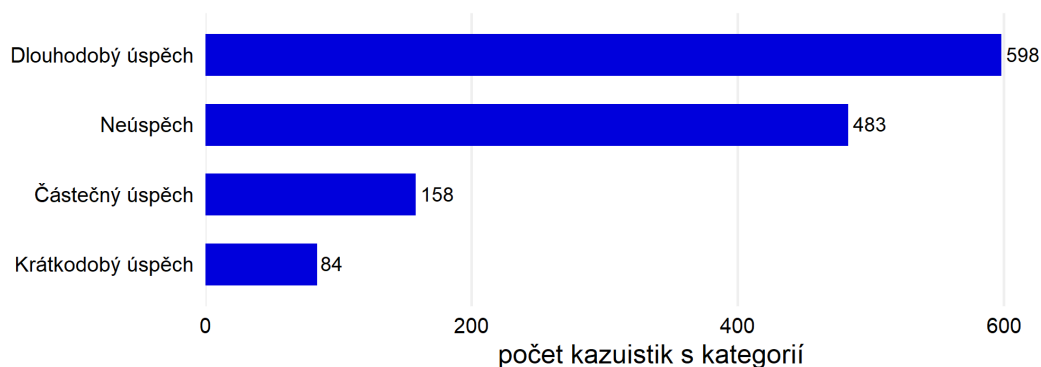
Obrázek 5.5. Nejčastější typy řešení v první etapě (počet kazuistik s kategorií; kódování je vícenásobné, hodnoty viz text). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Skladba kategorií naznačuje převahu postupů, které reagují na již vzniklou situaci. Finální kódovací schéma první etapy (V2) však samostatnou proaktivní kategorii nemá a reakci na incident zvýrazňuje i samotný žánr kazuistiky, tedy popis incidentu a jeho řešení. Ve Studii 3, která pracuje s jiným výběrem i jiným nástrojem, hodnotily anotátorky učitelská řešení v průměru jako mírně reaktivní (kapitola 7). Vysoko v pořadí stojí i spolupráce s odborníky, tedy krok, který překračuje hranici třídy a zapojuje poradenský systém školy. Tuto skladbu repertoáru komentuje diskuse v kapitole 9 pojmem „repertoár nejmenšího odporu“; ve Studii 3 slouží jako srovnávací pozadí pro řešení generovaná umělou inteligencí.

Vnímaný výsledek řešení je nejčastěji dlouhodobý úspěch (598 kazuistik) nebo naopak neúspěch (483), méně často částečný (158) či krátkodobý úspěch (84). Rozdělení je tedy bimodální (Obrázek 5.6), což pravděpodobně souvisí se sběrem podle zrcadlové šablony (kapitola 3): pisatelé odevzdávají úspěšně a neúspěšně řešenou situaci v páru, takže neúspěchy jsou v korpusu zastoupeny výrazněji než ve sbírkách, které zachycují jen úspěšné případy. Částečný a krátkodobý úspěch mezi krajními póly tvoří dohromady menšinu dopadů. I to může souviset se šablonou, která pisatele vede k volbě jednoznačně vydařené a jednoznačně nevy-

5. Studie 1: Korpus a jeho obsah

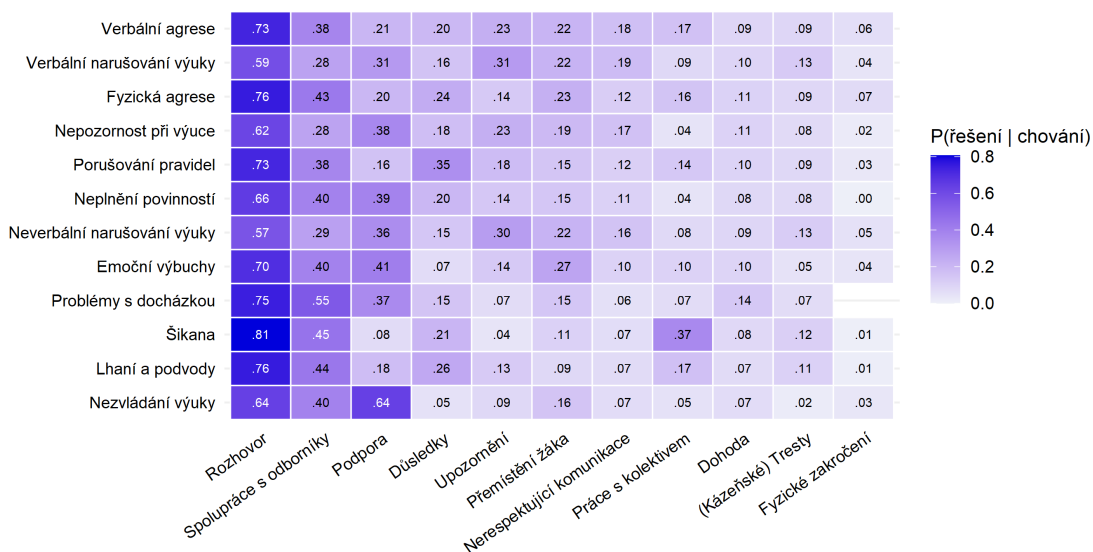
dařené situace; reálná praxe může být na částečné úspěchy bohatší, než korpus ukazuje. K této výhradě se vrací diskuse v kapitole 9.



Obrázek 5.6. Dopady řešení v první anotační etapě (počet kazuistik s kategorií). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

5.4 Jak se řešení pojí s chováním

Frekvence kategorií ukazují, jaká řešení kazuistiky zaznamenávají, ne však, v jakých situacích. Obrázek 5.7 proto zachycuje podmíněné podíly: pro každý typ chování podíl kazuistik s tímto chováním, v nichž se objevilo dané řešení. Jednotkou je kazuistika (kódování je vícenásobné, kapitola 4). Do analýzy vstupují typy chování s dostatečným počtem kazuistik (sebedestruktivní chování proto chybí) a podíly se čtou po řádcích. Jde o popis souběhů v kazuistikách, ne o doporučení ani kauzální vztah.



Obrázek 5.7. Podmíněné podíly řešení podle typu chování, $P(\text{řešení} \mid \text{chování})$: podíl kazuistik s daným chováním, v nichž se objevilo dané řešení. Řádky = typy chování (seřazeny podle četnosti; zobrazeny jen typy chování s dostatečným počtem kazuistik, sebedestruktivní chování proto chybí), sloupce = typy řešení; kódování je vícenásobné, řádkové podíly se proto nesčítají do 1. *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Obrázek ukazuje především rozšířenost rozhovoru: u všech typů chování je nejčastějším řešením, nebo se o první místo dělí s jiným, a nejvyšší podíl má u šikany, kde se objevuje v 81 % kazuistik. Jde o četnost zmínek, ne o pořadí kroků; analýza neurčuje, který krok byl první, druhý nebo třetí. Zastoupení dalších postupů se mezi typy situací liší. U šikany je nadprůměrně častá práce s kolektivem (37 % kazuistik se šikanou oproti výrazně nižším podílům, nejvýše kolem jedné šestiny, u ostatních typů chování), což odpovídá skupinové povaze jevu. U porušování pravidel je nadprůměrně častá práce s důsledky (35 %) a tam, kde pisatel popisuje nezvládnání výuky, se vedle rozhovoru rovným dílem objevuje podpora (64 %).

Rozdíly dobře ukazuje srovnání dvou častých typů chování. U verbálního narušování výuky, typického drobného incidentu, se vedle rozhovoru (59 %) objevují se srovnatelnou četností podpora žáka (31 %) a upozornění (31 %), tedy vedle individuální podpory i okamžitá korekce v běhu hodiny. U neplnění školních povinností, které má pomalejší dynamiku, okamžitá

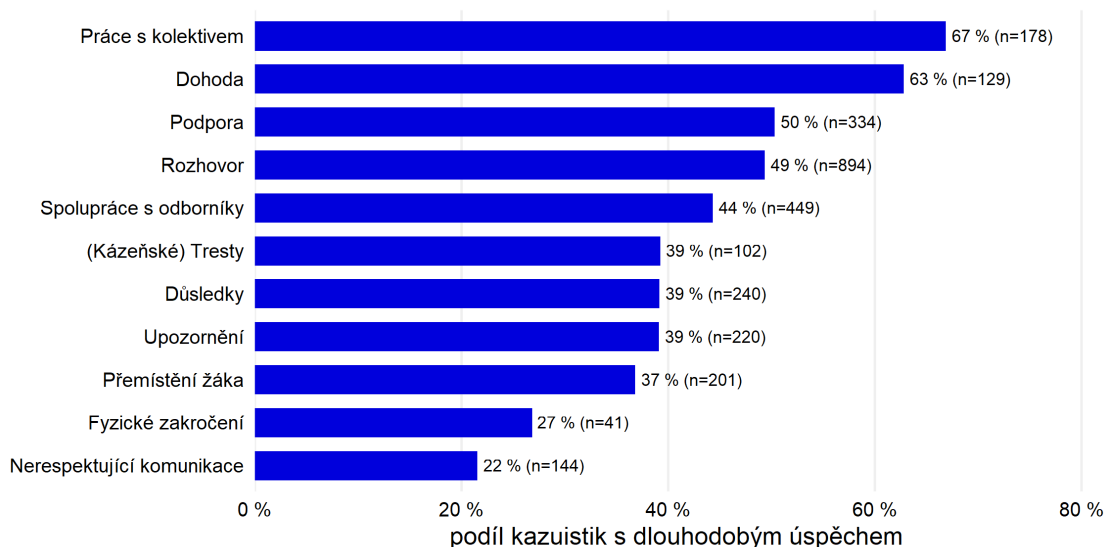
5. Studie 1: Korpus a jeho obsah

korekce ustupuje: upozornění klesá na 14 %, zatímco vedle rozhovoru (66 %) stojí spolupráce s odborníky (40 %) a podpora žáka (39 %). Kde situace nevyžaduje zásah v běhu hodiny, je tedy okamžitá korekce méně častá, kdežto individuální podpora a spolupráce s odborníky zůstávají časté. Repertoár se podle situace liší hlavně v tom, která řešení provázejí nejčastěji zaznamenaný rozhovor.

Samostatnou otázkou je, zda se repertoár mění tam, kde má žák zaznamenanou diagnózu. Protože diagnóza není typem chování, čteme ji z kurátorovaného pole korpusu, ne z anotace: vyplněné je u 530 kazuistik řezu, tedy u 39 %. Nejzřetelnější posun se týká podpory žáka, která se v těchto kazuistikách objevuje ve 37 % oproti 25 % v celém řezu; opačným směrem se posouvají důsledky (13 % oproti 18 %) a nerespektující komunikace (9 % oproti 11 %). Spolupráce s odborníky, tedy mezioborový krok, se v kazuistikách s diagnózou objevuje jen o málo častěji (36 % oproti 34 %). Srovnáváme přitom s celým řezem, jehož součástí jsou i kazuistiky s diagnózou. I zde jde o popis souběhů: kazuistiky s diagnózou se liší i skladbou situací, takže rozdíl v podílech nelze číst jako účinek diagnózy samotné.

5.5 Dopad podle typu řešení

Poslední deskriptivní pohled spojuje řešení s vnímaným výsledkem. Obrázek 5.8 ukazuje pro každý typ řešení podíl kazuistik s tímto řešením, které skončily dlouhodobým úspěchem. Kód neúspěchu z první etapy přitom zahrnuje nejen nezměněné chování, ale také normativní odmítnutí nevhodného prostředku (příloha A.1); ze samotného kódu tyto důvody oddělit nelze. Typ řešení si navíc učitelé nevybírají náhodně, ale podle situace, takže rozdíly mezi řešeními odrážejí i rozdílnou skladbu situací, v nichž se používají. Podíly proto popisují vztah anotačních kategorií, ne účinnost strategií, a nelze z nich vyvozovat, že by některé řešení úspěch způsobovalo.



Obrázek 5.8. Podíl kazuistik s dlouhodobým úspěchem podle typu řešení (deskriptivně; n = počet kazuistik s daným řešením a uvedeným dopadem). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Podíly se mezi řešeními liší. Nejvyšší podíl dlouhodobého úspěchu mají kazuistiky s prací s kolektivem (67 % ze 178 kazuistik), s dohodou (63 %) a s podporou žáka (50 %); rozhovor, nejčastěji zaznamenané řešení, leží uprostřed (49 %). Nejnižší podíly mají kázeňské tresty (39 %), přemístění žáka (37 %), fyzické zakročení (27 %) a nerespektující komunikace (22 %); podobný podíl jako tresty mají i důsledky a upozornění. U fyzického zakročení a nerespektující komunikace je nízký podíl zčásti dán už definicí kódu neúspěchu, který nevhodné prostředky řadí mezi neúspěchy (příloha A.1). Kód typu řešení navíc anotátorky přidělovaly s nízkou vzájemnou shodou (kapitola 6). Vztahová a kolektivní řešení se tedy v korpusu vyskytují spolu s příznivěji kódovanými výsledky než řešení represivní; nejde však o měření účinnosti.

Vzorec mohou spoluutvářet i mechanismy, které by kauzální čtení muselo vyloučit. Jedním je selekce situací: k práci s kolektivem nebo k dohodě učitel sahá typicky tam, kde situace ještě dohodu připouští, zatímco fyzické zakročení je z povahy věci vyhrazeno situacím eskalovaným, jejichž vyhlídky jsou horší bez ohledu na zvolený zásah. Dalším je pohled pisatele: výsledek popisuje pisatel, který situaci sám řešil, nebo ji na praxi sledoval u provázejícího

5. Studie 1: Korpus a jeho obsah

učitele, a vnímaná úspěšnost může být přikrášlena u řešení, s nimiž se ztotožňuje. Svou roli hraje i samotné kódování: jak bylo uvedeno výše, část vzorce vzniká už definicí kódu neúspěchu. Vliv pohledu pisatele částečně oslabuje zrcadlová šablona (kapitola 3), protože do korpusu aktivně přivádí i neúspěchy; selekci situací ani vliv kódování však sběr tohoto typu neodstraní. Vzorec proto předkládáme jako popis souběhů; silnější čtení by vyžadovalo uspořádání studie, které umožní srovnávat obdobné situace.

5.6 Shrnutí

Studie 1 popsala korpus jako rostoucí databázi, z níž se část kazuistik časem také stahuje; aktuální zveřejněný řez z května 2026 čítá 3 202 kazuistik. Kazuistiky jsou souvislé texty střední délky se čtyřdílnou strukturou, v níž je popis výsledku v mediánu nejkratší; věk žáků se soustřeďuje do let typických pro druhý stupeň základní školy. Obsah popisuje řez první anotační etapy ze srpna 2024: převažují v něm verbálně a fyzicky agresivní projevy a projevy narušující výuku a mezi řešeními rozhovor. Podmíněné podíly ukázaly, že se skladba řešení podle situace liší hlavně v tom, která řešení rozhovor provázejí. Vztahová a kolektivní řešení se v kazuistikách pojí s příznivěji kódovanými výsledky než řešení represivní; kód neúspěchu však zahrnuje i normativní posouzení prostředku, takže nejde o měření účinnosti. Na tento základ navazují otázky dalších studií: nakolik spolehlivě lze takový obsah kódovat jazykovým modelem (Studie 2), jak si učitelská řešení stojí ve srovnání s generovanými (Studie 3) a jak spolehlivě lze posoudit kvalitu kazuistik a dopad řešení (Studie 4).

6. Studie 2: Jazykový model jako anotátor

Druhá studie zjišťuje, nakolik se velký jazykový model při kódování obsahu kazuistik shoduje s vyškolenými lidskými anotátorkami (VO2). Omezuje se na dimenzi typu řešení, tedy na nejjemněji členěnou část schématu; kódování chování a dopadu modelem neověřuje. Kapitola postupuje od kódovacího schématu a shody mezi lidmi (6.2, 6.3) ke shodě modelu s lidmi souhrnně, po kategoriích i v jednotlivých záměnách (6.4) a končí kvalitativním pohledem na vybranou neshodu (6.5).

6.1 Východiska

Kapitola 5 ukázala, co kazuistiky obsahují, na řezu, který bylo možné ručně okódotovat. Ruční kódování je však úzké hrdlo: rostoucí korpus (kapitola 3) brzy přeroste kapacitu lidských anotátorek. Nabízí se proto použít jako anotátora jazykový model, který tímž kódovacím schématem zpracuje libovolně velký objem textu. To ovšem předpokládá, že se model s lidmi shoduje. Studie 2 proto srovnává lidské a modelové kódování na sdíleném vzorku kazuistik, a to u dimenze typu řešení. Její kategorie mají nejjemnější a nejvíce se překrývající hranice a shoda mezi lidmi je u ní, jak ukáže oddíl 6.3, nejvyšší. Výsledek proto nelze bez dalšího přenést na kódování chování a dopadu; to by vyžadovalo samostatné ověření. Na typech řešení přitom stojí ústřední pedagogický výklad knihy, repertoáry řešení, a právě u této dimenze je proto ověření shody nejdůležitější.

6.2 Metoda

Kódovací schéma a lidské kódování

Ve finálním kole první anotační etapy (kapitola 3) kódovalo pět vyškolených anotátorek kazuistiky do tří dimenzí: typ náročného chování, typ řešení a dopad řešení. Schéma vzniklo induktivně, otevřeným kódováním části korpusu, a bylo zpřesněno do kódovacího stromu (příloha A). Finální anotování probíhalo jako dělba práce: každá anotátorka kódovala jiné kazuistiky, aby se pokryl co největší rozsah korpusu, a vestavěný překryv sloužil ke kontrole shody. Spolehlivost proto počítáme na tomto překryvu; deskriptivní četnosti uvádí Studie 1. U typu chování vstupuje do výpočtu třináct behaviorálních kategorií. Kontextová položka původního stromu, která pouze zaznamenávala přítomnost diagnózy žáka, je stejně jako v deskriptivě Studie 1 vyřazena (oddíl 5.3 a příloha A).

Modelové kódování a srovnávací vzorek

Na srovnávacím vzorku kazuistik klasifikoval jazykový model gpt-4o (teplota nula) typy řešení do kategorií kódovací knihy, k nimž instrukce přidávala kategorii proaktivního řešení. Instrukce pro model (anglicky prompt) obsahovala vlastní, zkrácené definice kategorií a k nim doslovné příklady (příloha B.3). Definice se od kódovací knihy v některých kategoriích liší, nejvíce u podpory žáka. Model dostal pouze text řešení, anotátorky četly celou kazuistiku.

Srovnávací vzorek čítá 334 případů. Do vzorku byly zařazeny kazuistiky, u nichž lidské kódování určilo jediné dominantní řešení. Lidským kódem je kód z anotační matice první etapy, u většiny kazuistik tedy kód jediné anotátorky. U jednoho případu lidský kód chybí, analýza proto zahrnuje 333 úplných dvojic lidského a modelového kódu.

Srovnání pracuje s 12 kategoriemi řešení: jedenácti kategoriemi kódovací knihy první etapy a kategorií proaktivního řešení, kterou model dostal v instrukci ke klasifikaci (znění v příloze B.3, původ kategorie v příloze A).

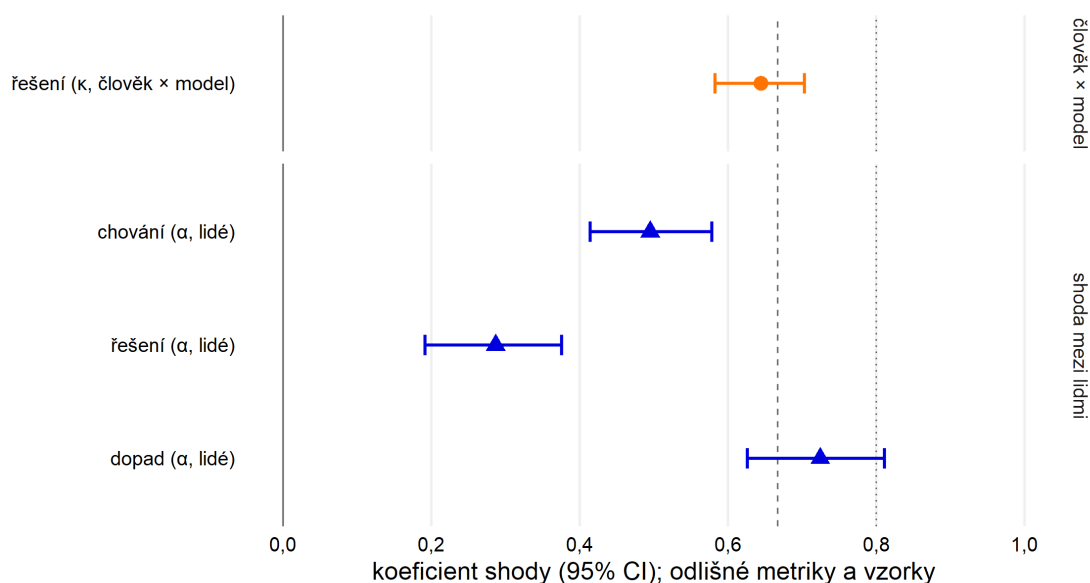
Míry shody

Shodu mezi lidmi měříme nominálním Krippendorffovým α na jednotkách kazuistika $\times$ anotátorka, a to na první kategorii, kterou anotátorka u kazuistiky uvedla. Protože schéma připouští více kategorií, doplňkově užíváme průměrný párový Jaccardův index přes všechny přiřazené kategorie. Shodu lidí a modelu vyjadřujeme přesnou shodou (hlavní kategorie modelu je totožná s kategorií řešení určenou v lidském kódování), rozšířenou shodou (lidská kategorie je mezi kategoriemi modelu) a Cohenovým κ . U α , přesné shody a κ uvádíme 95% intervaly spolehlivosti z bootstrapu přes kazuistiky (tisíc opakování).

6.3 Shoda mezi lidskými anotátorkami

Shoda pěti anotátorek na překryvu se mezi dimenzemi liší. U dopadu dosáhla $\alpha = 0,72$ (95% CI 0,63 až 0,81, na 85 společně kódovaných kazuistikách), tedy hodnoty mezi nižším a vyšším konvenčním prahem vyznačeným v Obrázku 6.1; dimenze dopadu má jen několik zřetelně odlišných kategorií. Anotátorky navíc při kódování dopadu viděly pisatelovo označení kazuistiky jako úspěšně, nebo neúspěšně řešené (příloha A.1), což mohlo shodu v této dimenzi zvýšit. U typu chování byla shoda pod nižším z prahů, $\alpha = 0,50$ (0,41 až 0,58), a u typu řešení nejnižší, $\alpha = 0,29$ (0,19 až 0,38). Obrázek 6.1 shrnuje tyto hodnoty spolu s hlavní mírou následujícího oddílu. Shodu modelu s lidmi je třeba číst na pozadí toho, jak se shodují lidé mezi sebou.

6. Studie 2: Jazykový model jako anotátor



Obrázek 6.1. Reliabilita kódování: Krippendorffovo α shody mezi anotátorkami po dimenzích a Cohenovo κ shody člověk $\times$ model u typu řešení (95% intervaly spolehlivosti z bootstrapu; svislice vyznačují konvenční prahy 0,667 a 0,80 pro Krippendorffovo α , u κ slouží jen k orientaci). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku](#)). Obě míry pocházejí z odlišného uspořádání (pět anotátorek na překryvu vs. lidský kód z anotační matice $\times$ model na srovnávacím vzorku) a figura je řadí vedle sebe pro orientaci, ne jako přímo srovnatelné hodnoty (oddíl 6.4).

Nižší shodu u typu řešení lze přičíst hlavně jemnosti a překryvnosti jeho kategorií. Totéž řešení lze oprávněně zařadit pod více blízkých kategorií, a protože α trestá každou neshodu stejně, jeho hodnota klesá. Přispívat může i nerovnoměrné rozdělení kategorií s převahou rozhovoru, které hodnotu α rovněž snižuje. Tomu odpovídá i průměrný párový Jaccardův index, který zohledňuje vícenásobné kódování a započítává částečný překryv: u řešení dosahuje 0,56, u chování 0,61 a u dopadu 0,86. Tato zjištění se týkají následujícího srovnání s modelem i celé knihy. Četnosti ze Studie 1 proto uvádíme na úrovni kategorií, ne jako přesné klasifikace jednotlivých kazuistik, a hlubší psychometrický výklad jednotlivých kódů typu řešení by přesahoval spolehlivost, kterou data poskytují.

6.4 Shoda jazykového modelu s lidmi

Na srovnávacím vzorku se hlavní kategorie modelu shodovala s kategorií řešení určenou v lidském kódování v 72 % případů (95% CI 68 % až 77 %). Připustíme-li rozšířenou shodu, tedy případy, kdy je lidská kategorie mezi kategoriemi přiřazenými modelem, stoupá shoda na 93 %. Cohenovo κ , které koriguje shodu očekávanou náhodou, dosahuje hodnoty 0,64 (95% CI 0,58 až 0,70), tedy těsně pod nižším z prahů, které Obrázek 6.1 vyznačuje pro orientaci. Nejde však o nezávislé ověření: srovnávací vzorek nebyl oddělen od příkladů v instrukci modelu. Po vyloučení případů, jejichž řešení se mezi příklady objevila, činí $\kappa = 0,61$ (podrobněji níže).

Z těchto výsledků nelze usuzovat, že model dosahuje lidské úrovně nebo ji překračuje. Nominální α pěti anotátovek na malém překryvu a κ mezi lidským kódem z anotační matice první etapy (u většiny kazuistik jde o kód jediné anotátorky) a modelem na jiném vzorku neodhadují tutéž veličinu, takže z jejich rozdílu nelze stanovit lidský strop. Neshody mohou souviset s nejednoznačností schématu i s chybami modelu; tento design jejich podíly neodděluje. U vzácných kategorií navíc zůstává odhad nejistý. Výsledek podporuje použití modelových kategorií jen pod lidskou kontrolou; jejich zobecnění by vyžadovalo nezávislé ověření.

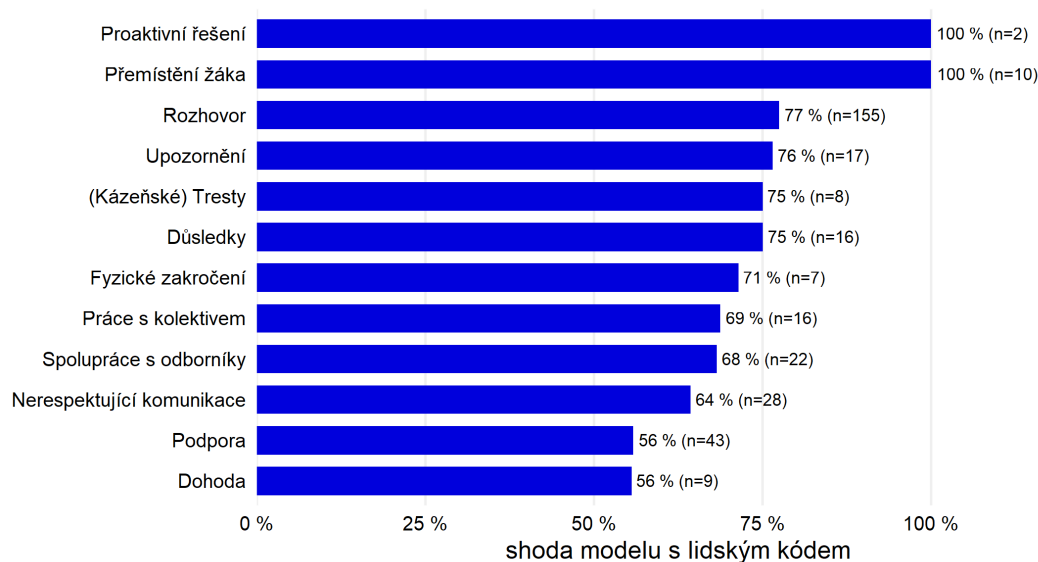
Překryv s příklady v instrukci

Mezi příklady v instrukci modelu jsou celé texty řešení 36 případů ze srovnávacího vzorku. Překryv jsme zjišťovali vyhledáním úplného znění řešení v části instrukce s příklady po sjednocení zápisu znaků, velikosti písmen a mezer; parafráze tak zachyceny nejsou.

Citlivostní analýza vylučuje tyto případy i případ bez lidského kódu. Na 297 případech činí přesná shoda 71 % a $\kappa = 0,61$ (95% CI 0,55 až 0,68). Úplné výsledky citlivostní analýzy jsou v doprovodném repozitáři (Nehyba, 2026). Vyloučení těchto případů neodstraňuje jiné možné vazby mezi vývojem instrukce a srovnávacím vzorkem; silnější tvrzení by vyžadovalo nový vzorek, který se při tvorbě instrukce nepoužil.

Shoda po kategoriích

Souhrnné κ nezachycuje, že se shoda modelu s lidským kódem mezi kategoriemi liší. Obrázek 6.2 proto rozkládá přesnou shodu podle lidské kategorie řešení. U zdaleka nejčastější kategorie, rozhovoru, se model shodl s lidmi v 77 % případů ($n = 155$); podobně vysokou shodu, i když na malém počtu případů, má i upozornění (76 %, $n = 17$) a důsledky (75 %, $n = 16$). Mezi kategoriemi s více než hrstkou případů má nejnižší shodu podpora žáka (56 %, $n = 43$), tedy kategorie, kterou kódovací kniha a instrukce modelu vymezují odlišně. Prakticky stejnou nízkou shodu má dohoda, u níž však jde jen o 9 případů. U kategorií s jednotkami případů uvádíme čísla jen pro úplnost; usuzovat z nich nelze.



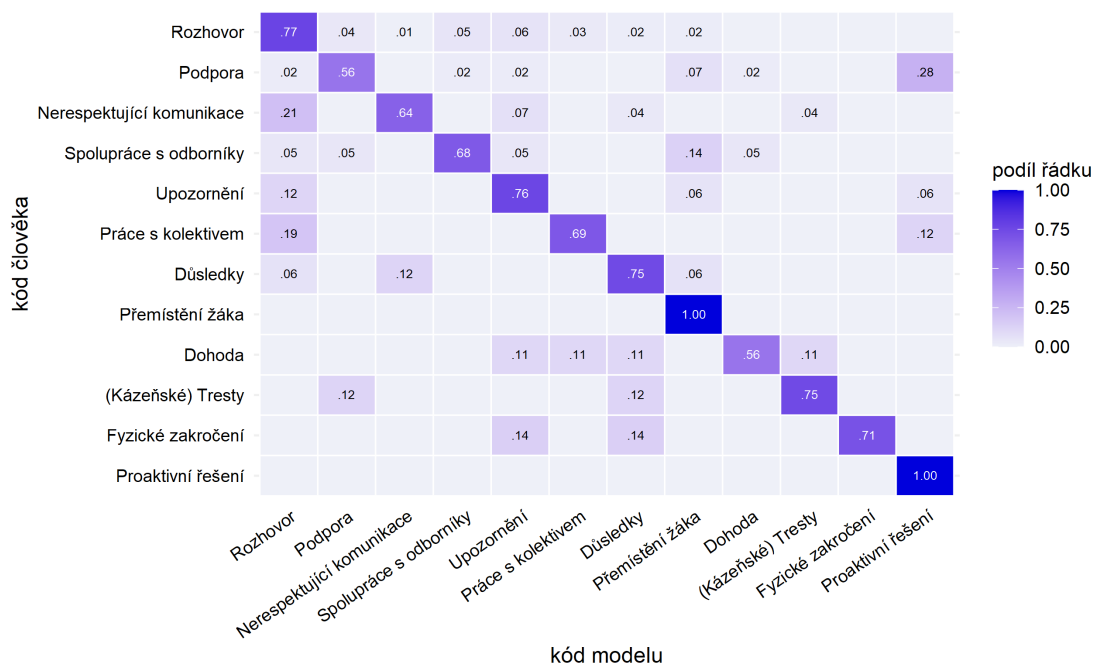
Obrázek 6.2. Přesná shoda modelu s lidským kódem po kategoriích řešení (n = počet kauzistik s danou lidskou kategorií). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Záměny mezi kategoriemi

Matice záměn (konfuzní matice, Obrázek 6.3) ukazuje, kam neshody směřují. Více než čtvrtina případů, které lidé kódovali jako podporu žáka, dostala od modelu kód proaktivního řešení. Tuto kategorii však měl v nabídce model, kdežto v lidském kódování první etapy ji

6.5 Kvalitativní pohled na vybranou neshodu

měla k dispozici jen jedna anotátorka (příloha A). Ve srovnání proto vystupuje téměř jen jako kód přiřazený modelem a část této záměny vyplývá z rozdílné nabídky kategorií, ne jen z jejich významové blízkosti. Pětinu případů, které lidé kódovali jako nerespektující komunikaci, zařadil model jako rozhovor; toto jednání může mít v kazuistikách podobu slovní výměny se žákem. Záměny tedy zčásti sledují významové hranice mezi kategoriemi. Zda jde o tytéž hranice, které snižovaly shodu mezi lidmi, z analýzy v oddílu 6.3 nevyplývá.



Obrázek 6.3. Matice záměn lidského a modelového kódu řešení (řádkově normalizované podíly; řádky = lidský kód, sloupce = kód modelu; prázdné buňky = podíl menší než 0,005). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku](#) ↗).

6.5 Kvalitativní pohled na vybranou neshodu

Ilustrujeme jeden případ; nejde o systematický rozbor příčin neshod.

„Vyučování pokračovalo a nic se neřešilo. Ten den se mi ozvala matka žáka, že takovéto jednání se jí nelíbí. Já jsem samozřejmě o té situaci neměla žádné bližší

6. Studie 2: Jazykový model jako anotátor

informace, tak jsem jako třídní napsala, že o tom bohužel nic nevím a ať se obrátí na tu konkrétní paní učitelku.“ (člověk: nerespektující komunikace; model: rozhovor, jako další kategorie nerespektující komunikace)

Případ ukazuje rozdíl mezi hodnotícím a doslovným čtením. Anotátorka jednání vyhodnotila jako nerespektující komunikaci, tedy kategorii, která nese normativní úsudek o kvalitě jednání. Model přiřadil jako hlavní kategorii rozhovor, tedy doslovný komunikační akt (výměnu zpráv), a nerespektující komunikaci uvedl až jako další kategorii. Podle přesné shody jde o neshodu, podle rozšířené shody o shodu. Hlavní kód modelu tu popisuje, co se stalo, kód anotátorky hodnotí, jak se to stalo. Souhrnné výsledky po kategoriích však neukazují, že by hodnotící kategorie byly pro model obecně nejtěžší.

Pro praxi strojového kódování to naznačuje, že nasazení modelu na velký korpus by mělo být doprovázeno lidskou kontrolou, zejména u hodnotících a významově širokých kategorií. Paušální míra shody takové rozlišení neumožňuje, rozklad po kategoriích ano. Ani vyšší shoda u některých kategorií však neznamená, že jsou modelové kódy správné.

6.6 Shrnutí

Studie 2 ukázala, že u typu řešení se hlavní kategorie přiřazená jazykovým modelem shodovala s lidským kódem v 72 % případů (Cohenovo $\kappa = 0,64$; rozšířená shoda 93 %). Nejde o nezávislé ověření, protože řešení části srovnávacích případů byla mezi příklady v instrukci modelu. Po jejich vyloučení činí $\kappa = 0,61$; ani tato dodatečná analýza není novým nezávislým testem. Lidským kódem je kód z anotační matice první etapy, kterou anotátorky pořizovaly dělbou práce, a výsledek platí jen pro dimenzi typu řešení. Lidské α z jiného uspořádání nelze použít jako horní mez shody, které by model mohl dosáhnout.

Rozklad po kategoriích ukázal, že shoda modelu s lidským kódem je nejvyšší u rozhovoru a u některých méně častých kategorií; mezi kategoriemi s více než hrstkou případů je nejnižší u podpory žáka. U málo četných kategorií jsou odhady nejisté. Matice záměn naznačuje,

že část záměn sleduje významové hranice schématu; záměna podpory a proaktivního řešení souvisí také s tím, že proaktivní řešení měl v nabídce model, kdežto z anotátorek jen jedna. Vybraná neshoda (oddíl 6.5) ilustruje rozdíl mezi hodnotícím a doslovným čtením jednání; o podílu chyb způsobených schématem nevypovídá. Model proto může sloužit k rozšíření obsahové analýzy pod lidskou kontrolou, lidské kódování však nenahrazuje; korpusová deskriptiva této knihy stojí výhradně na lidském kódování. Následující kapitola obrací pozornost od klasifikace řešení k jejich kvalitě a srovnává řešení generovaná modelem s řešeními učitelů.

7. Studie 3: Řešení generovaná umělou inteligencí versus řešení učitelů

Třetí studie odpovídá na otázku, jak se řešení náročných situací generovaná velkým jazykovým modelem liší od autentických řešení, která zvolili učitelé (VO3). Využívá druhou anotální etapu: zaslepené hodnocení dvojic řešení šesti vyškolenými anotátorkami na čtyřech pedagogických dimenzích a v úloze rozpoznávání pojmenovaných rámců řešení náročného chování. Kapitola popisuje materiál a uspořádání studie (7.2), uvádí spolehlivost hodnocení (7.3) a předkládá výsledky smíšených modelů (7.4).

7.1 Východiska

Kapitola 2 vymezila generování řešení jako jednu ze tří rolí, v nichž jazykový model v knize zkoumáme. Otevřenou otázkou zůstává, zda generované návrhy obstojí při hodnocení vhodnosti pro danou situaci a jak se od učitelských liší poměrem prevence a reakce, hodnotovou orientací a systémovostí zásahu. Přímé srovnání na autentických případech naráží na zjevný metodologický problém: kdyby hodnotitelé věděli, které řešení vytvořil stroj, jejich očekávání by hodnocení ovlivnila. Druhá etapa proto byla navržena jako zaslepená: anotátorky hodnotily obě řešení téže kazuistiky, aniž jim bylo sděleno, že jedno z nich je generované. Zda původ některých řešení odhadly, studie neověřovala.

7.2 Metoda

Materiál

Z korpusu (kapitola 5) bylo vybráno 324 kazuistik, které platforma v době výběru vedla v kategoriích fyzické a verbální agrese, s úplným popisem situace i učitelského řešení. Učitelským řešením rozumíme postup vyučujícího v situaci (učitele, nebo studujícího na praxi), jak jej v kazuistice zpětně zaznamenal pisatel. Ke každé kazuistice vygeneroval model gpt-4o alternativní řešení. Instrukce mu ukládala roli odborníka na náročné chování a řízení třídy a žádala co nejlepší postup pro popsanou situaci; učitelské řešení naproti tomu popisuje postup, který byl ve třídě skutečně použit. Instrukce dále požadovala přibližně stejnou délku jako u učitelského řešení, zachování gramatického rodu a minulý čas, aby povrchové rysy textu neprozrazovaly jeho původ. Šlo o jediné srovnávací řešení vytvořené jednotnou instrukcí (promptem); nejde o žádné ze šesti rámcovaných řešení, která nabízí platforma (kapitola 3). Znění instrukce uvádí příloha B (oddíl B.4).

Hodnocení

Každou dvojici řešení mělo hodnotit všech 6 anotátorek, a to nezávisle, zaslepeně a ve znárodném pořadí (krycí instrukce se o AI nezmiňovala). Každá anotátorka tak četla k téže situaci obě řešení, což zaslepení oslabuje. Každé řešení posuzovaly na čtyřech ordinálních pětibodových dimenzích: celková vhodnost pro danou situaci, orientace reaktivní versus proaktivní, orientace humanistická versus behaviorální a zaměření systémové versus situační. V analyzovaném souboru jsou všechny čtyři uloženy na škále od -2 do $+2$ (vhodnost se v kódovacím formuláři zadávala jako pětibodová a pro analýzu byla lineárně překódována). Vedle toho anotátorky označovaly, zda řešení rozpoznatelně odráží některý z pojmenovaných rámců (sociálně-emoční učení, pozitivní behaviorální podpora, nenásilná komunikace, restorativní praxe). Závěk proběhl ve dvou kalibračních kolech s rozbořem neshod (příloha C, Tabulka C.1).

Soubor hodnocení obsahuje 15 476 hodnoticích polí (kombinace kazuistiky, řešení, dimenze a anotátorky), z nichž 1 600 (přibližně desetina) zůstalo nevyplněno: 11 % polí u řešení generovaných a 10 % u učitelských. Chybění nebylo mezi anotátorky rozloženo rovnoměrně. Úplné uspořádání by mělo 15 552 polí; dalších 76 polí v datech zcela chybí, všechna u učitelských řešení. Celkem tedy chybí 1 676 hodnocení. Analýzy pracují se všemi vyplněnými odpověďmi. Zaslepení samo nezaručuje, že chybění je náhodné: vynechání mohlo souviset s vlastnostmi textu nebo s nejistotou anotátorky. Rozpis chybění podle zdroje, dimenze a anotátorky obsahuje doprovodný repozitář (Nehyba, 2026).

Analýza

Primárním modelem je lineární smíšený model (LMM) s pevným efektem zdroje řešení (referenční kategorie AI) a náhodnými průsečíky pro kazuistiku a anotátorku. Model tak zohledňuje opakovaná hodnocení týchž případů i systematické rozdíly v přísnosti hodnotitelů. Koeficient zdroje lze číst přímo v bodech škály; model však zachází s pětibodovou škálou, jako by vzdálenosti mezi sousedními stupni byly stejné.

Citlivostní analýzou je proto kumulativní logitový smíšený model (CLMM), který tento předpoklad nepotřebuje. Hodnocení v něm vystupuje jako uspořádaná kategoriální proměnná a model popisuje kumulativní pravděpodobnosti, tedy pravděpodobnost, že hodnocení nepřekročí daný stupeň škály, na čtyřech prazích mezi pěti stupni. Logit každé kumulativní pravděpodobnosti je lineární funkcí týchž členů jako v primárním modelu: pevného efektu zdroje řešení a náhodných průsečíků pro kazuistiku a anotátorku. Model předpokládá proporcionalitu šancí, tedy že jediný efekt zdroje posouvá všechny prahy stejně; tento předpoklad jsme samostatně netestovali. Exponenciála koeficientu zdroje je poměr šancí (odds ratio, OR) na hodnocení nad kterýmkoli prahem. OR menší než jedna znamená, že učitelské řešení má při hodnocení téže kazuistiky touž anotátorkou nižší šanci na vyšší stupeň škály než řešení generované, a to na kterékoli hranici mezi stupni; OR větší než jedna znamená šanci vyšší. Model odhadujeme metodou maximální věrohodnosti (balík ordinal, funkce clmm) a uvádíme Waldovy 95% intervaly spolehlivosti. Závěry se opírají o LMM; CLMM slouží k ověření,

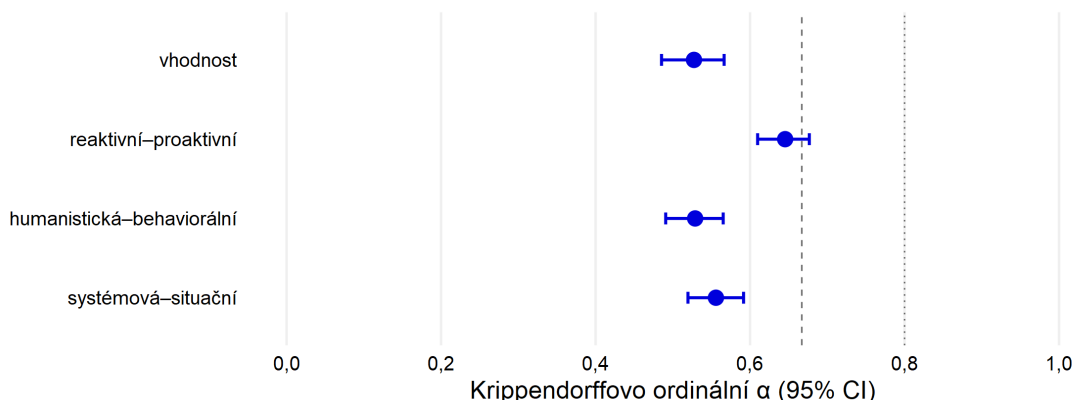
7. Studie 3: Řešení generovaná umělou inteligencí versus řešení učitelů

zda směr a statistická průkaznost efektů nezávisí na předpokladu stejných vzdáleností mezi stupni škály.

Protože uspořádání je párové uvnitř kazuistiky, druhou citlivostní analýzou je LMM s náhodným sklonem zdroje po kazuistikách, který připouští, že se rozdíl mezi zdroji případ od případu liší (oddíl 7.4). Třetí citlivostní analýza připouští, že se rozdíl liší mezi anotátorkami: LMM s náhodným sklonem zdroje po anotátorkách a doplňkově model s interakcí zdroje a anotátorky, z něhož vychází rozdíl učitel minus AI pro každou anotátorku zvlášť. Spolehlivost hodnocení vyjadřujeme Krippendorffovým ordinálním α s 95% intervaly spolehlivosti z opakovaného výběru kazuistik s vrácením (bootstrap). U rámců ji vyjadřujeme průměrným párovým Jaccardovým indexem, tedy podílem rámců, které dvojice anotátorek označila shodně, ze všech rámců, které označila alespoň jedna z nich.

7.3 Spolehlivost hodnocení

Shoda šesti anotátorek (jednotkou je řešení dané kazuistiky) činila $\alpha = 0,53$ (95% CI 0,49 až 0,57) pro vhodnost, $\alpha = 0,65$ (0,61 až 0,68) pro dimenzi reaktivní–proaktivní, $\alpha = 0,53$ (0,49 až 0,57) pro dimenzi humanistickou–behaviorální a $\alpha = 0,56$ (0,52 až 0,59) pro dimenzi systémovou–situační. Obrázek 7.1 hodnoty shrnuje. Všechny čtyři hodnoty leží pod nižším z konvenčních prahů vyznačených v obrázku; jednotlivá hodnocení jsou tedy zatížena značnou neshodou. Nejvyšší shody dosáhla dimenze reaktivní–proaktivní, snad proto, že její póly lze v textu řešení rozpoznat nejsnáze (zásah po incidentu versus práce s podmínkami). Smíšené modely v oddíle 7.4 zohledňují rozdílnou přesnost anotátorek; nízká shoda se v nich projeví širšími intervaly spolehlivosti, systematické zkreslení vázané na jeden zdroj řešení však neodhalí. Spolehlivost uvádíme z finálního anotování; k reliabilitě napříč etapami se vrací kapitola 9.



Obrázek 7.1. Reliabilita hodnocení: Krippendorffovo ordinální α podle dimenzí (95% intervaly spolehlivosti z bootstrapu přes kazuistiky; svislice vyznačují konvenční prahy 0,667 a 0,80). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku](#) ↗).

Hodnocené konstrukty jsou úsudkové: co je pro danou situaci „vhodné“, závisí na pedagogickém přesvědčení hodnotitelky, a čtyři osy vyžadují vážení protichůdných signálů v jednom textu. Výsledky proto vypovídají o hodnocené vhodnosti, tedy o úsudku vyškolených anotátorek, a ne o vhodnosti řešení samé.

U rozpoznávání rámců dosáhl průměrný párový Jaccardův index hodnoty 0,73 v konvenci, která shodu dvou prázdných označení počítá jako souhlas (obě anotátorky se shodly, že řešení žádný rámec nenese). Protože u většiny řešení žádný rámec označen nebyl, uvádíme i přísnější konvenci, která tyto dvojice vynechává; index v ní klesá na 0,12. Rozdíl obou konvencí ukazuje, jak silně hodnota indexu závisí na tom, zda se tyto dvojice do shody započítají.

7.4 Výsledky

Deskriptivní obraz

Tabulka 7.1 shrnuje průměrná hodnocení podle zdroje řešení. Jako hodnotící lze číst především vhodnost a s výhradou osu reaktivní–proaktivní (oddíl 2.3); na obou vyznívají průměry

7. Studie 3: Řešení generovaná umělou inteligencí versus řešení učitelů

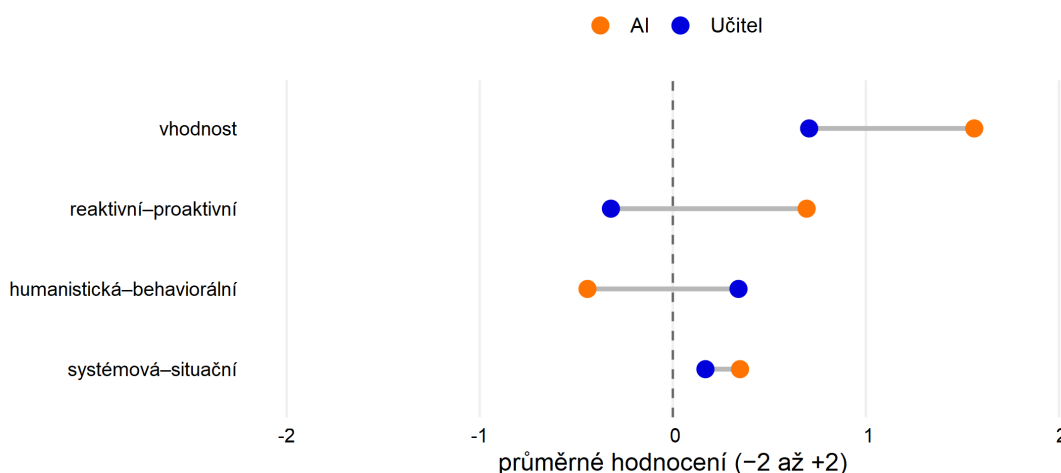
příznivěji pro generovaná řešení. Osa systémová–situační je popisná; osa humanistická–behaviorální je v manuálu vymezena zčásti hodnotově (oddíl 2.3), posun k humanistickému pólu proto znamená posun k postupu, který manuál popisuje jako vstřícný k potřebám žáka, ne nutně odklon od behaviorálních přístupů obecně. Na ose humanistická–behaviorální se generovaná řešení klonila k humanistickému pólu a učitelská k behaviorálnímu. Na ose systémová–situační se oba zdroje klonily k situačnímu pólu, generovaná řešení o něco výrazněji.

Tabulka 7.1. Průměrná hodnocení řešení podle zdroje (škály –2 až +2).

Dimenze	AI	Učitel
Vhodnost	1,56	0,71
Reaktivní–proaktivní	0,69	–0,32
Humanistická–behaviorální	–0,44	0,34
Systémová–situační	0,35	0,17

Poznámka. Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [výstupní tabulka ↗](#)).

Vedle rozdílů je podstatná i poloha průměrů na škále. Průměrná vhodnost generovaných řešení 1,56 leží blízko horního konce škály: anotátorky je hodnotily převážně jako vhodná. Blízkost stropu zároveň naznačuje možný stropový efekt, kvůli němuž může být rozdíl na horním konci škály podhodnocen. Učitelská řešení s průměrem 0,71 leží nad středem škály, na straně spíše vhodných řešení. Za špatná tedy hodnocena nebyla; generovaná řešení vedle nich dostávala ještě vyšší hodnocení.

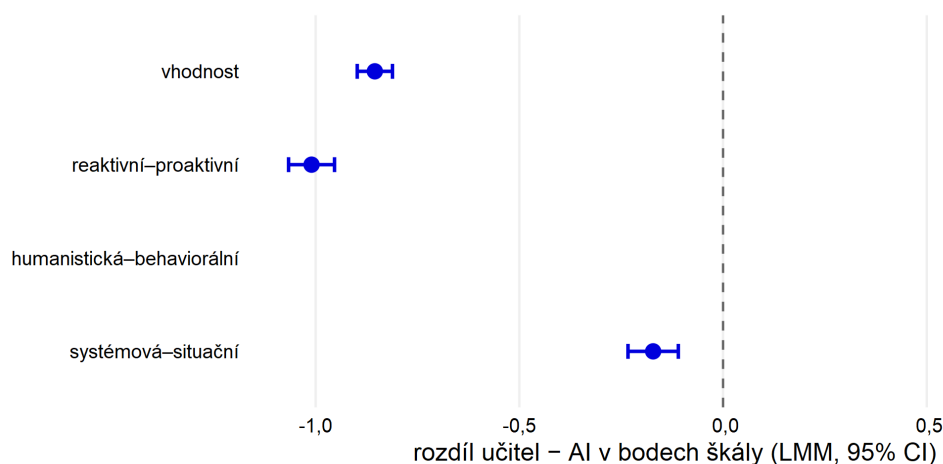


Obrázek 7.2. Průměrná hodnocení řešení podle zdroje (hodnoty viz Tabulka 7.1; oranžově řešení generovaná umělou inteligencí, modře řešení učitelů; u bipolárních dimenzí odpovídá záporný pól reaktivnímu, humanistickému a systémovému řešení). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Primární modely

Rozdíly přetrvávají i v lineárních smíšených modelech, které zohledňují opakovaná hodnocení týchž kazuistik a rozdílnou přísnost anotátorek. Učitelská řešení byla oproti generovaným hodnocena jako méně vhodná, s rozdílem $-0,85$ bodu škály (95% CI $-0,90$ až $-0,81$) a jako reaktivnější, s rozdílem $-1,01$ bodu ($-1,07$ až $-0,95$). Na ose humanistická–behaviorální se učitelská řešení posouvala k behaviorálnímu pólu o $0,77$ bodu ($0,71$ až $0,83$). Na ose systémová–situační byl rozdíl nejmenší: $-0,17$ bodu ($-0,23$ až $-0,11$). Učitelská řešení byla hodnocena o něco blíže systémovému pólu, oba zdroje se však klonily k situačním zásahům. Tento nejmenší rozdíl se mezi anotátorkami lišil (oddíl Citlivostní analýza níže).

7. Studie 3: Řešení generovaná umělou inteligencí versus řešení učitelů



Obrázek 7.3. Efekty zdroje řešení ze smíšených modelů: rozdíl učitel – AI v bodech škály s 95% intervalem spolehlivosti (hodnoty viz text; nula = žádný rozdíl mezi zdroji). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Rozpoznávání rámců

Anotátorky přiřadily celkem 874 rámcových štítků, z nich 72 % řešením generovaným umělou inteligencí. Podle logistického smíšeného modelu s náhodnými průsečíky pro kazuistiku a anotátorku mělo učitelské řešení nižší šanci, že v něm anotátorka rozpozná některý z pojmenovaných rámců: poměr šancí 0,31 (95% CI 0,26 až 0,38); šance učitelského řešení na rozpoznání rámce byla tedy zhruba třetinová. Podíl štítků a poměr šancí jsou různé míry: řešení může nést více štítků, kdežto logistický model sleduje jen to, zda anotátorka v řešení rozpoznala aspoň jeden rámec.

Instrukce pro generování žádný rámec nejmenovala; ukládala však modelu roli odborníka na řízení třídy a žádala co nejlepší postup (oddíl 7.2, příloha B). Rámcové rysy, které anotátorky v generovaných řešeních rozpoznávaly, tedy nevyplývají z pokynu uplatnit určitý rámec; mohou však souviset s touto rolí a s tímto požadavkem. Zjištění má dvě výhrady. Rozpoznání rámce je úsudek s nízkou shodou mezi anotátorkami (oddíl 7.3), takže vypovídá spíše o celkové tendenci než o spolehlivé klasifikaci jednotlivých řešení. A rozpoznaný rámec není totéž co věrné uplatnění přístupu; to by musely posoudit odbornice na příslušné

rámce.

Citlivostní analýza

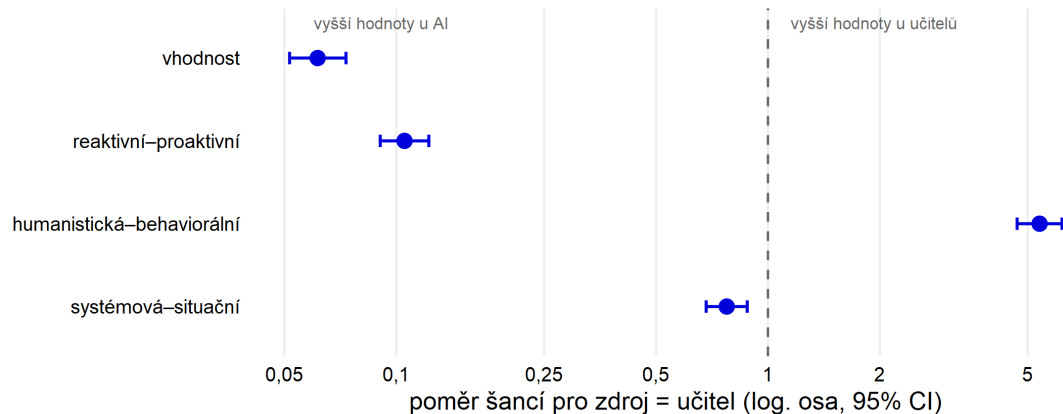
Kumulativní logitové smíšené modely (specifikace v oddíle 7.2) potvrzují směr všech čtyř efektů. Tabulka 7.2 uvádí poměry šancí s intervaly spolehlivosti, Obrázek 7.4 je vynáší na logaritmické ose.

Tabulka 7.2. Citlivostní CLMM: poměr šancí pro zdroj = učitel (referenční kategorie AI; Waldovy 95% intervaly spolehlivosti).

Dimenze	OR	95% CI
Vhodnost	0,06	0,05 až 0,07
Reaktivní–proaktivní	0,11	0,09 až 0,12
Humanistická–behaviorální	5,38	4,68 až 6,19
Systémová–situační	0,77	0,68 až 0,88

Poznámka. Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [výstupní tabulka ↗](#)).

7. Studie 3: Řešení generovaná umělou inteligencí versus řešení učitelů



Obrázek 7.4. Citlivostní CLMM: poměry šancí pro zdroj = učitel na logaritmické ose (95% Waldovy intervaly spolehlivosti; svislice OR = 1 značí žádný rozdíl mezi zdroji; hodnoty viz Tabulka 7.2). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Čtení tabulky ukážeme na vhodnosti: OR 0,06 znamená, že šance učitelského řešení na překročení kterékoli hranice mezi stupni škály (například na hodnocení „spíše vhodné“ či lepší oproti „vyvážené“ či horší) je při hodnocení téže kazuistiky touž anotátorkou mnohonásobně nižší než u řešení generovaného. Interval spolehlivosti je úzký a leží hluboko pod jedničkou; efekt tedy nezávisí na předpokladu stejných vzdáleností mezi stupni škály. Obdobně silný je efekt na ose reaktivní–proaktivní. Na ose humanistická–behaviorální míří opačným směrem (učitelská řešení mají vyšší šanci na hodnocení blíže behaviorálnímu pólu) a na ose systémová–situační je nejslabší, i jeho interval však leží celý pod jedničkou.

Druhá citlivostní analýza připouští, že se rozdíl mezi zdroji liší případ od případu. V LMM s náhodným sklonem zdroje po kazuistikách nevyšel u žádné ze čtyř dimenzí singulární (degenerovaný) odhad rozptylu a odhady efektů se od primárního modelu prakticky neliší: u vhodnosti $-0,86$ bodu (95% CI $-0,93$ až $-0,78$), na ose systémová–situační, kde je efekt nejslabší, $-0,16$ bodu ($-0,27$ až $-0,05$), s intervalem stále celým pod nulou. Závěry tedy nezávisí ani na předpokladu, že je rozdíl mezi zdroji ve všech kazuistikách stejný.

Třetí citlivostní analýza připouští, že se rozdíl mezi zdroji liší mezi anotátorkami. U vhodnosti, na ose reaktivní–proaktivní a na ose humanistická–behaviorální měl rozdíl u

všech šesti anotátorek týž směr a odhady se od primárních modelů prakticky nelišily (u osy humanistická–behaviorální s degenerovaným odhadem rozptylu sklonu; u vhodnosti $-0,86$ bodu, 95% CI $-0,95$ až $-0,77$). Na ose systémová–situační se anotátorky rozcházely: rozdíl učitel minus AI se u jednotlivých anotátorek pohyboval od $-0,87$ do $0,39$ bodu a v modelu, který tuto variabilitu připouští, činil $-0,14$ bodu s intervalem od $-0,44$ do $0,16$, který zahrnuje nulu. Rozdíl na této ose proto nelze považovat za stabilní výsledek; závisí na tom, kdo hodnotil.

Analýza omezená na úplné dvojice (obě řešení téže kazuistiky hodnocená touž anotátorkou na téže dimenzi) zachovává směr všech čtyř rozdílů. U vhodnosti činí rozdíl učitel minus AI $-0,86$ bodu ($-0,90$ až $-0,82$). Výsledek oslabuje obavu, že hlavní směr závisí na neúplných dvojicích; náhodnost chybení však nedokazuje.

7.5 Shrnutí

Studie 3 v zaslepeném uspořádání porovnála 324 dvojic řešení, které hodnotilo šest anotátorek na čtyřech dimenzích. Podle primárních smíšených modelů byla řešení generovaná umělou inteligencí hodnocena jako vhodnější a proaktivnější, učitelská jako behaviorálnější. Citlivostní ordinální modely (Tabulka 7.2) i model s náhodným sklonem zdroje po kazuistikách daly u všech efektů týž směr, s intervaly spolehlivosti celými mimo hodnotu nulového rozdílu; nejmenší rozdíl, na ose systémová–situační, se však neukázal stabilní napříč anotátorkami. Shoda anotátorek ležela u všech dimenzí pod nižším konvenčním prahem (oddíl 7.3); náhodné efekty zohledňují závislost hodnocení, neodstraňují však veškerou chybu měření. Přepočtení na úplných dvojicích směr rozdílů zachoval, náhodnost chybějících odpovědí však nedokazuje. Pojmenované rámce rozpoznávaly anotátorky častěji v generovaných řešeních, na něž připadly necelé tři čtvrtiny všech rámcových štítků (oddíl 7.4). Učitelská řešení byla hodnocena jako reaktivnější a behaviorálnější, což je v souladu s interpretačním pojmem „repertoáru nejmenšího odporu“ (oddíl 2.3), jemuž dává obsah Studie 1.

Výklad výsledků má několik omezení, k nimž se vrací kapitola 9. Hodnocena byla textová

7. Studie 3: Řešení generovaná umělou inteligencí versus řešení učitelů

podoba řešení, ne jeho proveditelnost v reálné třídě. Učitelská řešení jsou zpětným popisem postupu, který byl ve třídě použit v reálném čase a pod tlakem situace; generovaná řešení vznikla bez těchto omezení a na instrukci navrhnout co nejlepší postup z pozice odborníka. Srovnává se tedy popis skutečného jednání s navrženým postupem. Přes požadovanou paritu délky a stylu nelze zcela vyloučit, že hodnocení ovlivnily jazykové rysy generovaného textu. Dvě ze šesti anotátorek navíc část týchž kazuistik dříve kódovaly v první etapě (příloha C, Tabulka C.1), takže některá učitelská řešení mohly znát. Konečně se výsledky vážou k jedinému modelu (gpt-4o) a jediné instrukci; nelze je proto bez dalšího zobecňovat na jazykové modely obecně ani na jiné způsoby generování. Studie tedy nedokládá, že by generovaná řešení byla lepší než učitelská. Ukazuje, že generovaná řešení mají v tomto uspořádání textové vlastnosti, které vyškolené anotátorky spojují s vhodným, proaktivním a rámcově ukotveným postupem.

Pro uvažovanou podpůrnou roli takových systémů lze tento výsledek spojit se Studií 1. Ta popsala repertoár, v němž podle skladby kategorií převažují postupy reagující na vzniklou situaci. Ve Studii 3 anotátorky hodnotily generovaná řešení jako proaktivnější a častěji v nich rozpoznávaly pojmenované rámce; v tomto hodnocení se tedy generovaná řešení blížila postupům, které jsou v zaznamenaném repertoáru méně zastoupené. K situačním zásahům se přitom klonila řešení z obou zdrojů. Generované řešení tak může učiteli nad konkrétní situaci nabídnout kontrast k zaznamenanému postupu, aniž by nahrazovalo jeho úsudek o proveditelnosti; zda učitel takový postup už zná, studie nezjišťovala. Zda takový kontrast repertoár skutečně rozšiřuje, by musela ověřit intervenční studie, kterou tato kniha neprovádí. Výsledky Studie 3 pro ni vymezují realistická očekávání i rizika, mimo jiné to, že textová přesvědčivost není totéž co pedagogická proveditelnost. V následující kapitole (Studie 4) se pozornost přesouvá od řešení k samotným kazuistikám a k tomu, jak spolehlivě lze posoudit jejich kvalitu.

8. Studie 4: Kvalita kazuistik, spolehlivost jejího posouzení a predikce úspěchu řešení

Čtvrtá studie obrací pozornost od řešení k samotným kazuistikám. Ptá se, jakou mají kvalitu, jak spolehlivě ji lze posoudit a nakolik dokáže jazykový model předvídat úspěšnost řešení (VO4). Využívá třetí anotační etapu, v níž dvě vyškolené anotátorky posuzovaly téměř celý zveřejněný řez korpusu ve třech dimenzích: kvalita kazuistiky, dopad řešení a vhodnost řešení. Kapitola popisuje materiál a průběh etapy včetně rekalice uprostřed sběru (8.2), rozdělení hodnocení (8.3), spolehlivost posouzení včetně srovnání částí etapy před rekalicí a po ní (8.4), spojení s první etapou, které dovoluje ptát se, jaké typy řešení se pojí s úspěchem (8.5), a predikční experiment s jazykovým modelem (8.6).

8.1 Východiska

Předchozí studie pracovaly s kazuistikami jako s daným materiálem. Studie 1 ukázala, že popis výsledku je systematicky nejkratší částí kazuistiky (oddíl 5.2); právě z něj přitom vychází posouzení dopadu, které je ve Studii 4 cílem predikce. Studie 3 ukázala, že i vyškolené posuzovatelky se na úsudkových škálách shodují jen středně (oddíl 7.3). Třetí etapa se proto zaměřuje na meze obou předchozích studií (oddíl 3.5): ptá se, které kazuistiky vůbec umožňují interpretaci, jak dopadlo popsání řešení a jak vhodné bylo, a především jak spolehlivě lze tyto tři úsudky učinit.

8. Studie 4: Kvalita kazuistik, spolehlivost jejího posouzení a predikce úspěchu řešení

Kvalitou kazuistiky se zde rozumí její použitelnost: zda text poskytuje dost informací pro pochopení situace a interpretaci řešení a zda popsané jednání učitele neporušuje práva dítěte (příloha A.3). Dopad řešení pak není měřený výsledek, ale posouzení toho, co pisatel o výsledku napsal, tedy čtení jeho retrospektivní písemné výpovědi. Spolehlivost obou úsudků je proto vlastností dvojice posuzovatelka $\times$ text. Uvádíme ji koeficienty shody i hrubou shodou; u silně nevyvážených kategorií doplňujeme koeficient odolný vůči nevyváženosti (Gwet, 2008), protože κ a α mohou být v takovém případě nízké i při vysoké hrubé shodě (Feinstein & Cicchetti, 1990; kapitola 4).

8.2 Metoda

Materiál a anotátorky

Materiálem byl zveřejněný řez korpusu z května 2026 čítající 3 202 kazuistik (kapitola 5); hodnocení pokrývají téměř všechny z nich. Etapu provedly dvě anotátorky, v knize pod pseudonymy A2 a A12; obě prošly předchozími etapami: A2 kódovala už v první etapě, A12 přibyla ve druhé (příloha C). Kazuistiky byly rozděleny na dvě poloviny s vestavěným překryvem, takže každá anotátorka hodnotila přibližně tři pětiny řezu a část kazuistik hodnotily obě. Anotátorka A12 ohodnotila 1 947 kazuistik a anotátorka A2 celkem 1 938 kazuistik; dohromady jde o 3 185 různých kazuistik s jednotným identifikátorem. Dvojitě hodnocený překryv čítá 693 kazuistik. Kromě toho je v datech 7 hodnocených textů, které se nepodařilo spárovat s jednotným identifikátorem kazuistiky (kapitola 4); v rozděleních zůstávají, do překryvu a do spojení s první etapou nevstupují. Počet 3 192 je proto jen horní mezí počtu hodnocených kazuistik: platí za předpokladu, že každý z těchto textů představuje jiný případ.

Nástroje

Znění operacionalizace všech tří dimenzí uvádí příloha A.3. Kvalita kazuistiky je binární: plochá (nepoužitelná), nebo plná. Kód plochosti zahrnuje i normativní posouzení jednání učitele, nejde tedy jen o úplnost textu. Dopad řešení má tři kategorie: neúspěch, krátkodobý nebo částečný úspěch a dlouhodobý úspěch; oproti první etapě jsou částečný a krátkodobý úspěch sloučeny. Vhodnost řešení se hodnotí na čtyřbodové škále bez středu, od zcela nevhodného po zcela vhodné; s pětibodovou vhodností druhé etapy se nesměšuje (kapitola 4). Vedle kategorií vyplňovaly anotátorky tři volné komentáře: jazykový komentář, povinné zdůvodnění plochosti a další komentář.

U ploché kazuistiky předepisuje manuál dopad a vhodnost nevyplňovat. Data však obsahují výjimky: dopad je vyplněn u 29 plochých záznamů A12 a u 16 záznamů A2, vhodnost u 33 záznamů A12 a u 17 záznamů A2. Hlavní analýza tyto hodnoty ponechává; citlivostní analýza omezuje shodu dopadu a vhodnosti na případy, které obě anotátorky označily jako plné.

Průběh etapy a rekaliibrace

Etapa měla stejnou posloupnost jako obě předchozí: zácvek na šesti kazuistikách, první pilot na 19 kazuistikách, druhý pilot na 49 kazuistikách se společným rozbořem neshod a teprve poté finální anotování (příloha C). Na druhém pilotu byla shoda vysoká: hrubá shoda u dopadu i vhodnosti dosahovala zhruba devadesáti procent nebo více a κ leželo nad 0,8 (příloha A.3).

Během dlouhého finálního kódování se však výklad škály vhodnosti u obou anotátorek rozešel. Při průběžné kontrole v polovině července 2026, kdy měly obě anotátorky ohodnoceno 239 společných kazuistik, byla u vhodnosti hrubá shoda jen 45 % a vážená $\kappa = 0,42$. Rozdíl měl jasný směr: anotátorka A2 dávala v průměru 3,49 bodu, anotátorka A12 2,95, a tam, kde se lišily, dávala A2 vyšší známku ve 120 ze 128 případů. Výzkumný tým na to reagoval rekaliibrací. Anotátorky dostaly rekaliibrační podklad: pro každý stupeň škály kotevní kazuistiky,

8. Studie 4: Kvalita kazuistik, spolehlivost jejího posouzení a predikce úspěchu řešení

na nichž se shodly, a seznamy sporných kazuistik k projednání. U vhodnosti jich bylo 40, další se týkaly dopadu a kvality. Do týdne se anotátorky sešly s koordinátorkou týmu, sporné případy probraly a zbytek řezu dokódovaly podle sjednoceného výkladu škály. Rekalibrace uprostřed sběru je změnou protokolu; rozdíly mezi částmi etapy před ní a po ní popisuje oddíl 8.4. Všechny kazuistiky z těchto seznamů jsou z primárního výpočtu spolehlivosti vyloučeny, protože po projednání už nejde o nezávislá dvojí měření; dále je označujeme jako sporné kazuistiky.

Uzavření etapy

Sběr skončil 8. září 2026. Analýzy kapitoly vycházejí z finálních tabulek obou anotátorek k tomuto dni (řez ze září 2026); seznam kazuistik řezu je uložen v doprovodném repozitáři (Nehyba, 2026). Výsledek schůzky se promítl přímo do dat: u projednaných kazuistik změnila anotátorka A12 svou známku ve 28 případech, anotátorka A2 ve 2 a ve 23 případech se finální známky obou anotátorek shodují. Finální tabulky tak u sporných kazuistik obsahují vedle původních i hodnocení změněná po společném projednání. Spolehlivost proto uvádíme primárně bez sporných kazuistik a pro srovnání i na plném překryvu (oddíl 8.4).

Analýza

Shodu anotátorek vyjadřujeme hrubou shodou, Cohenovým κ (u vhodnosti váženým kvadraticky), Krippendorffovým α a Gwetovým AC1; u vhodnosti uvádíme i podíl hodnocení lišících se nejvýše o jeden stupeň a AC2 (váženou obdobu koeficientu AC1) s kvadratickými vahami. Interval spolehlivosti pocházejí z bootstrapu přes kazuistiky, tedy z opakovaného náhodného výběru kazuistik s vrácením (tisíc opakování). Primární výpočet vylučuje z překryvu 64 sporných kazuistik (sporných u některé ze tří dimenzí, z toho 40 u vhodnosti); výpočet na plném překryvu slouží jako citlivostní analýza. Při spojení s první etapou (oddíly 8.4 a 8.5) se kazuistiky párují výhradně přes jednotný identifikátor (kapitola 4); z první etapy

vstupuje primární kategorie chování a řešení a dopad převedený ze čtyř kategorií na tři sloučením částečného a krátkodobého úspěchu. Vztah typu řešení k dopadu popisujeme podíly a intervaly spolehlivosti z bootstrapu přes kazuistiky; jde o popis souběhů, ne o kauzální model. Uspořádání predikčního experimentu popisuje oddíl 8.6.

8.3 Rozdělení hodnocení

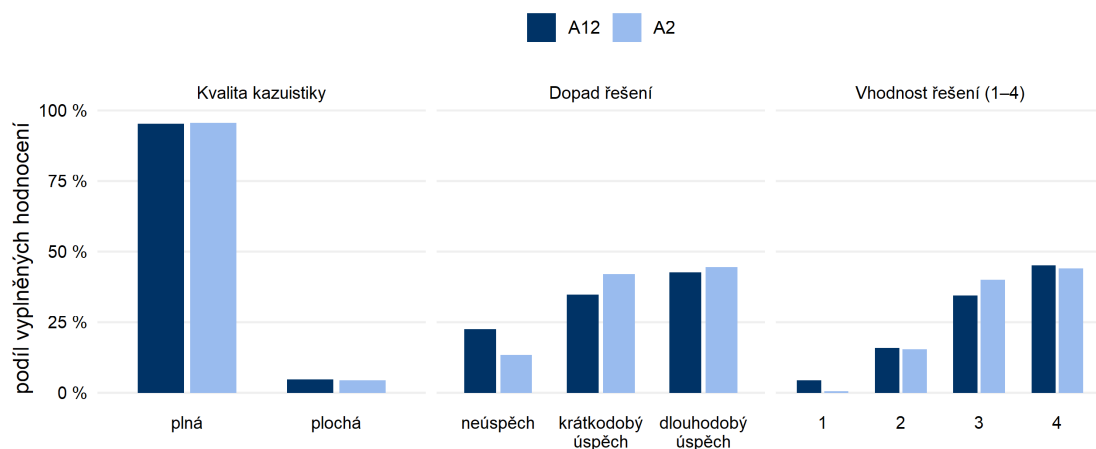
Tabulka 8.1 a Obrázek 8.1 shrnují rozdělení všech tří dimenzí zvlášť pro každou anotátorku. Rozdělení pocházejí převážně z různých kazuistik, protože každá anotátorka hodnotila jinou polovinu řezu; rozdíl mezi nimi tedy směřují rozdíl mezi případy a rozdíl v přísnosti. Oddělit obojí lze jen na překryvu (oddíl 8.4).

Tabulka 8.1. Rozdělení hodnocení ve třetí etapě podle anotátorky (počet kazuistik a podíl z vyplněných hodnocení dané dimenze).

Dimenze a kategorie	A12	A2
Kvalita: plná	1 855	1 852
Kvalita: plochá	92 (4,7 %)	86 (4,4 %)
Dopad: neúspěch	425 (23 %)	251 (13 %)
Dopad: krátkodobý nebo částečný úspěch	655 (35 %)	784 (42 %)
Dopad: dlouhodobý úspěch	803 (43 %)	833 (45 %)
Vhodnost 1 (zcela nevhodné)	85 (4,5 %)	11 (0,6 %)
Vhodnost 2 (spíše nevhodné)	300 (16 %)	288 (15 %)
Vhodnost 3 (spíše vhodné)	650 (34 %)	748 (40 %)
Vhodnost 4 (zcela vhodné)	852 (45 %)	822 (44 %)

8. Studie 4: Kvalita kazuistik, spolehlivost jejího posouzení a predikce úspěchu řešení

Poznámka. Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [výstupní tabulka ↗](#)). Anotátorky hodnotily převážně různé kazuistiky (překryv viz oddíl 8.4).



Obrázek 8.1. Rozdělení hodnocení ve třetí etapě podle anotátorky (podíl z vyplněných hodnocení; hodnoty viz Tabulka 8.1). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Podle obou anotátorek je korpus převážně použitelný. Plochých kazuistik označila A12 92 (4,7 %) a A2 86 (4,4 %). Zdůvodnění plochosti napsaly téměř vždy: A12 u 86 z 92 plochých, A2 u 83 z 86. Nejčastějším důvodem je chybějící řešení nebo výstup, tedy kazuistika, která končí popisem situace; méně často nesrozumitelná struktura textu a jen ojediněle jednání učitele porušující práva dítěte. Chybějící hodnocení dopadu a vhodnosti se týkají převážně plochých kazuistik. Například dopad chybí anotátorce A12 u 64 kazuistik, z nichž 63 jsou ploché, a anotátorce A2 u 70 kazuistik, z nichž je plochých všech 70.

U dopadu se anotátorky liší v tom, jak často kódují neúspěch: A12 u 23 % posouzených kazuistik, A2 u 13 %. Dlouhodobý úspěch kódují obě zhruba u dvou pětín. Vhodnost je u obou vychýlená ke kladnému pólu: stupně 3 a 4 pokrývají u A12 80 % a u A2 84 % hodnocení, stupeň 1 je vzácný. V praxi se tak škála využívá spíš jako trojstupňová.

Volné komentáře zachycují informace, které kódy nezachytí; anotátorky je však psaly velmi nerovnoměrně. Jazykový komentář napsala A12 u 38 kazuistik, A2 u 348; další komentář A12 u 21 a A2 u 258 kazuistik. V jazykových komentářích převažují překlepy a pravopis. U

46 kazuistik A2 a 14 kazuistik A12 komentář žádá překlad do češtiny (zpravidla ze slovenštiny) nebo vytýká anglické označení žáka („pupil“) v českém textu. Tuto jazykovou vrstvu korpusu uvádí kapitola 9 mezi omezeními. Ve 145 dalších komentářích anotátorky A2 stojí značka nejistoty „??“, nejčastěji u vhodnosti. Okruh nejistých případů je tedy širší, než jaký obsahoval rekalibrační podklad. Rozdíl v počtu komentářů přičítáme spíše pracovnímu stylu anotátorek než kazuistikám, a jako proměnnou jej proto nepoužíváme.

Anotátorky přitom posuzovaly souvislé texty: medián délky popisu situace je 664 znaků, anamnézy 478, řešení 610 a výstupu 410 znaků. Výstup je i zde nejkratší částí kazuistiky, stejně jako ve Studii 1.

8.4 Spolehlivost posouzení

Tabulka 8.2 shrnuje shodu obou anotátorek na dvojité kódovaném překryvu, primárně bez sporných kazuistik a pro srovnání na plném překryvu. U dopadu a vhodnosti mění vyloučení sporných kazuistik hodnoty jen málo; u vhodnosti je primární hodnota dokonce nepatrně nižší. U kvality je rozdíl větší: na plném překryvu je hrubá shoda 97 % a $\kappa = 0,67$. Vyloučené kazuistiky byly vybrány právě jako sporné, a primární výpočet proto shodu u kvality spíše nadhodnocuje. Malý rozdíl u dopadu a vhodnosti přitom nedokládá, že hodnocení sporných kazuistik zůstala po schůzce nezávislá. Obrázek 8.2 ukazuje matice záměn na plném překryvu.

Tabulka 8.2. Shoda anotátorek A12 a A2 na dvojité kódovaném překryvu třetí etapy (první tři řádky: primární výpočet bez sporných kazuistik z rekalibračního podkladu, v závorce 95% intervaly spolehlivosti z bootstrapu přes kazuistiky; řádky „plný překryv“: včetně těchto kazuistik).

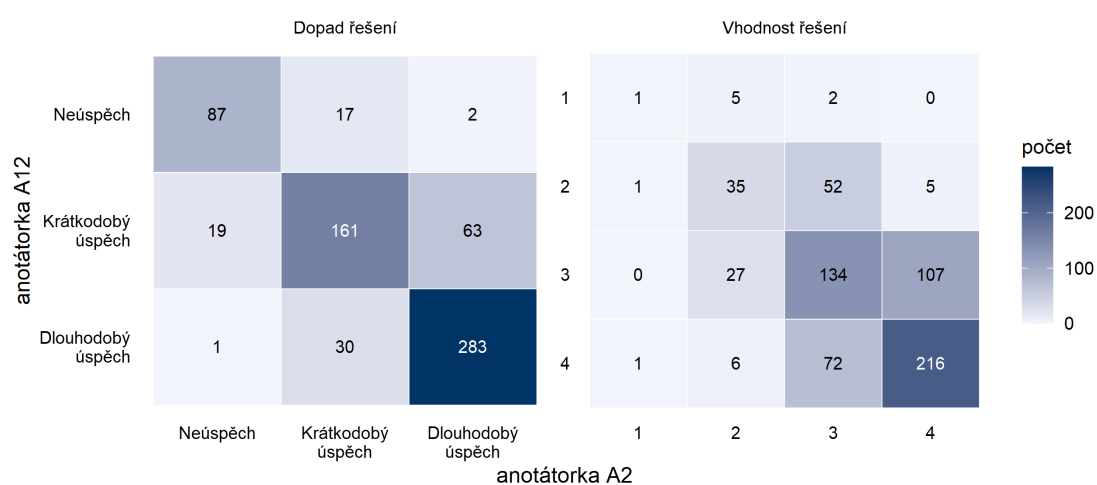
Dimenze	n	Hrubá shoda	κ (95% CI)	Gwet AC1
Kvalita	629	98 %	0,77 (0,64 až 0,88)	0,98
Dopad	603	81 %	0,69 (0,63 až 0,74)	0,73
Vhodnost (vážená	603	58 %	0,52 (0,46 až 0,58)	0,47

κ)

8. Studie 4: Kvalita kazuistik, spolehlivost jejího posouzení a predikce úspěchu řešení

Dimenze	n	Hrubá shoda	κ (95% CI)	Gwet AC1
Kvalita, plný překryv	693	97 %	0,67	0,96
Dopad, plný překryv	663	80 %	0,67	0,71
Vhodnost, plný překryv	664	58 %	0,53	0,48

Poznámka. Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [výstupní tabulka ↗](#)). U vhodnosti je κ kvadraticky vážené; Krippendorffovo α se od κ liší nejvýše v setinách.



Obrázek 8.2. Matice záměn na plném překryvu: řádky = anotátorka A12, sloupce = anotátorka A2; na diagonále je shoda. *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Kvalita

V primárním výpočtu se anotátorky na kvalitě shodly v 98 % kazuistik. Zde se projevuje paradox shody: plochých kazuistik je tak málo, že $\kappa = 0,77$ leží výrazně pod hrubou shodou a má široký interval spolehlivosti, zatímco AC1 = 0,98, méně citlivý na nevyváženost kategorií, zůstává blízko hrubé shodě. Posouzení použitelnosti je tedy mezi anotátorkami vysoce

shodné a podíl plochých kazuistik kolem pěti procent není artefaktem jedné posuzovatelky. Shoda sama ovšem nedokládá správnost posouzení vůči vnějšímu kritériu.

Dopad

U dopadu se anotátorky shodly v 81 % kazuistik, $\kappa = 0,69$. Neshody mají zřetelný vzorec (Obrázek 8.2). Zdaleka nejčastěji kóduje A12 krátkodobý úspěch tam, kde A2 kóduje dlouhodobý; záměny neúspěchu s krátkodobým úspěchem jsou méně časté a záměny neúspěchu s dlouhodobým úspěchem ojedinělé. Většina neshod se tedy týká trvalosti úspěchu: A2 hodnotí tentýž popis výstupu častěji jako trvalý úspěch. Neúspěch přitom obě anotátorky kódují na překryvu prakticky stejně často, takže rozdíl v podílu neúspěchu v Tabulce 8.1 odráží spíše rozdíl mezi hodnocenými polovinami korpusu než rozdílnou přísnost.

Vhodnost

Nejnižší shodu má vhodnost: hrubá shoda 58 %, vážená $\kappa = 0,52$. Neshoda je však téměř vždy posun o jediný stupeň: v 98 % kazuistik se známky liší nejvýše o jeden bod a AC2 s kvadratickými vahami dosahuje 0,87. Protichůdná hodnocení, kdy jedna anotátorka hodnotí řešení jako nevhodné a druhá jako zcela vhodné, se prakticky nevyskytují. Zůstává systematický posun: A2 hodnotí výš u 149 kazuistik, A12 u 105; průměry na překryvu jsou 3,30 u A12 a 3,37 u A2, rozdíl 0,07 bodu (95% CI 0,01 až 0,12). Vhodnost je hodnotový soud. Pracujeme s ní proto jako s hodnocením dvou vyškolených posuzovatelek, stejně jako ve Studii 3, a neklademe ji na jednu stupnici s pětibodovou vhodností druhé etapy.

Citlivostní analýza omezená na kazuistiky, které obě anotátorky označily jako plné, zahrnuje 593 úplných dvojic dopadu i vhodnosti. Dopad v ní dosahuje $\kappa = 0,69$ a vhodnost vážené $\kappa = 0,50$, obraz shody je tedy podobný. Ani tato analýza však neodděluje normativní složku kódování od popsaného výsledku.

Rozdíly mezi částmi etapy kolem rekaliibrace

Průběh etapy popisuje srovnání finálních tabulek se stavem dat z 22. července 2026, tedy z doby rekaliibrační schůzky, před zapracováním dohodnutých oprav sporných kazuistik a před veškerým pozdějším hodnocením. Jde o popis časových segmentů, nikoli odhad kauzálního účinku rekaliibrace. Srovnání po jednotlivých záznamech dělí finální řez na tři části: hodnocení pořízená do tohoto data a od té doby nezměněná (1 249 řádků), hodnocení pořízená až po tomto datu (2 598 řádků) a zpětné změny, tedy řádky ohodnocené do tohoto data a později změněné (38 řádků, z toho 37 u sporných kazuistik; ve 31 případech se změnila vhodnost).

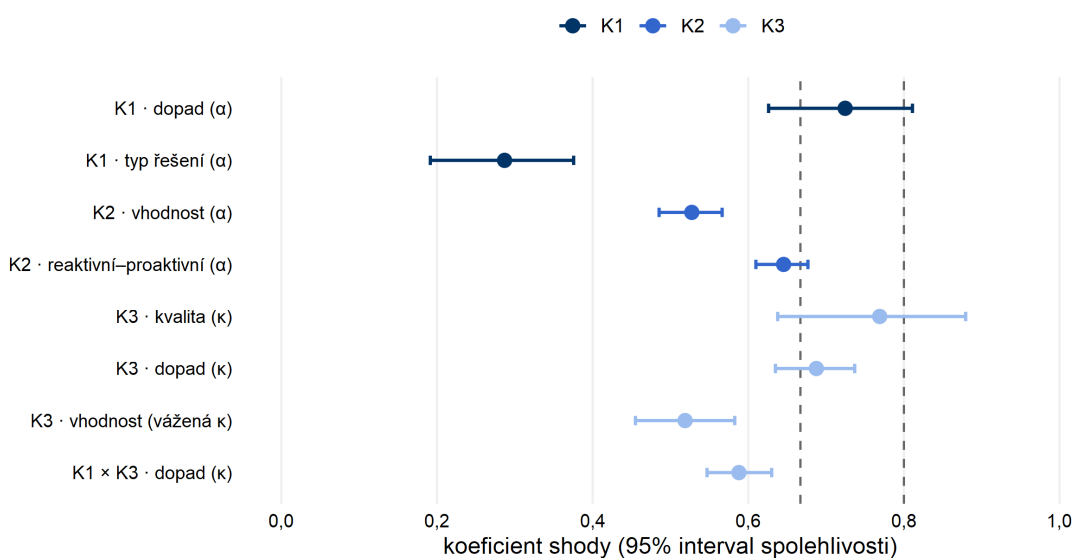
Pro interpretaci je podstatné nerovnoměrné rozvržení práce. K témuž datu měla anotátorka A2 ohodnoceno 1 048 kazuistik a mezi nimi celý pozdější překryv, zatímco anotátorka A12 měla ohodnoceno 239. Na celém překryvu tedy pocházejí hodnocení A2 z doby před rekaliibrací, kdežto A12 překryv z větší části dokódovala až po ní. Ještě v tomto dřívějším stavu ohodnotily obě anotátorky 203 kazuistik překryvu; dalších 454 ohodnotila A12 až po rekaliibraci a 35 svých dřívějších hodnocení na překryvu později změnila. Hodnocení A2 se přitom v celé tabulce změnila jen ve třech případech. Srovnání obou částí překryvu je proto srovnáním anotátorky A12 před rekaliibrací a po ní vůči nezměněným hodnocením A2.

Obě části se liší hlavně průměrným rozdílem v hodnocení vhodnosti. V části před rekaliibrací se anotátorky u vhodnosti shodly v 52 % kazuistik, průměry byly 3,04 u A12 a 3,47 u A2, rozdíl 0,43 bodu (95% CI 0,36 až 0,51), a tam, kde se lišily, hodnotila A2 výš u 90 kazuistik proti šesti. V části po rekaliibraci byla shoda 60 %, průměry 3,39 a 3,32, rozdíl –0,07 bodu (95% CI –0,13 až 0,00) a neshody byly vyrovnanější: A12 hodnotila výš u 101 kazuistik, A2 u 72. V pozdější části je průměrný rozdíl menší a opačně orientovaný. Vážená κ se však mezi částmi téměř neliší: 0,55 před rekaliibrací a 0,52 po ní. Výsledky odpovídají změně průměrného nastavení škály bez obdobného růstu vážené shody; samostatný účinek rekaliibrace z nich odvodit nelze. U dopadu je κ v obou částech srovnatelná (0,64 před a 0,67 po rekaliibraci). U kvality je srovnatelná hrubá shoda (96 % a 97 %), κ kvality je však v menší části před rekaliibrací velmi nejisté.

Srovnání má své meze. Obě části překryvu tvoří jiné kazuistiky a ty z první části byly pro anotátorku A12 vůbec první prací v etapě, takže se do rozdílu promítá i pokračující zácvik. Hodnocení A2 navíc pochází v obou částech z doby před rekalicací; měří se tedy posun jedné anotátorky vůči pevnému měřítku, ne sblížení obou. Překryv v Tabulce 8.2 obsahuje kazuistiky z obou částí, a jeho hodnoty proto nepopisují stav po rekalicí.

Shoda napříč etapami a opakované posouzení dopadu

Obrázek 8.3 shrnuje koeficienty tří etap. Etapy se liší nejen úsudkovými úlohami, ale i počtem kategorií, rozdělením štítků, vzorky a použitými koeficienty. Pořadí hodnot proto nelze číst jako test náročnosti úsudku ani jako společnou stupnici kvality měření.



Obrázek 8.3. Shoda anotátek napříč třemi anotačními etapami a mezi etapami ($K1 \times K3$): koeficienty s 95% intervaly spolehlivosti; svislice vyznačují konvenční prahy 0,667 a 0,80. Koeficienty pocházejí z různých uspořádání (pět anotátek na překryvu v první etapě, šest ve druhé, dvě ve třetí) a figura je řadí vedle sebe pro orientaci. *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Třetí etapa také umožňuje porovnat posouzení dopadu týchž textů s první etapou. Spojení přes identifikátor dává 926 kazuistik; po sloučení částečného a krátkodobého úspěchu a vy-

8. Studie 4: Kvalita kazuistik, spolehlivost jejího posouzení a predikce úspěchu řešení

loučení chybějících a prázdných štítků zbývá 885 úplných dvojic. Všechna hodnocení ze třetí etapy pocházejí v tomto souboru od A12, hodnocení z první etapy od pěti jiných anotátorek. Nejde tedy o opakované hodnocení touž osobou, ale ani o replikaci dvěma nezávislými týmy. Anotátorky první etapy navíc při kódování dopadu viděly pisatelovo označení kazuistiky jako úspěšně, nebo neúspěšně řešené (příloha A.1), což je při srovnání třeba vzít v úvahu. Shoda mezi etapami činí 73 % , $\kappa = 0,59$. Výsledek vypovídá o opakovatelnosti kódování v tomto souboru, sám o sobě však nedokládá konstruktovou validitu ani účinnost popsaného řešení.

8.5 Typ řešení, hodnocený dopad a vhodnost

Spojení obou etap dovoluje položit otázku, kterou Studie 1 mohla klást jen uvnitř jedné etapy (oddíl 5.5): jaké typy řešení se pojí s dlouhodobým úspěchem a s vyšší hodnocenou vhodností, když typ řešení určila první etapa a dopad i vhodnost později posoudila A12 bez znalosti dřívějších kódů. Tabulka 8.3 a Obrázek 8.4 uvádějí výsledky pro 917 kazuistik s primární kategorií řešení z první etapy. Jednotkou je opět dvojice kazuistika $\times$ anotátorka třetí etapy; uvedeny jsou kategorie s alespoň dvaceti kazuistikami.

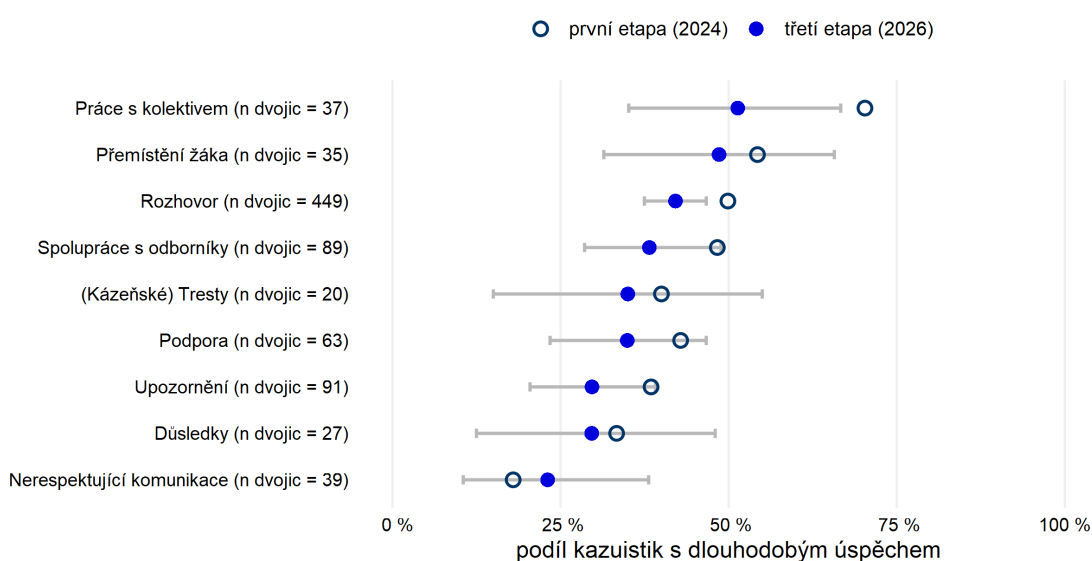
Tabulka 8.3. Typ řešení z první etapy (primární kategorie) a hodnocení dopadu na týchž úplných dvojicích oběma etapami: podíl dlouhodobého úspěchu podle první a podle třetí etapy (u třetí etapy s 95% intervalem spolehlivosti z bootstrapu přes kazuistiky) a průměrná vhodnost (1–4) ze třetí etapy.

Typ řešení (K1)	Celkem / úplné		Úspěch podle K3	
	dvojice	Úspěch podle K1	(95% CI)	Vhodnost (K3)
Práce s kolektivem	38 / 37	70 %	51 % (35 % až 67 %)	3,38
Přemístění žáka	35 / 35	54 %	49 % (31 % až 66 %)	3,20
Rozhovor	463 / 449	50 %	42 % (37 % až 47 %)	3,25
Spolupráce s odborníky	93 / 89	48 %	38 % (29 % až 49 %)	3,31
(Kázeňské) tresty	20 / 20	40 %	35 % (15 % až 55 %)	2,75
Podpora	66 / 63	43 %	35 % (23 % až 47 %)	3,43
Důsledky	28 / 27	33 %	30 % (12 % až 48 %)	2,78

8.5 Typ řešení, hodnocený dopad a vhodnost

Typ řešení (K1)	Celkem / úplné		Úspěch podle K3	
	dvojice	Úspěch podle K1	(95% CI)	Vhodnost (K3)
Upozornění	97 / 91	38 %	30 % (20 % až 39 %)	2,75
Nerespektující komunikace	45 / 39	18 %	23 % (11 % až 38 %)	2,10

Poznámka. Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [výstupní tabulka ↗](#)). Kategorie s méně než dvaceti kazuistikami (dohoda, fyzické zakročení) vynechány. Druhé číslo ve sloupci rozsahu je společným jmenovatelem obou podílů úspěchu, první označuje všechny dostupné kazuistiky kategorie. Vhodnost využívá vlastní dostupná hodnocení. Přesné jmenovatele uvádí výstupní tabulka.



Obrázek 8.4. Podíl kazuistik s dlouhodobým úspěchem podle primárního typu řešení z první etapy, posouzený oběma etapami na týchž úplných dvojicích (plné body = třetí etapa s 95% intervalem spolehlivosti z bootstrapu přes kazuistiky, prázdné = první etapa; jen kategorie s alespoň dvaceti kazuistikami; hodnoty viz Tabulka 8.3). *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Podíly dlouhodobého úspěchu se mezi kategoriemi liší, ale jejich pořadí není v obou etapách totožné. Pozdější hodnocení nejsou všeobecně přísnější: například u nerespektující komunikace je podíl dlouhodobého úspěchu vyšší ve třetí etapě. Úplné dvojice dovolují porovnat tytéž případy, neoddělují však vliv anotátorky, času a kódovacího schématu. Celková aso-

8. Studie 4: Kvalita kazuistik, spolehlivost jejího posouzení a predikce úspěchu řešení

ciace typu řešení s dopadem je slabá (Cramérovo $V = 0,13$) a sama nic neříká o příčinách hodnoceného dopadu. Intervaly malých kategorií jsou široké; rozdíly mezi jednotlivými kategoriemi jsme netestovali přímými kontrasty, a netvrdíme proto, že jsou průkazné.

Interpretaci dále omezuje několik okolností. Volba strategie závisí na situaci. Šablona sběru žádá od pisatele kazuistiku úspěšného i neúspěšného řešení (oddíl 3.3), takže rozdělení dopadu je zčásti dáno zadáním. Kazuistika je retrospektivní písemná výpověď, vícekrokové řešení zastupuje v analýze jediná primární kategorie a anotátorky první etapy při kódování dopadu viděly pisatelovo označení úspěchu (oddíl 8.4). Kód neúspěchu z první etapy i vyřazení plochých případů navíc obsahují normativní složku (příloha A); to, že texty v obou etapách četly různé anotátorky, tento problém neodstraňuje. Zjištění tedy popisuje souvislost anotačních kategorií. Ověření účinnosti strategií by vyžadovalo oddělené měření přijatelnosti a výsledku, ideálně v intervenčním designu.

8.6 Predikce dopadu řešení jazykovým modelem

Druhá část VO4 se ptá, zda jazykový model z popisu situace, anamnézy a řešení odhadne dopad řešení tak, jak jej anotátorky kodovaly z popisu výstupu. Model dostává jen tyto tři části kazuistiky, bez výstupu. Test úniku informace hledal ve vstupu doslovný začátek výstupu; parafráze, pozdější části výstupu ani jiné významové nápovědy proto vyloučit nelze (příloha B.6). Model předpovídá jednu ze tří kategorií dopadu podle definic třetí etapy (příloha A.3), v základní variantě bez příkladů a v doplňkové variantě s pevnou sadou šesti příkladů z kazuistik mimo hodnoticí množinu. Znění instrukce pro model (anglický prompt) uvádí příloha B.6.

Experiment běžel na jazykových modelech nasazených v národní akademické infrastruktuře e-INFRA CZ. Texty kazuistik tak neopouštějí akademickou infrastrukturu a výzkum méně závisí na komerčních službách (Ollion et al., 2024). Predikce ovšem neběží na témže modelu jako Studie 2 a 3, a výsledky se proto napříč studiemi nesrovnávají.

Čtyři kandidátní modely (kimi-k3, glm-5.3, mistral-medium-3.5 a qwen3.5-int4) byly nejprve porovnány na vývojové množině, kterou tvoří kazuistiky mimo dvojitě kódovaný překryv. Primárním modelem se stal kimi-k3, který na ní dosáhl nejvyšší vyvážené přesnosti. Na hodnoticí množině jsme pro srovnání vyhodnotili i druhý model v pořadí, glm-5.3, a jen u něj také variantu s příklady v instrukci. Hodnoticí množinu tvoří dvojitě kódovaný překryv třetí etapy bez sporných kazuistik a bez 2 kazuistik, u nichž test úniku informace zachytil popis výstupu v zadání. Hodnoticí množina tak čítá 601 kazuistik. Předpovědi srovnáváme s hodnocením každé anotátorky zvlášť a s podmnožinou kazuistik, na jejichž dopadu se obě shodly.

Protože třídy dopadu jsou nevyvážené, uvádíme vyváženou přesnost, macro-F1 (průměr měř F1 jednotlivých tříd; F1 spojuje podíl správných předpovědí dané třídy s podílem zachycených případů této třídy), prostou přesnost a κ . Prázdná nebo neplatná odpověď modelu se počítá jako chybná předpověď a zůstává ve jmenovateli; při výpočtu κ tvoří samostatnou kategorii, která se s žádným lidským hodnocením neshoduje.

Shodu anotátorek, které popis výstupu znaly, uvádíme jako orientační srovnání, nikoli jako horní mez predikce: lidé a model mají k dispozici jiné informace a konsenzuální podmnožina má jiné složení než celý překryv. Člověka a model v téže úloze by bylo možné porovnat jen s lidskou predikcí ze stejných vstupů.

Model kimi-k3 vygeneroval na celé hodnoticí množině 13 neplatných odpovědí, každá z obou variant modelu glm-5.3 po 7; na konsenzuální podmnožině 11, respektive 6. Tabulka 8.4 uvádí metriky, v nichž jsou neplatné odpovědi započteny jako chybné. Pokrytí, intervaly spolehlivosti všech metrik a κ vypočtené jen z platných odpovědí obsahuje zpráva o analýze a tabulka predikčních výsledků v doprovodném repozitáři (Nehyba, 2026).

Tabulka 8.4. Predikce dopadu řešení jazykovým modelem na hodnoticí množině (dvojitě kódovaný překryv bez sporných kazuistik z rekalibračního podkladu; „shoda obou“ = podmnožina, na níž se anotátorky shodly). Referenční rámec tvoří shoda anotátorek mezi sebou a hádání většinové třídy.

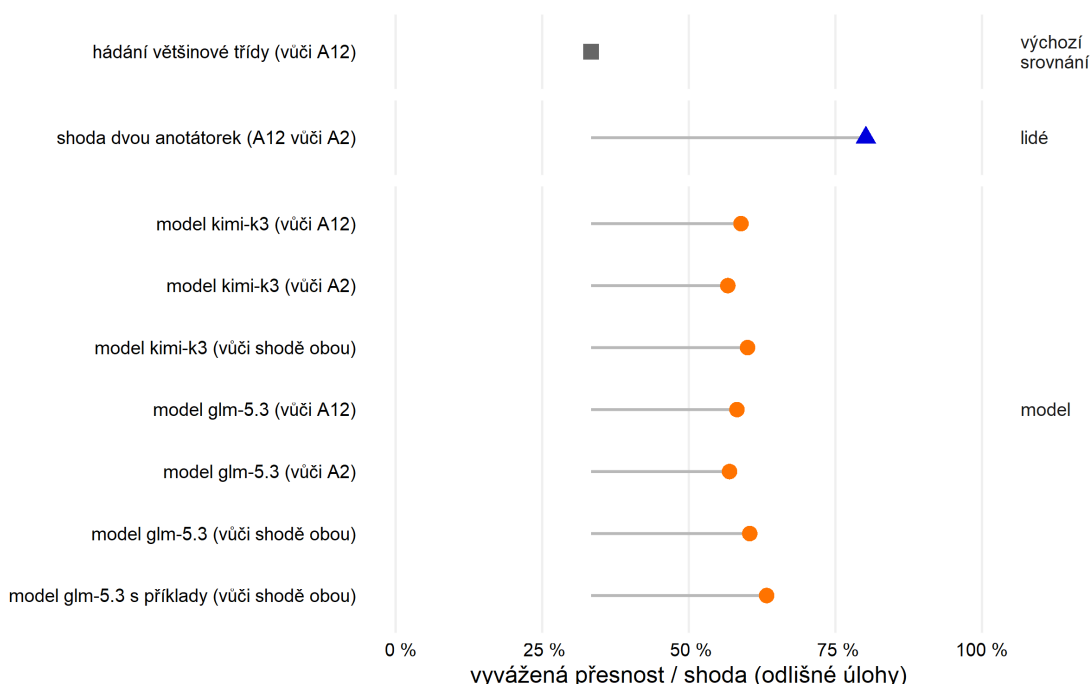
8. Studie 4: Kvalita kazuistik, spolehlivost jejího posouzení a predikce úspěchu řešení

Srovnání	n	Přesnost	Vyvážená		
			přesnost	macro-F1	κ
kimi-k3 × A12	601	52 %	59 %	0,54	0,29
kimi-k3 × A2	601	49 %	57 %	0,51	0,25
kimi-k3 × shoda obou	487	52 %	60 %	0,55	0,30
glm-5.3 × shoda obou	487	52 %	60 %	0,55	0,30
glm-5.3 s příklady × shoda obou	487	56 %	63 %	0,58	0,34
A12 × A2 (čtení výstupu)	601	81 %	80 %	0,80	0,69
hádání většinové třídy (vůči A12)	601	49 %	33 %	–	0

Poznámka. Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [výstupní tabulka ↗](#)). Anotátorky popis výstupu při hodnocení viděly, model nikoli; poslední dva řádky proto nejsou predikce, ale referenční rámec.

Vyvážená přesnost modelu kimi-k3 vůči společnému štítku obou anotátorek činí 60 %, tedy víc než třetina, kterou dává hádání většinové třídy. Neměří však správnost předpovědi skutečného výsledku ve třídě: cílem je štítek odvozený z pisatelova vyprávění. Obrázek 8.5 ukazuje predikci vedle lidského čtení popsaného výstupu.

8.6 Predikce dopadu řešení jazykovým modelem



Obrázek 8.5. Vyvážená přesnost předpovědi dopadu proti dvěma referenčním bodům: hádání většinové třídy (vůči A12) a shodě anotátorek mezi sebou (A12 vůči A2). Panely: oranžově model, modře lidé, šedě výchozí srovnání (hádání většinové třídy); vodorovná úsečka měří vzdálenost od této výchozí hodnoty. *Poznámka.* Vlastní analýza; kód a data v doprovodném repozitáři (Nehyba, 2026; [kód obrázku ↗](#)).

Párové bootstrapové porovnání na týchž případech dává rozdíl vyvážených přesností (glm-5.3 minus kimi-k3) 0,004 s intervalem od $-0,039$ do $0,048$. U modelu glm-5.3 činí rozdíl varianty s příklady a bez nich 0,029 s intervalem od $-0,011$ do $0,065$. Oba intervaly zahrnují nulu. Výhodu žádné varianty tedy prokázat nelze; zároveň z toho neplyne, že jsou modely rovnocenné nebo že úloha narazila na informační strop.

Model kimi-k3 nejčastěji předpovídá prostřední kategorii: krátkodobý nebo částečný úspěch u 52 % kazuistik, zatímco anotátorky jej kódují u 35 % a 31 %. Dlouhodobý úspěch naopak předpovídá jen u 22 % kazuistik, anotátorky jej kódují u 49 % a 53 %. Tomu odpovídá i citlivost po třídách, tedy podíl případů dané třídy, které model předpověděl správně. Vůči hodnocení anotátorky A12 činí u neúspěchu 77 %, u dlouhodobého úspěchu jen 38 %. Proč model dlouhodobý úspěch zachycuje hůř, ze samotné matice předpovědí určit nelze.

8. Studie 4: Kvalita kazuistik, spolehlivost jejího posouzení a predikce úspěchu řešení

Předpovědi lze ještě rozdělit podle toho, zda se anotátorky shodly. U kazuistik se shodou odpovídala předpověď jejich společné známce v 52 % případů; u kazuistik, kde se anotátorky lišily, odpovídala jedné z obou známek v 89 %. Druhé kritérium je mírnější, protože za zásah počítá dvě ze tří kategorií, a často by ho splnila i náhodná předpověď; s přesností vůči společnému štítku ho proto nelze přímo srovnávat.

Odpověď na druhou část VO4 je tedy omezená: předpovědi modelů se v tomto souboru shodují s kódovaným dopadem jen částečně. Z výsledků nelze určit, zda vstupní text potřebnou informaci neobsahuje, nebo ji použité modely a instrukce nevyužívají. Pro doporučování postupů ve třídě taková validace nestačí.

8.7 Shrnutí

Třetí etapa pokrývá 3 185 kazuistik s jednotným identifikátorem a 7 dalších hodnocených textů bez něj. Posouzení použitelnosti je mezi anotátorkami vysoce shodné; shoda však sama neznamená správnost vůči vnějšímu kritériu a primární výpočet bez sporných kazuistik ji u kvality spíše nadhodnocuje. Shoda dopadu je nižší a neshody se týkají hlavně trvalosti úspěchu. Nejnížší je shoda vhodnosti, hodnocení se však téměř vždy liší nejvýše o jeden stupeň. Rozdíly mezi částmi etapy před rekalicací a po ní neurčují samostatný účinek rekalicace. Srovnání s první etapou je jiným čtením týchž textů jednou pozdější anotátorkou a souvislosti typů řešení s hodnoceným dopadem nejsou ověřením jejich pedagogické účinnosti.

Predikční experiment ukazuje omezenou shodu předpovědí s kódovaným dopadem, nikoli informační strop úlohy. Neplatné odpovědi zůstávají ve výpočtu a intervaly párových rozdílů modelů zahrnují nulu. Lidské čtení výstupu nelze považovat za predikci za stejných podmínek. Následující kapitola s těmito omezeními počítá při spojení pedagogické a metodologické linie knihy.

Integrace

9. Diskuse

Čtyři studie této knihy kladly čtyři otázky nad jedním korpusem. Tato kapitola jejich výsledky propojuje. Nejprve shrnuje, co která studie zjistila a s jakou jistotou (9.1). Poté rozvíjí tři témata, která studie spojují: repertoár nejmenšího odporu jako interpretační pojem pro převahu snadno dostupných reakcí v zaznamenaných situacích (9.2), otázku, co lze jazykovému modelu v pedagogickém výzkumu svěřit (9.3), a spolehlivost lidského posouzení (9.4). Následují omezení (9.5), důsledky pro přípravu učitelů, pro platformu Edustories a pro výzkum s jazykovými modely (9.6) a směry dalšího výzkumu (9.7).

9.1 Oblouk čtyř studií

Studie 1 vybudovala a popsala korpus, Studie 2 porovnávala kódování typů řešení jazykovým modelem s lidským kódováním, ve Studii 3 lidé zaslepeně hodnotili řešení vygenerovaná modelem a Studie 4 zjišťovala, jak shodně lidé posuzují kvalitu kazuistik a úspěšnost řešení a nakolik lze dopad řešení předpovědět jazykovým modelem. Tabulka 9.1 shrnuje hlavní zjištění a míru jistoty, s níž je kniha předkládá.

9. Diskuse

Tabulka 9.1. Výzkumné otázky, hlavní zjištění a míra jistoty.

Otázka	Hlavní zjištění	Míra jistoty
VO1: korpus a jeho obsah (Studie 1)	Zveřejněný řez čítá 3 202 kazuistik se čtyřdílnou strukturou. V repertoáru dominuje rozhovor (v 922 z 1 359 kazuistik lidsky kódovaných v řezu ze srpna 2024) a další reaktivní kroky; vztahová a kolektivní řešení se pojí s příznivěji kódovaným výsledkem.	Vysoká pro rozsah a strukturu korpusu; četnosti platí na úrovni kategorií a omezuje je střední až nízká shoda anotátorek (kapitola 6); souběhy řešení a dopadu jsou popis, ne příčina, a kód neúspěchu zahrnuje i normativní posouzení prostředku.
VO2: model jako anotátor (Studie 2)	Shoda s lidským kódem typu řešení $\kappa = 0,64$ na úplných dvojicích; vzorek nebyl oddělen od vývoje instrukce pro model (anglicky prompt), nejde tedy o nezávislé ověření. Po dodatečném vyloučení překryvů s příklady v instrukci $\kappa = 0,61$.	Omezená validace; lidskou shodu z jiného uspořádání nelze použít jako horní mez shody modelu.
VO3: generovaná versus učitelská řešení (Studie 3)	Učitelská řešení hodnocena jako méně vhodná (rozdíl $-0,85$ bodu), reaktivnější a behaviorálnější; šance na rozpoznání rámce třetinová (OR 0,31).	Vysoká pro směr rozdílů, který potvrdily i citlivostní modely, v textové podobě řešení jednoho modelu a jedné instrukce; nevypovídá o proveditelnosti.

Otázka	Hlavní zjištění	Míra jistoty
VO4: kvalita, spolehlivost, predikce (Studie 4)	Posouzení použitelnosti kazuistiky s vysokou vzájemnou shodou ($\kappa = 0,77$ bez kazuistik projednaných při rekalibraci, $AC1 = 0,98$), dopadu s nižší shodou ($\kappa = 0,69$), jejíž neshody se soustřeďují do rozlišení trvalosti úspěchu, vhodnosti se střední shodou (vážená $\kappa = 0,52$); shoda dopadu při opakovaném čtení týchž textů jinou anotátorkou $\kappa = 0,59$. Předpověď dopadu jazykovým modelem dosáhla vyvážené přesnosti 60 % proti 33 % u hádání většinové třídy; anotátorky, které popis výsledku četly, se mezi sebou shodly s vyváženou přesností 80 % (jiná úloha, nejde o předpověď).	Vyloučení sporných kazuistik shodu spíše nadhodnocuje, nejvíce u kvality (oddíl 8.4); shoda mezi etapami vznikla za jiných podmínek čtení (příloha A.1); vzájemná shoda není správnost vůči externímu kritériu; predikce je omezená a lidé na rozdíl od modelu popis výsledku četli.

Poznámka. Hodnoty z výsledků analýz studií 1 až 4 (kapitoly 5 až 8); intervaly spolehlivosti a podmínky výpočtu uvádějí příslušné kapitoly.

Výsledky studií spolu souvisejí věcně i metodologicky. Zaznamenaná řešení ze Studie 1 tvoří ve Studii 3 srovnávací pozadí pro řešení generovaná modelem a ve Studii 4 se k týmž kazuistikám vrací opakované čtení. Každý výsledek o modelu je vztažen k lidskému posouzení, jehož shodu kniha měří. Model byl nejprve porovnán s lidským kódováním typů řešení, u něhož byla nízká i shoda anotátorek mezi sebou (kapitola 6); poté lidé zaslepeně hodnotili řešení, která model vygeneroval; nakonec byla změřena shoda lidských úsudků o kvalitě kazuistik a dopadu řešení a s kódovaným dopadem byla porovnána předpověď modelu. Žádný z těchto kroků nestanovuje horní hranici lidské ani modelové spolehlivosti.

9.2 Repertoár nejmenšího odporu

Kapitola 2 zavedla repertoár nejmenšího odporu jako interpretační pojem pro převahu snadno dostupných reakcí v zaznamenaných situacích. Studie mapují četnosti a vlastnosti textů; neměří však úplný repertoár jednotlivce, jeho znalosti, časový tlak ani vznik konkrétního rozhodnutí. Mechanismy učňovství pozorováním a situační dostupnosti proto zůstávají hypotézami pro další výzkum.

Obsah pojmu dávají výsledky Studie 1. Kazuistiky zapsali převážně studující učitelství na praxi. Rozhovor se objevuje ve dvou třetinách kódovaných kazuistik, po něm následují spolupráce s odborníky, podpora, důsledky a upozornění. Samostatnou proaktivní kategorii finální schéma neobsahovalo: proaktivní řešení bylo při vývoji schématu převedeno mezi podřazené kódy podpory žáka (příloha A.1). Chybějící kategorie proto nesvědčí o tom, že by proaktivní jednání chybělo.

Studie 3 doplnila srovnání s řešeními generovanými modelem. Šest anotátorek zaslepeně hodnotilo učitelské řešení vedle řešení vygenerovaného pro tutéž situaci; šlo o kazuistiky, které platforma v době výběru vedla v kategoriích fyzické a verbální agrese (kapitola 7). Učitelské řešení bylo reaktivnější (rozdíl $-1,01$ bodu) a behaviorálnější (o $0,77$ bodu) a šance, že v něm anotátorky rozpoznají některý z pojmenovaných rámců řešení náročného chování, byla oproti generovanému řešení zhruba třetinová.

Ve Studii 4 posoudila část kazuistik kódovaných ve Studii 1 znovu pozdější anotátorka. Vzor ze Studie 1 se v tomto srovnání potvrdil jen zčásti. Podíly dlouhodobého úspěchu se mezi kategoriemi řešení liší, celková souvislost typu řešení s dopadem je však slabá a pořadí kategorií není v obou etapách stejné (Tabulka 8.3). Záleží také na vzorku a na tom, zda se počítá jen primární kategorie řešení. Přemístění žáka patřilo v celém kódovaném řezu k nejméně úspěšným kategoriím (oddíl 5.5), na úplných dvojicích s primární kategorií však mezi ně nepatří podle žádné z obou etap. Na týchž dvojicích činil podíl dlouhodobého úspěchu u práce s kolektivem podle první etapy 70 % a podle třetí 51 %, u rozhovoru 50 % a 42 %, u upozornění 38 % a 30 %. U těchto kategorií jsou podíly třetí etapy nižší, u všech kategorií to

však neplatí (Tabulka 8.3).

Možné příčiny takového repertoáru popisuje kapitola 2 dvěma mechanismy, s nimiž jsou data knihy v souladu. Podle teze o učňovství pozorováním (Lortie, 1975) si budoucí učitelé osvojují postupy, které viděli z lavice, například napomenutí, domluvu, trest nebo poslání za dveře. Proaktivní práce může být z lavice méně viditelná a časový tlak může zvýhodnit dostupné postupy. Repertoár nejmenšího odporu je tedy hypotetický výklad vztahu mezi profesním učením a podmínkami jednání. V pojetí, které nabízí Swidler (1986), jde o zásobu, z níž se strategie skládají, ne o volbu mezi známými alternativami. S tímto výkladem se shodují domácí studie, které reaktivní strategie zachytily přímým pozorováním praxe (Vlčková et al., 2020) a v situacích začínajících učitelů (Karasová & Nehyba, 2025), i přehledové studie, které proaktivním a strukturovaným strategiím přiznávají větší přínos (Korpershoek et al., 2016; Simonsen et al., 2008).

Reaktivní repertoár neznamena, že zaznamenaná řešení zůstávají u jednoho aktéra; spolupráce s odborníky patří k nejčastějším krokům. I ona je však v kazuistikách popsána jako reakce na konkrétní událost, k níž šablona kazuistiky každé řešení vztahuje. Vzorec dopadů zůstává popisem souběhů, jak upozorňují kapitoly 5 a 8. Typ zaznamenaného řešení není na situaci nezávislý, trest se mohl objevovat spíše v situacích s horšími vyhlídkami a dopad posuzuje čtenář pisatelova zpětného záznamu. Ani opakované čtení jinou anotátorkou kauzální výklad nezakládá; navíc probíhalo bez pisatelova označení úspěchu, které anotátorky první etapy viděly (příloha A.1). Kódy první etapy také částečně spojují výsledek s přijatelností prostředku, a bez jejich oddělení proto vztah kategorií nelze pokládat za nezávislou validaci účinnosti.

9.3 Co lze jazykovému modelu svěřit

Knihy prověřuje jazykové modely ve třech rolích: při kódování typů řešení, při generování alternativních řešení a při předpovědi dopadu; v predikci přitom šlo o jiné modely než ve Studiích 2 a 3 (oddíl 8.6). Ve Studii 2 dosáhla shoda modelu s lidským kódem typu řešení

$\kappa = 0,64$. Srovnávací vzorek však nebyl oddělen od vývoje instrukce pro model, takže nejde o nezávislé ověření. Po dodatečném vyloučení kazuistik, jejichž řešení byla mezi příklady v instrukci, činí $\kappa = 0,61$. Přenos poznatků z krátkých anglických textů (Gilardi et al., 2023; Törnberg, 2025) na tento korpus je proto podpořen jen omezeně. Část neshod může souviset se schématem: vybraná neshoda v oddílu 6.5 ilustruje rozdíl mezi hodnotícím a doslovným čtením jednání, o podílu takových neshod však nevypovídá.

Při generování řešení vyšel model v zaslepeném lidském hodnocení příznivě. Jeho řešení byla hodnocena jako vhodnější a proaktivnější a anotátorky v nich častěji rozpoznávaly rysy pojmenovaných rámců. Instrukce žádný rámec nejmenovala, ukládala však modelu roli odborníka na řízení třídy a žádala co nejlepší postup (oddíl 7.4); rozpoznání rámce je navíc úsudek s nízkou shodou (oddíl 7.3). Závěr, že „AI radí lépe než učitel“, z toho nevyplyvá. Hodnocen byl text, ne jednání ve třídě; model odpovídal bez časového tlaku a bez odpovědnosti za následky; výsledky se vážou k jedinému modelu a jediné instrukci. Lze z nich vyvodit jen to, že generované řešení nabízí učiteli nad konkrétní situací kontrast k zaznamenanému postupu. Zda učitel takový postup už zná, studie nezjišťovala; zda kontrast jeho repertoár rozšiřuje, může ukázat až intervenční studie.

Posuzování vhodnosti řešení modelem kniha přímo netestovala, a z lidské shody proto nelze usuzovat na úspěch ani selhání modelu v této roli. Studie 4 však ukazuje, proč by takové ověření bylo obtížné. I dvě vyškolené anotátorky posuzují vhodnost řešení jen se střední shodou (vážená $\kappa = 0,52$) a ve druhé etapě šest anotátorek s $\alpha = 0,53$. Model posuzující vhodnost by tedy byl ověřován proti málo spolehlivému měřítku a jeho „shoda s lidmi“ by vypovídala hlavně o tom, se kterou anotátorkou se shoduje.

Dopad řešení, jak jej pisatel popsal, posuzují lidé s vyšší shodou ($\kappa = 0,69$ uvnitř etapy, $\kappa = 0,59$ mezi etapami); nejistější je rozlišení trvalosti úspěchu (oddíl 8.4). Dopad je zároveň veličinou, kterou model ve Studii 4 předpovídal. Predikce má jinou logiku než posouzení: anotátorky popis výsledku četly, model jej záměrně nedostal. Srovnání proto neříká, zda model předvídá lépe, nebo hůře než člověk; shoda dvou anotátorek je jen kontextovou referencí, ne měřítkem téže úlohy (oddíl 8.6). Model kimi-k3 dosáhl vyvážené přesnosti 60 %

vůči kódu dopadu, na němž se obě anotátorky shodly. Intervaly párových rozdílů mezi modely a variantami zahrnují nulu (oddíl 8.6), výsledek tedy nedokládá ekvivalenci modelů ani informační strop úlohy. Chybí lidská předpověď při stejné informaci a měření skutečného dopadu ve třídě; pro doporučení v jednotlivých případech proto tato validace nestačí.

Z toho plynou opatrná doporučení. Klasifikaci obsahu velkého korpusu na úrovni kategorií lze modelu svěřit jen po validaci na vlastních datech proti lidskému kódování se změřenou shodou a s lidskou kontrolou u hodnotících a významově širokých kategorií (kapitola 6; Reiss, 2023; Ziems et al., 2024). Generovaná řešení lze zkoušet jako kontrastní materiál k reflexi; jejich přínos musí teprve ověřit intervenční studie. Posuzování vhodnosti bez lidské reference ani rozhodování o jednotlivém žákovi modelu svěřit nelze. Protože poskytovatelé modely průběžně mění (Ollion et al., 2024), platí každý takový výsledek jen pro použitý model a instrukci (příloha B).

9.4 Spolehlivost lidského posouzení v průběhu etap

Spolehlivost lidského posouzení se lišila podle posuzovaného konstruktů a schématu a lišily se i hodnoty z pilotních a finálních kol, i když nejde o přímo srovnatelná měření. Uprostřed třetí etapy se navíc posunulo nastavení škály vhodnosti u jedné anotátorky, zatímco vážená shoda zůstala prakticky stejná (oddíl 8.4).

Obrázek 8.3 staví koeficienty všech tří etap vedle sebe. Nejvyšší jsou u binární kvality kazuistiky, u níž primární výpočet bez sporných kazuistik shodu spíše nadhodnocuje, a u dopadu posuzovaného uvnitř jedné etapy, nižší u škál vhodnosti, u orientace reaktivní–proaktivní a u dopadu porovnávaného mezi etapami a nejnižší u jemného schématu typů řešení. Koeficienty se však liší typem, počtem kategorií, vzorky i posuzovatelkami, takže jejich pořadí není samostatným testem náročnosti úsudku (oddíl 8.4). U typu řešení shodu pravděpodobně snižovala jemnost a překryvnost kategorií spolu s jejich nevyváženým zastoupením (kapitola 6). Naznačuje to Jaccardův index 0,56, který započítává i částečný překryv, zatímco α hodnotí každou neshodu stejně.

9. Diskuse

Ve třetí etapě dosáhl druhý pilot na necelé padesátce kazuistik κ nad 0,8 u dopadu i vhodnosti, finální kolo na stovkách kazuistik však $\kappa = 0,69$ a vážené $\kappa = 0,52$. Hodnoty nejsou přímo srovnatelné a rozdíl může souviset s tím, že pilotní shoda se měří krátce po společném rozboru, na malém vzorku a s čerstvou instrukcí, kdežto finální po měsících samostatné práce. Spolehlivost proto kniha vykazuje z finálních kol, a kde má obě hodnoty, uvádí obě. Totéž platí pro druhou etapu, jejíž dvě kalibrační kola byla malá a společně rozebíraná; i zde se kniha opírá jen o finální kolo.

Uprostřed třetí etapy se hodnocení vhodnosti obou anotátorek rozešlo v průměru o půl stupně škály. Po rekaliibraci byl průměrný rozdíl menší, vážená shoda však obdobně nevzrostla; samostatný účinek rekaliibrace z tohoto uspořádání určit nelze (oddíl 8.4).

Z těchto pozorování plynou praktická doporučení pro výzkum s lidskými posuzovateli. V teorii reliability nejsou nová (Krippendorff, 2018; Hallgren, 2012), u dlouhých vícefázových anotačních etap nad pedagogickým textem, v nichž se nastavení škál může posouvat, se však ukázala jako potřebná. Koeficient shody je vhodné uvádět vždy spolu s hrubou shodou a u nevyváženě zastoupených kategorií také s mírou, která je vůči nevyváženosti odolná (paradox shody; Feinstein & Cicchetti, 1990; Gwet, 2008). Spolehlivost je vhodné vykazovat z finálních dat, ne z pilotu. Zásah do protokolu uprostřed sběru, jako byla rekaliibrace, je třeba popsat a kvantifikovat. Projednané případy je vhodné vyloučit z primárního výpočtu, protože po projednání už nejde o nezávislá měření; výsledek je však třeba porovnat s výpočtem na plném souboru, protože vyloučení sporných případů může shodu nadhodnotit (oddíl 8.4). Zpětné přepisy hodnocení je vhodné v datech označovat, aby bylo možné určit, která hodnocení vznikla před zásahem, která po něm a která byla zpětně přepsána. Ve třetí etapě tak šlo odlišit změnu průměrného rozdílu hodnocení od vývoje vážené shody; časová souvislost s rekaliibrací však neznamená, že změnu způsobila.

9.5 Omezení

Omezení knihy plynou z povahy materiálu, z uspořádání studií a z použitých jazykových modelů.

Materiál je korpus dobrovolně psaných kazuistik, které pravděpodobně nadreprezentují dramatické situace (kapitola 5). Dopad posuzuje čtenář podle toho, co pisatel o výsledku napsal, a zrcadlová šablona (kapitola 3) může pisatele vést k volbě jednoznačně vydařených a nevydařených situací. V první etapě tvořily částečné a krátkodobé úspěchy jen menšinu dopadů (oddíl 5.3), ve třetí etapě však anotátorky sloučenou kategorii krátkodobého nebo částečného úspěchu přiřazovaly často (Tabulka 8.1). Zastoupení těchto mezistupňů tedy závisí i na schématu a posuzovateli. Korpus roste a kniha pracuje s datovanými řezy; každé číslo platí pro svůj řez. Část kazuistik je psána slovensky nebo nese anglické zástupné označení žáka; tuto jazykovou vrstvu kniha popisuje jen z komentářů anotátorek (kapitola 8). Kazuistiky, jejichž text byl v minulosti zkrácen při extrakci, kniha z analýz závislých na délce textu vylučuje (kapitola 4). Kódovací schémata se mezi etapami liší a kniha je nesměšuje; jedinou převodní operací je sloučení částečného a krátkodobého úspěchu pro srovnání dopadu mezi etapami. Toto srovnání navíc neproběhlo za stejných podmínek čtení, protože pisatelovo označení kazuistiky jako úspěšné, nebo neúspěšně řešené viděly jen anotátorky první etapy (příloha A.1).

Uspořádání studií je popisné a srovnávací, ne intervenční, a souběhy typů řešení s dopadem proto nezakládají kauzální tvrzení. Zaslepené hodnocení ve Studii 3 mělo chránit před zkreslením očekáváním. Zaslepení však oslabovalo, že každá anotátorka četla k téže situaci obě řešení a že dvě z nich mohly část učitelských řešení znát z první etapy. Instrukce pro generování sice požadovala paritu délky a stylu, nelze však vyloučit, že anotátorky generovaný text rozpoznaly podle jazykových rysů (kapitola 7). Hodnocen byl navíc text, ne jednání ve třídě. Třetí etapu nesly jen dvě anotátorky a rekaliibrace uprostřed sběru byla změnou protokolu. Části etapy před rekaliibrací a po ní se liší i souborem kazuistik, takže samostatný účinek rekaliibrace z jejich srovnání odvodit nelze. Kazuistiky také prošly etapami nerovnoměrně a

část jich prošla několika (oddíl 4.5, příloha C), což omezuje čtení překryvů jako nezávislých vzorků.

Studie 2 a 3 pracovaly každá s jedním modelem a jednou instrukcí (příloha B). Predikční experiment zahrnul ve vývoji více kandidátních modelů a při hodnocení dva modely, včetně varianty s příklady v instrukci. Výsledky nelze zobecnit na jiné instrukce, jazyky, poskytovatele ani budoucí verze modelů. V predikčním srovnání lidé popis výsledku četli, kdežto model jej záměrně nedostal; lidská předpověď dopadu za stejných podmínek, tedy bez znalosti výsledku, k dispozici není.

9.6 Důsledky

Pro přípravu učitelů a školní praxi

Z obrazu repertoáru nejmenšího odporu plyne pro přípravu učitelů podnět pracovat s konkrétními situacemi a vedle zaznamenaného postupu zvažovat alternativy. Z dat však nelze určit, zda rozdíly souvisejí s neznalostí, časovým tlakem, dostupnými prostředky, nebo s vlastnostmi situace. Kazuistiky a generované kontrasty mohou sloužit jako materiál k reflexi; jejich účinek na jednání musí ověřit intervenční studie. Pozorované souvislosti kategorií řešení s kódovaným dopadem neopravňují bez dalšího ověření doporučovat některou strategii jako účinnější.

Pro platformu Edustories

Platforma nabízí generovaná řešení v šesti přístupech (kapitola 3). Studie 3 však použila jednu instrukci bez předepsaného rámce, a není tedy ověřením žádného z generátorů platformy (příloha B.5). Studie 4 měřila lidské posouzení použitelnosti kazuistik, automatický filtr kvality netestovala. Pro platformu z toho plynou návrhy, které je třeba ověřit: kontrola úplnosti kazuistiky před generováním, označení výstupu jako materiálu k reflexi a sběr

zpětné vazby o jeho využití. Spolehlivost případného automatického filtru by bylo nutné vyhodnotit zvlášť.

Pro jazykové modely v pedagogickém výzkumu

Z výsledků knihy plyne pro výzkum s jazykovými modely tento postup: ověřit model na vlastních datech proti lidskému kódování se změřenou shodou, uvádět shodu modelu s lidmi vždy vedle shody lidí mezi sebou, tam, kde model generuje, používat zaslepené lidské hodnocení, a zaznamenávat verze modelů i instrukcí. Doprovodný repozitář (Nehyba, 2026) tento postup zpřístupňuje i s odvozenými daty bez textů kazuistik; hlavní statistické výsledky z nich lze přepočítat bez nového volání modelu (oddíl 4.8).

9.7 Další výzkum

Prvním směrem dalšího výzkumu je intervenční studie, která by ověřila, zda učitelé, kteří k vlastní situaci dostanou generované řešení v jiném rámci, volí v dalších situacích proaktivnější kroky. Korpus a platforma pro ni poskytují materiál. Nástroj pro měření proaktivity dalších kroků by však bylo nutné teprve ověřit a velikost možného účinku tato kniha neodhaduje.

Druhým směrem je spolehlivost posouzení v čase. Ve třetí etapě se ukázalo, že se nastavení škály může během měsíců samostatné práce posunout, aniž se to projeví ve vážené shodě; samostatný účinek rekaliibrace ze srovnání částí etapy určit nelze. Příští etapa by proto měla mít časový záznam každého hodnocení, průběžné kontrolní body a označené přepisy, aby se vývoj shody dal sledovat přímo z dat. Totéž uspořádání dovolí ověřit, zda pilotní shoda předpovídá finální.

Třetím směrem je rozšíření na další modely a jazyky. Opakování Studie 2 a predikčního experimentu s dalšími modely, například provozovanými lokálně nebo od jiných poskytovatelů, by ukázalo, nakolik výsledky závisejí na úloze a nakolik na jednom poskytovateli. Anglická

datová sada Edustories-en (kapitola 3) navíc umožňuje zjišťovat, zda model kóduje tentýž obsah v obou jazycích shodně, a tedy měřit, jak se shoda modelu liší mezi jazyky; vliv jazyka by přitom bylo třeba oddělit od vlivu překladu, jímž anglická sada vznikla.

9.8 Shrnutí

Kniha propojuje popis korpusu, kódování jazykovým modelem, zaslepené hodnocení alternativních řešení a opakované lidské posouzení s předpovědí dopadu. Odlišuje přitom četnosti od příčin volby strategií, shodu od správnosti a hodnocení textu od účinku jednání. Kódování modelem má jen omezenou retrospektivní validaci, generované alternativy byly hodnoceny příznivěji než řešení učitelů a předpověď dopadu zůstává nejistá. Žádný z těchto výsledků nestanovuje obecnou horní mez lidské shody ani informační strop předpovědi. Výklad výsledků omezují i změny protokolu, překryvy s příklady v instrukci, neúplné odpovědi a normativní složka kategorií.

10. Závěr

Kniha vycházela ze dvou propojených otázek: co vypovídají kazuistiky náročného chování o tom, jak čeští učitelé takové chování zvládají, a nakolik lze jejich výzkum ve velkém rozsahu svěřit jazykovému modelu. Zkoumají je čtyři studie nad jedním korpusem projektu Edustories. Tato kapitola shrnuje odpovědi na čtyři výzkumné otázky, pojmenovává přínos knihy a naznačuje, kam práce směřuje.

10.1 Odpovědi na výzkumné otázky

Na první otázku, jak vybudovat a popsat rozsáhlý korpus autentických kazuistik a co zachycuje, odpovídá Studie 1. Korpus vzniká sběrem kazuistik, převážně při pedagogických praxích studujících, podle zrcadlové šablony, která od pisatelů žádá úspěšně i neúspěšně řešenou situaci. Texty procházejí strukturací, anonymizací a redakční kontrolou; zveřejněný řez z května 2026 čítá 3 202 kazuistik. Lidsky kódovaný řez korpusu zachycuje široké spektrum náročného chování od narušování výuky po šikanu a sebedestruktivní chování. V repertoáru řešení dominuje rozhovor a další reaktivní kroky; vztahová a kolektivní řešení se v kódovaném řezu pojí s příznivějším vnímaným výsledkem než řešení represivní. Korpus je zároveň časovou řadou řezů; čísla knihy proto platí vždy pro konkrétní datovaný řez.

Na druhou otázku, nakolik se jazykový model shodne s vyškolenými lidmi při deduktivní klasifikaci obsahu, odpovídá Studie 2 pro typ řešení. Na úplných dvojicích srovnávacího vzorku dosáhla shoda modelu s lidským kódem typu řešení $\kappa = 0,64$; protože vzorek nebyl oddělen

od vývoje instrukce pro model, nejde o nezávislé ověření. Po dodatečném vyloučení kazuistik, jejichž řešení se objevila mezi příklady v instrukci, činí $\kappa = 0,61$. Krippendorffovo α shody mezi anotátorkami, zjištěné v jiném uspořádání, nelze použít jako horní mez shody, které by model mohl dosáhnout. Kategorie přiřazené modelem proto vyžadují průběžnou kontrolu a nenahrazují lidský úsudek nad jednotlivým případem.

Na třetí otázku, jak se liší řešení generovaná modelem od autentických řešení učitelů, odpovídá Studie 3 zaslepeným hodnocením 324 dvojic řešení z kazuistik, které platforma v době výběru vedla v kategoriích fyzické a verbální agrese. Učitelská řešení byla hodnocena jako méně vhodná (rozdíl $-0,85$ bodu škály) a jako reaktivnější a behaviorálnější. Šance, že v nich anotátorky rozpoznají některý z pojmenovaných rámců řešení náročného chování, byla zhruba třetinová oproti řešením generovaným. Výsledek platí pro textovou podobu řešení v těchto kazuistikách, pro jediný model (gpt-4o) a jedinou instrukci. Nedokládá, že „AI radí lépe než učitel“, ale že generované řešení může nabídnout kontrast k zaznamenanému postupu. Úplný repertoár ani znalosti jednotlivých učitelů studie neměřila.

Na čtvrtou otázku odpovídá Studie 4. Na dvojité posouzeném překryvu 693 kazuistik rozlišuje použitelnost textu, popsany dopad a vhodnost řešení. Po vyloučení sporných případů projednaných při rekaliibraci vykazuje posouzení použitelnosti vysokou vzájemnou shodu, na plném překryvu nižší; bez vnějšího kritéria však nelze určit, zda je správné. U dopadu činí $\kappa = 0,69$, u vhodnosti je shoda nižší. Dřívější hodnocení dopadu se s jednou pozdější anotátorkou shodují na úrovni $\kappa = 0,59$; nejde o vnější replikaci. Rozdíly mezi částmi etapy časově souvisejí s rekaliibrací, její samostatný účinek však nelze oddělit od zácvičku a od změny posuzovaných případů.

Při předpovědi dopadu bez jeho popisu dosahuje model kimi-k3 vyvážené přesnosti 60 %, tedy zřetelně nad hádáním většinové třídy, které dává 33 %. Interval párových rozdílů mezi modely i mezi variantou s příklady v instrukci a bez nich zahrnují nulu. Výsledek tedy nedokládá rovnocennost modelů ani to, že by z textu nebylo možné dosáhnout vyšší přesnosti. Shoda lidských posuzovatelek, které popis výsledku četly, je jen kontextovou referencí; jde o jinou úlohu než predikce.

10.2 Přínos knihy

Přínos knihy je trojí. Empirický: kniha popisuje učitelské postupy zaznamenané v rozsáhlém souboru autentických českých situací. Převahu snadno dostupných reakcí označuje interpretačním pojmem repertoár nejmenšího odporu. Neměří však znalosti učitelů ani příčiny volby postupů. Souvislosti řešení s kódovaným úspěchem vycházejí z kategorií dopadu první anotační etapy, které částečně zahrnují i normativní posouzení přijatelnosti řešení; nezávislým měřítkem účinnosti postupů proto nejsou.

Metodologický: jazykové modely byly na jednom korpusu autentických textů v jiném jazyce než angličtině prověřeny ve třech rolích a v každé srovnány s lidským posouzením, jehož shoda byla měřena. Z výsledků vyplývá návrh rozdělení rolí: klasifikace obsahu na úrovni kategorií po ověření na vlastních datech, generování kontrastních řešení jako materiálu k reflexi a posuzování vhodnosti jen s lidskou referencí, protože tuto roli studie přímo netestovaly. Studie 4 zároveň ukázala, že pilotní shoda nemusí odpovídat shodě ve finálním kole a že se během dlouhé etapy může posunout nastavení hodnotící škály, aniž se změní vážená shoda. Proto je vhodné spolehlivost vykazovat z finálních dat, uvádět ji spolu s hrubou shodou a zaznamenat každý zásah do protokolu.

Infrastrukturní: kniha dokumentuje korpus jako časovou řadu datovaných řezů a tři anotační etapy s popsánými schématy a maticí pokrytí. Doprovodný repozitář obsahuje analýzy, tabulky výsledných hodnot, skripty obrázků a odvozená data bez textů kazuistik a bez osobních údajů. Hlavní statistické výsledky lze přepočítat z odvozených dat a uložených predikcí bez opětovného spuštění modelu.

10.3 Směřování

Z výsledků plynou tři směry další práce. Intervenční studie by měla prověřit, zda generované kontrastní řešení nad vlastní situací repertoár učitelů skutečně rozšiřuje. Korpus a platforma Edustories pro ni poskytují materiál; nástroj pro měření takové změny by však bylo nutné

10. Závěr

teprve ověřit a velikost intervenčního účinku tato kniha neodhaduje. Příští anotační etapa by měla zaznamenávat čas hodnocení, mít kontrolní body a označovat změny hodnocení, aby bylo možné vývoj spolehlivosti sledovat přesně. Opakování Studie 2 a predikčního experimentu s dalšími modely i poskytovateli, spolu s anglickou datovou sadou Edustories-en, by ukázalo, nakolik jsou výsledky vlastností úlohy a nakolik vlastností jednoho modelu a jednoho jazyka.

Ve čtyřech studiích se ukázaly možnosti kontrolovaného použití jazykových modelů v pedagogickém výzkumu i hranice dosavadních ověření. Praktická doporučení proto rozlišují shodu, správnost, predikci a pedagogickou účinnost. Pro další použití je podstatné tuto odlišnost zachovat a nehledat jediný koeficient, který by vyjadřoval univerzální hranici.

Literatura

- Acosta, J., Chinman, M., Ebener, P., Malone, P. S., Phillips, A., & Wilks, A. (2019). Evaluation of a whole-school change intervention: Findings from a two-year cluster-randomized trial of the Restorative Practices Intervention. *Journal of Youth and Adolescence*, 48(5), 876–890. <https://doi.org/10.1007/s10964-019-01013-2> ↗
- Adriani, P. A., Hino, P., Taminato, M., Okuno, M. F. P., Santos, O. V., & Fernandes, H. (2024). Non-violent communication as a technology in interpersonal relationships in health work: A scoping review. *BMC Health Services Research*, 24, čl. 289. <https://doi.org/10.1186/s12913-024-10753-2> ↗
- Bendl, S. (2001). *Školní kázeň: Metody a strategie* [School discipline: Methods and strategies]. ISV.
- Bendl, S. (2011). *Kázeňské problémy ve škole* [Discipline problems at school] (2., aktualiz. a dopl. vyd.). Triton.
- Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A. K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J. C., & Durlak, J. A. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development*, 94(5), 1181–1204. <https://doi.org/10.1111/cdev.13968> ↗
- Černochová, M., Selcuk, H., Cohen, A., Voštinár, P., Cohen, G., Azaria Yaakov, A., Jeřábek, T., Beneš, M., & Škrinářová, J. (2025). Generative artificial intelligence (GenAI) in K-12 edu-

cation: National policy perspectives from the Czech Republic, Israel, Latvia, and Slovakia. *Pedagogika*, 75(4), 381–400. <https://doi.org/10.14712/23362189.2025.4924> ↗

Černý, M. (2023, 18. dubna). *ChatGPT ve školní praxi: nástroj oborové didaktiky* [ChatGPT in school practice: A tool for subject didactics]. Metodický portál RVP.CZ. <https://clanky.rvp.cz/clanek/23455/CHATGPT-VE-SKOLNI-PRAXI-NASTROJ-OBOROVE-DIDAKTIKY.html> ↗

Černý, M. (2024). Gramotnost pro umělou inteligenci: koncepty, přístupy a otevřené otázky [Literacy for artificial intelligence: Concepts, approaches and open questions]. *e-Pedagogium*, 24(2), 7–16. <https://doi.org/10.5507/epd.2024.009> ↗

Darling-Hammond, S., Fronius, T. A., Sutherland, H., Guckenburg, S., Petrosino, A., & Hurley, N. (2020). Effectiveness of restorative justice in US K-12 schools: A review of quantitative research. *Contemporary School Psychology*, 24(3), 295–308. <https://doi.org/10.1007/s40688-020-00290-0> ↗

Durlak, J. A., Mahoney, J. L., & Boyle, A. E. (2022). What we know, and what we need to find out about universal, school-based social and emotional learning programs for children and adolescents: A review of meta-analyses and directions for future research. *Psychological Bulletin*, 148(11–12), 765–782. <https://doi.org/10.1037/bul0000383> ↗

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x> ↗

Epinat-Duclos, J., Foncelle, A., Quesque, F., Chabanat, E., Duguet, A., Van der Henst, J.-B., & Rossetti, Y. (2021). Does nonviolent communication education improve empathy in French medical students? *International Journal of Medical Education*, 12, 205–218. <https://doi.org/10.5116/ijme.615e.c507> ↗

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, 43(6), 543–549. <https://doi.org/10.1016/>

[0895-4356\(90\)90158-L ↗](#)

Felcmanová, L., Hečková, L., Myšková, L., Němec, Z., Krejčová, L., Winkler, P., Dvořáková, K., Houška, P., Franke, H., Korbel, M., Kubičková, A., Šimáčková Laurenčíková, K., Nosál, I., Salomonová, M., Smrž, J., Turková, K., Zatloukal, T., Andrys, O., Pražáková, D., ... Borkovcová, I. (2021). *Přístupy k náročnému chování dětí a žáků ve školách a školských zařízeních a možnosti jeho řešení: Metodické doporučení* [Approaches to challenging behaviour of children and pupils in schools and school facilities and options for addressing it: Methodological recommendation]. Česká školní inspekce. https://www.csicr.cz/CSICR/media/Elektronicke-publikace/2021/MD_Pristupy_k_narocnemu_chovani_deti_a_zaku/html5/index.html ↗

Felcmanová, L., Kubičková, A., & Vlachová, P. (2023). *Pozitivní podpora a vhodné reakce na chování žáků: Ověřené pedagogické strategie pro práci s žáky* [Positive support and appropriate responses to pupil behaviour: Evidence-based teaching strategies]. Society for All. <https://www.societyforall.cz/pozitivni-podpora-a-vhodne-reakce-na-chovani-zaku> ↗

Fronius, T., Darling-Hammond, S., Persson, H., Guckenburg, S., Hurley, N., & Petrosino, A. (2019). *Restorative justice in U.S. schools: An updated research review*. WestEd Justice and Prevention Research Center. <https://eric.ed.gov/?id=ED595733> ↗

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30), čl. e2305016120. <https://doi.org/10.1073/pnas.2305016120> ↗

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61(1), 29–48. <https://doi.org/10.1348/000711006X126600> ↗

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023> ↗

Havlíčková, M., Šťastný, V., Boudová, S., & Basl, J. (2025). *Národní zpráva TALIS 2024* [TALIS 2024 national report]. Česká školní inspekce. <https://www.csicr.cz/cz/Dokumenty/Publikace-a-ostatni-vystupy/Narodni-zprava-TALIS-2024> ↗

Honzík, L., Huclová, M., Königsmarková, S., & Vrbová, V. (2025). Některá omezení generativní umělé inteligence při řešení logických problémů [Some limitations of generative artificial intelligence in solving logical problems]. *Pedagogika*, 75(4), 341–353. <https://doi.org/10.14712/23362189.2025.4920> ↗

Horner, R. H., & Sugai, G. (2015). School-wide PBIS: An example of applied behavior analysis implemented at a scale of social importance. *Behavior Analysis in Practice*, 8(1), 80–85. <https://doi.org/10.1007/s40617-015-0045-4> ↗

Karasová, J., & Nehyba, J. (2023). Student-centered teacher responses to student behavior in the classroom: A systematic review. *Frontiers in Education*, 8, čl. 1156530. <https://doi.org/10.3389/feduc.2023.1156530> ↗

Karasová, J., & Nehyba, J. (2025). Novice teachers' classroom behaviour management: Situations, responses and impact on student behaviour. *British Educational Research Journal*, 51(5), 2116–2141. <https://doi.org/10.1002/berj.4166> ↗

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, čl. 102274. <https://doi.org/10.1016/j.lindif.2023.102274> ↗

Kopecký, K., Kříž, M., Krejčí, V., Mašát, M., & Voráč, D. (2025). Využití generativní umělé inteligence jako nástroje podpory učení se zvláštním zaměřením na jazykovou a literární výchovu [Using generative artificial intelligence to support learning with a specific focus on language and literature education]. *Pedagogika*, 75(4), 471–484. <https://doi.org/10.14712/23362189.2025.4911> ↗

- Kopecký, K., Szotkowski, R., Voráč, D., Krejčí, V., & Dobešová, P. (2023). *České školy a umělá inteligence – výzkumná zpráva* [Czech schools and artificial intelligence – research report]. Centrum prevence rizikové virtuální komunikace, Pedagogická fakulta Univerzity Palackého v Olomouci. <https://www.e-bezpeci.cz/index.php/ke-stazeni/vyzkumne-zpravy/163-ceske-skoly-a-umela-inteligence-2023/file>
- Kopecký, K., & Voráč, D. (2024). *Čeští žáci a umělá inteligence – výzkumná zpráva* [Czech pupils and artificial intelligence – research report]. Centrum prevence rizikové virtuální komunikace, Pedagogická fakulta Univerzity Palackého v Olomouci. <https://e-bezpeci.cz/index.php/ke-stazeni/vyzkumne-zpravy/190-cesti-zaci-a-umela-inteligence-2024/file>
- Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, 86(3), 643–680. <https://doi.org/10.3102/0034654315626799>
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4. vyd.). SAGE.
- Kubíčková, A., & Felcmanová, L. (2021). Možnosti podpory duševního zdraví žáků [Opportunities for pupils' mental health support]. *Pedagogická orientace*, 31(1), 70–96. <https://doi.org/10.5817/PedOr2021-1-70>
- Lee, A., & Gage, N. A. (2020). Updating and expanding systematic reviews and meta-analyses on the effects of school-wide positive behavior interventions and supports. *Psychology in the Schools*, 57(5), 783–804. <https://doi.org/10.1002/pits.22336>
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. University of Chicago Press.
- Lukas, J., & Lojdová, K. (2018). Řízení třídy: přístupy, oblasti, strategie [Classroom management: Approaches, areas, strategies]. *Pedagogika*, 68(2), 155–172. <https://doi.org/10.14712/23362189.2017.1180>
- Ministerstvo školství, mládeže a tělovýchovy. (2010). *Metodické doporučení k primární prevenci rizikového chování u dětí, žáků a studentů ve školách a školských zařízeních* [Me-

methodological recommendation on primary prevention of risk behaviour among children, pupils and students in schools and school facilities] (č. j. 21291/2010-28). <https://msmt.gov.cz/vzdelavani/socialni-programy/metodicke-dokumenty-doporuceni-a-pokyny> ↗

Ministerstvo školství, mládeže a tělovýchovy. (2023). *Rámcový vzdělávací program pro základní vzdělávání* [Framework educational programme for basic education]. https://www.edu.cz/wp-content/uploads/2023/07/RVP_ZV_2023_cista_verze.pdf ↗

Ministerstvo školství, mládeže a tělovýchovy. (2025). *Rámcový vzdělávací program pro základní vzdělávání* [Framework educational programme for basic education] (revidované znění, opatření ministra č. j. MSMT-20704/2024-5). <https://prohlednout.rvp.cz/zakladni-vzdelavani> ↗

Národní pedagogický institut ČR. (2023). *Doporučení, jak pracovat s umělou inteligencí* [Recommendations on how to work with artificial intelligence]. NPI ČR. <https://www.npi.cz/aktuality/74837-doporuceni-jak-pracovat-s-umelou-inteligenci> ↗

Nehyba, J. (2026). *Edustories: doprovodný repozitář habilitační monografie* [Edustories: Companion repository of the habilitation monograph] [Datová sada a kód]. GitHub. <https://github.com/JanNehyba/edustories-habilitation> ↗

Nehyba, J., Karasová, J., Škarková, L., Fico, M., Košatková, M., Košatka, D., & Štefánik, M. (2024). Edustories: analýza kazuistik o náročném chování žáků [Edustories: An analysis of case studies on challenging student behaviour]. In A. Rozkovicová & J. Pivarč (Eds.), *Spolupráce oborů pro podporu kvality vzdělávání: Sborník příspěvků z XXXII. výroční konference České asociace pedagogického výzkumu* (s. 256–260). MSD.

Nehyba, J., Karasová, J., & Štefánik, M. (2024). *Edustories-en* [Datová sada]. Hugging Face. <https://huggingface.co/datasets/MU-NLPC/Edustories-en> ↗

Nehyba, J., & Štefánik, M. (2023). Applications of deep language models for reflective writings. *Education and Information Technologies*, 28(3), 2961–2999. <https://doi.org/10.1007/s10639-022-11254-7> ↗

- OECD. (2025). *Results from TALIS 2024: The state of teaching*. OECD Publishing. <https://doi.org/10.1787/90df6235-en> ↗
- Ollion, É., Shen, R., Macanovic, A., & Chatelain, A. (2024). The dangers of using proprietary LLMs for research. *Nature Machine Intelligence*, 6(1), 4–5. <https://doi.org/10.1038/s42256-023-00783-6> ↗
- Payne, G., & Hustler, D. (1980). Teaching the class: The practical management of a cohort. *British Journal of Sociology of Education*, 1(1), 49–66. <https://doi.org/10.1080/0142569800010104> ↗
- Reiss, M. V. (2023). *Testing the reliability of ChatGPT for text annotation and classification: A cautionary remark* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2304.11085> ↗
- Rosenberg, M. B. (2015). *Nonviolent communication: A language of life* (3. vyd.). PuddleDancer Press.
- Rosenberg, M. B. (2022). *Nenásilná komunikace: Řeč života* [Nonviolent communication: A language of life] (N. Garciová & M. Záleta, Přel.; 5., rev. a dopl. vyd.). Portál.
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, 31(3), 351–380. <https://doi.org/10.1353/etc.0.0007> ↗
- Swidler, A. (1986). Culture in action: Symbols and strategies. *American Sociological Review*, 51(2), 273–286. <https://doi.org/10.2307/2095521> ↗
- Šedřová, K., Švaříček, R., & Šalamounová, Z. (2012). *Komunikace ve školní třídě* [Communication in the school classroom]. Portál.
- Ševčíková, J., Buchtová, T., & Plischke, J. (2023). Nekázeň žáků v reflexích začínajících učitelů v České republice (a možnosti jejich ovlivnění) [Pupil indiscipline in the reflections of novice teachers in the Czech Republic (and options for influencing them)]. *Paidagogos*, (1–2), čl. 1. <https://paidagogos.net/issues/2023/1-2/article.php?id=1> ↗

Törnberg, P. (2025). Large language models outperform expert coders and supervised classifiers at annotating political social media messages. *Social Science Computer Review*, 43(6), 1181–1195. <https://doi.org/10.1177/08944393241286471> ✓

Valenta, J. (2006). *Osobnostní a sociální výchova a její cesty k žákovi* [Personal and social education and its paths to the pupil]. AISIS.

Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research*, 54(2), 143–178. <https://doi.org/10.3102/00346543054002143> ✓

Vlčková, K., Lojdová, K., Lukas, J., Mareš, J., Škarková, L., Kohoutek, T., Květon, P., & Ježek, S. (2020). *Řízení třídy: studenti učitelství a jejich provázející učitelé* [Classroom management: Student teachers and their mentor teachers]. Masarykova univerzita. <https://doi.org/10.5817/CZ.MUNI.M210-9652-2020> ✓

Vyhláška č. 27/2016 Sb., o vzdělávání žáků se speciálními vzdělávacími potřebami a žáků nadaných. (2016). *Sbírka zákonů České republiky*, částka 10. <https://www.zakonyprolidi.cz/cs/2016-27> ✓

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370> ✓

Zákon č. 82/2015 Sb., kterým se mění zákon č. 561/2004 Sb., o předškolním, základním, středním, vyšším odborném a jiném vzdělávání (školský zákon), ve znění pozdějších předpisů, a některé další zákony. (2015). *Sbírka zákonů České republiky*, částka 37. <https://www.zakonyprolidi.cz/cs/2015-82> ✓

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), *Advances in Neural Information Processing Systems* 36 (s. 46595–46623). Curran Associates. <https://doi.org/10.52202/075280-2020> ✓

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2024). Can large language models transform computational social science? *Computational Linguistics*, 50(1), 237–291.
https://doi.org/10.1162/coli_a_00502 ↗

Příloha A. Kódovací manuály

Příloha uvádí kódovací nástroje všech tří anotačních etap v podobě, v jaké je anotátorky používaly; definice kategorií a škálových bodů citujeme v původním znění. Vývoj nástrojů mezi verzemi popisujeme výslovně, protože se verze schémat v analýzách nesměšují (kapitola 4).

A.1 První etapa: kódovací kniha obsahového kódování

Kódovací kniha vznikala induktivně: v zácvičné fázi (zima 2023/2024) anotátorky otevřeně kodovaly menší vzorky kazuistik do vlastních listů a z těchto kódů byl postupně sestaven kódovací strom (verze V1), který byl po kalibračních kolech zpřesněn do finální verze V2. Verze V2 přidala anotační kódy (zkratky kategorií), upřesnila definice a část kategorií V1 zrušila nebo převedla mezi podřazené kódy: u typů chování „Neuposlechnutí pokynu“, „Stranění se kolektivu“, „Opouštění třídy“ a „Užívání návykových látek“, u řešení „Opouštění třídy“, „Zjišťování“ a „Proaktivní řešení“. Definice neúspěchu ve V2 nově zahrnuje i použití nepřijatelných prostředků. Finální anotování (kapitola 4) probíhalo podle V2 se závazným kódovacím postupem:

1. Při kódování kódujeme pouze to, co je explicitně napsáno, snažíme se vystihnout jen to, co přímo souvisí s problémem/řešením. [...]
2. U kódů k řešení situací je to stejné - kóduji pouze to, co učitel explicitně zmiňuje, že použil jako konkrétní řešení (ne, co by si myslel, že by fungovalo; ne co se stalo o pár měsíců později).

3. Pozor na to, aby byly v kategoriích napsány přesné názvy kódů.
4. Pozor na to, jak zní definice kategorie, kterou ke kódu dáváte. Vždy si jí dobře přečtěte a zvažte, jestli i ta definice sedí s tím, co je v textu.

[...]

1. Flat/deep - budeme kódovat jako flat, pokud v kazuistice chybí nějaká část, pokud kazuistika z nějakého jiného důvodu nejde použít (je příliš stručná, není ze školního kontextu, nejedná se o popis problémové situace atd.). [...]
2. Výstup je správný - Budete zaškrtnávat Ano/ne - podle toho, jestli ve vedlejším sloupci ++ a -- sedí na danou kazuistiku. ++ znamená, že je kazuistika alespoň do nějaké míry úspěšně vyřešená. -- Znamená, že kazuistika není úspěšně vyřešená.
 - Tzn. ve chvíli, kdy jsem okódovala kazuistiku jako Dlouhodobý úspěch/krátkodobý úspěch nebo částečný úspěch a mám tam u ní ++, pak je výstup správný. Ve chvíli, kdy kóduji kazuistiku jako neúspěch a je tam ++, pak výstup správný není, atd. Hodnoty ++ a -- nijak nepřepisujeme, necháváme je tak, jak jsou.
3. Co všechno je ještě neúspěch - kazuistika není úspěšně vyřešená i v případě, že učitel použil nevhodné strategie, čímž teď myslím strategie typu násilí, šikana, ... něco, co nepatří do hodnot vzdělávacího systému. V takových případech kódujte rovnou jako "Neúspěch" a v případě, že u ní bude ++, tak píšeme, že výstup správný není, protože to má být --.
 - Doplnila jsem tento popis i do kódovacího stromu: Neúspěch: Chování žáka nebo daná situace se nijak nezměnila. Nebo učitel použil k vyřešení situace neakceptovatelné a nevhodné způsoby (fyzické násilí, psychické násilí, nerespektující strategie) - ne v souladu

s hodnotami chováním učitele.

Poznámka. Kódovací strom první etapy, list Kódovací postup.

Anotátorky první etapy tedy při kódování dopadu viděly pisatelovo označení kazuistiky jako úspěšně (++) , nebo neúspěšně (--) řešené a ověřovaly, zda s ním jejich kód souhlasí; v listech třetí etapy toto označení nebylo. To je třeba vzít v úvahu při čtení shody dopadu v první etapě i při srovnání dopadu mezi etapami (oddíly 6.3 a 8.4).

Tabulka A.1. Analyticky použité kategorie typů náročného chování (finální verze V2).

Kategorie	Kód	Definice
Fyzická agrese	FAG	Žák/žákyně fyzicky napadá/poškozuje/ubližuje nebo hrozí ublížením spolužákům, učiteli, nebo ničí majetek školy nebo věci spolužáků/učitele (hází věcmi, rozbíjí věci, vyhrožuje nožem, kope do věcí, pere se, bije,...)
Verbální agrese	VAG	Žák záměrně používá verbální výraz/y, které jsou vulgární nebo mají za cíl ublížit (ponížit,...) učiteli/spolužákům.
Šikana	SIK	Dlouhodobé záměrné verbální nebo neverbální ubližování spolužákovi/spolužačce/učiteli.

Kategorie	Kód	Definice
Emoční výbuchy	EV	Afektivní - Vyafektované chování, krátkodobá silná a prudká emoční reakce na podnět při sníženém sebeovládání. Žák nezvládne nápor emocí. Může se to projevit tak, že je vybíjí agresivním chováním ven (udeří spolužáka, hází předměty, ničení nábytku školy,...)
Lhaní a podvody	LH	Žák/žákyně záměrně poskytuje nepravdivé informace učiteli/rodičům.
Nevěnování se výuce/Nepozornost při výuce	NEV	Žák/žákyně se nevěnuje výuce - nedává pozor, není aktivní, běhá po třídě, ignoruje pokyny učitele, spí, dívá se z okna, atd.
Neplnění školních povinností/nepřipravenost na výuku	NEP	Žák/žákyně neplní své školní povinnosti (učení se, nošení pomůcek,...)
Nezvládání výuky	NEZ	Dítěti brání absence dovednosti zvládat výuku.
Verbální narušování výuky	VER	Žák/žákyně narušuje hodinu verbálními projevy.

Kategorie	Kód	Definice
Porušování třídních/školních pravidel	PRA	Žák porušuje třídní nebo školní pravidla
Neverbální narušování výuky	NEVER	Žák/žákyně narušuje vyučování neverbálními chováními (interakcí s věcmi, hlučnými pohyby,...)
Sebedestruktivní chování	SEB	Žák/žákyně si záměrně ubližuje/hrozí ublížením.
Problémy s docházkou	DOC	Dívka chodila za školu.

Poznámka. Kódovací strom první etapy, list Problémové chování.

Kategorii Problémy s docházkou vymezuje strom obecněji u podřazeného kódu Nechození do školy: „Opakovaná nepřítomnost žáka na hodinách. Žák se záměrně vyhýbá škole a nedochází do ní, chodí za školu.“ Strom obsahoval ještě položku DIA, jejíž definice popisovala přítomnost diagnózy žáka, nikoli jeho chování. V deskriptivě i výpočtu shody je proto vyřazena a údaj o diagnóze se čte z kurátorovaného kontextového pole korpusu (oddíly 5.3, 5.4 a 6.2).

Tabulka A.2. Kategorie typů řešení (finální verze V2).

Kategorie	Kód	Definice
Rozhovor	ROZ	Řešení problému rozhovorem (s rodiči, se žákem, s třídou,...). Cíleně vedený rozhovor za účelem zjištění více informací a řešení dané situace.

Kategorie	Kód	Definice
Dohoda	DOH	Dohoda učitele s žákem/třídou na společně přijatelném řešení.
Práce s kolektivem	KOL	Učitel řeší situaci skrze podporu vztahů ve třídě, čímž se snaží situaci vyřešit a zároveň předcházet dalším
Spolupráce s odborníky	SPO	Učitel vyhledává pomoc jinde - mezi odborníky v rámci školy i mimo školu.
Upozornění	UPO	Učitel popisuje žákovo chování, které je pro něj rušivé. Žák je verbálně upozorněn na rušivé chování.
Přemístění žáka	PRE	Učitel řeší situaci tak, že žák/žákyně opouští danou situaci a je přemístěn(a) v rámci třídy, školy nebo systému.
Nerespektující komunikace	NER	Učitel volí formu reakce skrze verbální komunikační bloky, např. vyhrožování, křik, oplácení chování, rozkaz...

Kategorie	Kód	Definice
(Kázeňské) Tresty	TRE	Učitel řeší situaci formou trestu, který není přímým důsledkem žákova chování. Příčina-následek není - trest nesouvisí přímo s chováním žáka. Učitel pojmenovává zásah jako trestající.
Důsledky	DUS	Žák/žákyně plní přirozené důsledky, které vyplývají z jeho chování. Učitel poskytl informaci, která seznamuje žáka s tím, že toto je důsledek, který se stane, když bude v nekázní pokračovat. (sociálně vyjednané)
Fyzické zakročení	FYZ	Učitel fyzicky reaguje na chování žáka.
Podpora	POD	Žák/žákyně dostává speciální formu individuální podpory, např. skrze asitenta nebo IVP.

Poznámka. Kódovací strom první etapy, list Řešení situací.

Poznámka ke kategorii Proaktivní řešení. Ve verzi V1 byla „Proaktivní řešení“ samostatnou kategorií s definicí: „Učitel volí strategie, kterými předchází dalšímu potenciálnímu náročnému chování žáků a které vrací pozornost žáků směrem k výuce.“ Ve verzi V2 byla

přesunuta mezi podřazené kódy kategorie Podpora. Do klasifikační instrukce pro jazykový model (příloha B) přesto vstoupila jako samostatná položka s vlastní definicí, a model ji proto v korpusu přiřazuje (kapitola 6). V lidském kódování první etapy ji jako samostatnou kategorií měla ve výběrovém seznamu jen jedna anotátorka. V korpusové deskriptivě (kapitola 5) ji proto mezi lidské kategorie řešení nezařazujeme (kmenová sada řešení má 11 položek); v porovnání s modelem (kapitola 6) ji ponecháváme jako kategorii predikovanou modelem. V datech se ojediněle vyskytují i štítky mimo strom (například „humor“) a překlepové varianty; deskriptivní analýzy je převádějí na kanonické znění (kapitola 5).

Tabulka A.3. Kategorie dopadu řešení (finální verze V2).

Kategorie	Kód	Definice
Neúspěch	NEU	Chování žáka nebo daná situace se nijak nezměnila. Nebo učitel použil k vyřešení situace neakceptovatelné a nevhodné způsoby (fyzické násilí, psychické násilí, nerespektující strategie) - ne v souladu s hodnotami chováním učitele.
Částečný úspěch	CU	Chování žáka nebo daná situace se změnila jen částečně.
Krátkodobý úspěch	KU	Situace se změnila pouze na krátkou dobu a poté se chování začalo opět opakovat.
Dlouhodobý úspěch	DU	Situaci se podařilo vyřešit v danou chvíli a chování/situace se již neopakovala.

Poznámka. Kódovací strom první etapy, list Dopad řešení.

Kategorie NEU spojuje popsané nezměněné chování s hodnotovým odmítnutím prostředku a z výsledného kódu nelze určit, který z těchto důvodů rozhodl. Ani ve třetí etapě není plo-

chost pouze technickou neúplností: zahrnuje také eticky nepřijatelné jednání. S těmito konstrukčními omezeními počítají kapitoly 4, 5 a 8.

A.2 Druhá etapa: manuál hodnocení kvality řešení

Manuál druhé etapy definuje pojmy tří bipolárních dimenzí, čtyři pojmenované rámce řešení náročného chování a škálové body. Anotátorky hodnotily každé řešení (učitelské i generované, zaslepeně; kapitola 7) na čtyřech škálách a označovaly rozpoznané rámce.

Tabulka A.4. Vymezení pólů bipolárních dimenzí.

Pojem	Definice
Reaktivní řešení	Okamžitá reakce na aktuální problém či konflikt ve třídě. Cílem je rychle obnovit řád a minimalizovat negativní dopady na průběh výuky.
Proaktivní řešení	Preventivní opatření a postupy, které předcházejí vzniku problémového chování či náročných situací. Učitel je plánuje a zavádí dopředu, aby žáky učil principy prosociálního chování a aby ve třídě vytvořil pozitivní prostředí a jasná pravidla.
Systémové řešení	Komplexní rámec, který platí pro celou třídu nebo školu. Přináší jasně definované zásady, postupy a jednotné přístupy ke zvládání chování, na nichž se podílí vedení, učitelé i žáci (využívá další aktéry jako školní psycholog, metodik prevence, školní programy, rodiče...)

Pojem	Definice
Situační řešení	Okamžitá a flexibilní reakce v rámci konkrétní situace. Učitel volí postup podle aktuálních podmínek a potřeb žáků s cílem minimalizovat narušení výuky a vyřešit konflikt nebo nežádoucí chování na místě.
Humanisticky orientované	V centru pozornosti je žák (jeho potřeby, motivace, emoční stav). Učitel buduje vztah založený na empatii, respektu a sebereflexi. Učitel vědomě pěstuje respektující a empatický přístup, dává žákovi prostor vyjádřit se, vede ho k uvědomění si dopadu svého chování a společně s ním hledá možná řešení.
Behaviorálně orientované	Vychází z behaviorálních teorií učení, kde se posiluje žádoucí chování nebo potlačuje nežádoucí chování pomocí odměn (pochval, výhod) či trestů (sankcí). Řešení je jasně nastavené (učitel se rozhodnul a řešení nastavil), nebere ohledy na pocity/potřeby žáků, je direktivní směrem od učitele.

Poznámka. Kódovací manuál druhé etapy, list Popis definic.

Škály. Celková vhodnost byla ve formuláři zadávána jako pětibodová (1 – Zcela nevhodné, 2 – Spíše nevhodné, 3 – Vyvážené, 4 – Spíše vhodné, 5 – Zcela vhodné); v analyzovaném souboru je lineárně překódována na –2 až +2 jako ostatní dimenze (kapitola 7). Tři bipolární dimenze užívají pětibodovou škálu –2 až +2, jejíž krajní a střední body manuál definuje takto (výběr):

Reaktivní vs. Proaktivní

- 2 - Čistě reaktivní: „Řešení se zaměřuje výhradně na okamžitou reakci na vzniklý problém či konflikt, aniž by zahrnovalo preventivní opatření či dlouhodobou vizi.“
- 0 - Vyvážené: „Řešení kombinuje okamžitou reakci s preventivními kroky...”
- +2 - Čistě proaktivní: „Řešení je plně orientované na prevenci – učitel plánuje a zavádí opatření před vznikem problému...”

Systémové vs. Situační (konverzační)

- 2 - Plně systémové: „Řešení vychází z jednotného, jasně definovaného rámce a pravidel, platných pro celou třídu nebo školu...”
- +2 - Plně situační (konverzační): „Řešení je zcela situační a flexibilní, přizpůsobuje se okamžitým potřebám a podmínkám, aniž by se opíralo o přísně daný systém.”

Humanisticky vs. Behaviorálně orientované

- 2 - Plně humanisticky orientované: „Řešení se zcela soustředí na individuální potřeby, motivaci a emoční stav žáků...”
- +2 - Plně behaviorálně orientované: „Řešení je zcela orientované na behaviorální techniky, využívá jasně definovaných odměn a sankcí...”

Poznámka. Kódovací manuál druhé etapy, list se škálami. Orientace pólů (záporný pól = reaktivní/systémové/humanistické) odpovídá znaménkům výsledků v kapitole 7.

Rámce. Manuál definuje čtyři pojmenované rámce, které anotátorky rozpoznávaly ve znění řešení: PBIS (pozitivní behaviorální intervence a podpora; systémový přístup s jasnými pravidly, konzistentní pozitivní zpětnou vazbou a předem určenými kroky při porušení pravidel), SEL (sociálně-emoční učení; dlouhodobý rozvoj emočních a vztahových dovedností, sebe-regulace a odpovědného rozhodování), NVC (nenásilná komunikace; respektující dialog ve čtyřech krocích pozorování, pocity, potřeby, prosba) a restorativní praxe (v manuálu „resto-

rativní spravedlnost“; náprava škody a obnovení vztahů namísto trestu). Teoretické ukotvení rámců podává kapitola 2.

A.3 Třetí etapa: operacionalizace kvality kazuistik

Třetí etapa posuzovala kvalitu kazuistiky, dopad řešení a vhodnost řešení. Nástroje byly definovány takto:

Tabulka A.5. Kvalita kazuistiky (binární posouzení).

Kategorie	Definice
Plochá (nepoužitelná)	Kazuistika neposkytuje dostatek informací pro pochopení situace ani pro interpretaci řešení. Nebo učitel použil k vyřešení situace neakceptovatelné a nevhodné způsoby (fyzické násilí, psychické násilí, vysoce nerespektující strategie) - ne v souladu s hodnotami chováním učitele a právy dítěte.
Použitelná (dostatečná - skvělá)	Kazuistika poskytuje dostatek informací pro základní pochopení situace a pro interpretaci řešení.

Poznámka. Kódovací manuál třetí etapy. V anotačních datech je kladná kategorie zapisována štítkem „Plná“; v textu proto užíváme označení plná a plochá.

Tabulka A.6. Dopad řešení ve třetí etapě.

Kategorie	Kód	Definice
Neúspěch	NEU	Chování žáka nebo daná situace se nijak nezměnila - ne k lepšímu.
Krátkodobý/částečný úspěch	KU	Situace se změnila pouze na krátkou dobu a nebo se změnila pouze část situace, případně se chování začalo opět opakovat.
Dlouhodobý úspěch	DU	Situaci se podařilo vyřešit v danou chvíli a chování/situace se již neopakovala.

Poznámka. Kódovací manuál třetí etapy.

Dopad řešení je ve třetí etapě tříkategorální. Oproti první etapě (Tabulka A.3) jsou kategorie Částečný a Krátkodobý úspěch sloučeny a definice neúspěchu nezahrnuje hodnotové odmítnutí prostředku; to je ve třetí etapě součástí definice ploché kazuistiky (Tabulka A.5). Dopady z obou etap proto nelze přímo slučovat bez explicitního převodu (kapitola 4).

Tabulka A.7. Vhodnost řešení ve třetí etapě (čtyřbodová škála bez středního bodu).

Stupeň	Definice
1 – Zcela nevhodné	Řešení vůbec nereflektuje specifické potřeby žáka. Je příliš náročné nebo neadekvátní, což může vést k dalším potížím či zhoršení situace.

Stupeň	Definice
2 – Spíše nevhodné	Řešení obsahuje zásadní nedostatky – přestože se snaží reagovat na problém, chybí mu dostatečná přizpůsobivost či citlivost vůči individuálním potřebám žáka, což snižuje jeho efektivitu.
3 – Spíše vhodné	Řešení je většinou přiměřené a respektuje individuální potřeby žáka. Nabízí jak preventivní, tak reaktivní přístupy a většinou dobře vyvažuje náročnost s podporou, byť s drobnými výhradami.
4 – Zcela vhodné	Řešení je ideálně sladěno s požadavky dané situace. Perfektně kombinuje prevenci a adekvátní reakci, respektuje individuální charakteristiky žáka a přispívá k pozitivnímu rozvoji.

Poznámka. Kódovací manuál třetí etapy. S pětibodovou vhodností druhé etapy se tato škála nesměšuje (kapitola 4).

Vedle kategorií anotátorky vyplňovaly tři textové komentáře (jazykový komentář, komentář k plochosti s povinným zdůvodněním, další komentář).

Postup etapy a pilotní shoda. Manuál popisuje plánovaný postup etapy a shodu z druhé pilotáže takto:

Budeme kódovat takto:

- děláme zaškolení na 5-10 kazuistikách (jen vysvětlení pojmů, nácvik kódování)
- poté budou anotátorky kódovat 20 kazuistik. Projdeme společně každou z nich a nesrovnalosti vyjasníme a dle toho upravíme kódovací manuál (viz

- operacionalizace pojmů výše)
- poté budou anotátorky kódovat 50 kazuistik - spočte se shoda [...]. Shoda vyšla ve všech případech nad 80%, někde i nad 90% [...]
- kóduje se 1950 kazuistik, z toho 695 je překryv pro výpočet shody
- poté se opět spočte shoda [...]

Výsledky druhé pilotáže zaznamenává manuál slovensky:

1. Sloupec: Kvalita kazuistiky:

Zhoda: 97.96%

[...]

2. Sloupec: Dopad řešení:

Zhoda: 93.88%

Cohen's kappa: 0.8960

3. Sloupec: Vhodnost:

Zhoda: 89.80%

Cohen's kappa: 0.8131

Poznámka. Kódovací manuál třetí etapy. Jde o pilotní hodnoty z přípravy etapy; shodu na finálních datech uvádí Studie 4 (kapitola 8).

Příloha B. Prompty jazykových modelů

Příloha uvádí znění promptů (zadání neboli instrukcí, které jazykový model dostává) u kroků zpracování, na nichž stojí analýzy knihy, spolu s použitým modelem. Oddíl B.5 navíc vymezuje generovaná řešení platformy, z nichž analýzy knihy nevycházejí. Postup reprodukce popisuje oddíl 4.8.

Tabulka B.1. Kroky zpracování s jazykovými modely a zdroje znění promptů.

Krok zpracování	Model	Zdroj znění	Oddíl
Extrakce a strukturace kazuistik	gpt-4o (dříve gpt-4-1106-preview)	extrakční notebook projektu	B.1
Anonymizace (třetí stupeň)	součást promptu extrakce	tamtéž	B.2
Klasifikace řešení (model v roli anotátora)	gpt-4o	klasifikační notebook projektu	B.3
Generace srovnávacího řešení (druhá etapa)	gpt-4o	rukopis článku o srovnání řešení a návrh studie	B.4
Rámcovaná řešení platformy	jazykový model platformy	kód platformy	B.5

Krok zpracování	Model	Zdroj znění	Oddíl
Predikce dopadu řešení jazykovým modelem	kimik3 a glm-5.3 (e-INFRA CZ)	skript predikčního experimentu v doprovodném repozitáři (Nehyba, 2026)	B.6

B.1 Extrakce a strukturace kazuistik

Jazykový model dělí odevzdané texty do čtyř sekcí podle strukturačního promptu, který zároveň obsahuje třetí stupeň anonymizace (oddíl B.2). Strukturaci provádí model gpt-4o s teplotou nula; totéž znění se dříve používalo s modelem gpt-4-1106-preview s pojistkou opakování (tři pokusy, práh délky výstupu, mírné zvýšení teploty). Vstupem je text po strojové anonymizaci.

System:

- Jsi expert na univerzitě, který má za úkol detailně přepsat slovo od slova a následně upravit texty (kazuistiky) popisující problémové chování žáků na škole.
- Uživatel (User) ti poskytne celý text kazuistiky.

User:

- Doslovně transkribuj tento vstupní text „{text}“ kazuistiky tak, aby odpovídal původnímu obsahu co nejpřesněji, a poté tento text rozděl do čtyř specifických sekcí podle následujících kritérií:
 - V sekci 'description' umístí text od začátku 'Podrobného popisu (vzniku) situace' až do 'Anamnéza'.
 - V sekci 'anamnesis' umístí text od 'Anamnéza' až do 'popis řešení'.

- V sekci 'solution' umístí text popisující řešení situace, začínající od 'popis řešení' až do 'Výsledek'.
- V sekci 'outcome' umístí text popisující dlouhodobý výsledek od 'Výsledek' až do konce narativního popisu situace.
- Pokud vstupní text kazuistiky neobsahuje rozdělení na Podrobný popis (vzniku) situace, Anamnéza a podobně, pak veškerý vstupní text pouze rozděl podle svého nejlepšího uvážení do jednotlivých sekcí.
- V sekcích odstraň úvodní zadání jako: "Podrobný popis (vzniku) situace na úrovni chování", "Anamnéza žáka/žáků nebo třídy (max. 2 normostrany, chronologicky)", "Podrobný popis řešení problematického chování", "Výsledek řešení (krátkodobě-jak to probíhalo hned po incidentu a dlouhodobě-zda se to nějak odrazilo v dalších hodinách, chronologicky, max. 2 normostrany)".
- Každou sekci začni smysluplným slovem, kterým obvykle začínají věty v českém jazyce, například jako "Žák", "Učitel", "Žáci", "Situace" a podobně.
- Ve výstupu v sekcích použij maximum doslovných citací ze vstupního textu kazuistiky.
- Zkontroluj, že jsi transkriboval (transcribe) veškerý vstupní text kazuistiky, pokud ne, tak jej doplň o chybějící části textu.
- Celkově by měly mít všechny sekce dohromady stejný počet tokenů jako vstupní text kazuistiky.
- Svůj výstup poskytni v tomto formátu: ("description": "", "anamnesis": "", "solution": "", "outcome": ""), každá sekce je na novém řádku.
- Z výstupu a citací odstraň pouze všechna jména, příjmení.
- Z výstupu a citací odstraň pouze zkratky jmen a příjmení (například jako: H., J., K., P., M., M.B., a podobně), tak aby text byl stále smysluplný popis situace.
- [ANONYMIZED] často označuje nějaké jméno nebo příjmení nebo město,

ulici, nahraď tuto frázi obecným slovem jako například osobní zájmeno nebo žák, žačka a podobně.

- Ve výstupu nesmí být nikde fráze [ANONYMIZED] a fráze "Do jaké míry".
- Při řešení úkolu postupuj krok za krokem.
- Když se budeš držet těchto instrukcí, tak budeš pochválený a dostaneš nepředstavitelnou odměnu.

Poznámka. Extrakční notebook projektu.

B.2 Anonymizace v promptu

Úplný trojstupňový anonymizační protokol i schválení etickou komisí uvádí příloha D.2; zde uvádíme jeho část, která se týká promptů. Druhý stupeň není prompt, nýbrž kód: rozpoznávání pojmenovaných entit službou NameTag (ÚFAL, model Czech Named Entity Corpus 2.0) nahrazuje entity typů osobních jmen, příjmení, měst, ulic a dalších (typy „pf“, „pp“, „pc“, „gc“, „pm“, „ps“, „gs“, „gu“) řetězcem [ANONYMIZED]. Třetí stupeň je vestavěn přímo do strukturačního promptu (B.1): instrukce ukládá jazykovému modelu odstranit jména a jejich zkratky a nahradit zbylé značky [ANONYMIZED] obecnými výrazy.

B.3 Klasifikace řešení jazykovým modelem (model v roli anotátora)

Deduktivní klasifikaci typů řešení (Studie 2) provedl model gpt-4o s teplotou nula. Tato varianta promptu byla vybrána z několika pilotovaných podle shody s lidským kódováním týchž kazuistik, z nichž je sestaven srovnávací soubor; srovnání s lidským kódováním proto není nezávislým testem (kapitola 6). Vstupem bylo pouze znění řešení; prompt obsahuje vlastní zkrácená znění definic kategorií, která se od kódovací knihy (příloha A) v některých kategoriích liší, a ke každé kategorii několik doslovných příkladů. Srovnávací soubor s lidským kódováním čítá 334 párů a pracuje s 12 kategoriemi (kapitola 6).

System:

- Jsi expert na klasifikaci textu.
- Uživatel (User) ti poskytne popis navrženého řešení kazuistiky o náročném chování žáků.
- Tvým úkolem je přesně klasifikovat řešení kazuistiky podle předdefinovaných kategorií, které poskytne uživatel.

User:

- Analyzuj situaci krok za krokem (Let's Think Step by Step).
- Analyzuj podrobně řešení nevhodného chování žáka, žáků: ###{solution}### a přiřaď nejlepší možnou následující kategorii k řešení z kazuistiky.
- K přiřazování použij pouze tyto kategorie:
 - Rozhovor: Řešení chování pouze pomocí rozhovoru s různými účastníky (žák, rodiče, třída, učitelé, vedení školy).
 - Dohoda: Dohoda nebo kompromis mezi učitelem a žákem/třídou na společně přijatelném řešení.
 - Proaktivní řešení: Strategie učitele zaměřené na předcházení náročnému chování a zaměření pozornosti žáků na výuku.
 - Práce s kolektivem: Podpora vztahů ve třídě jako prostředek k řešení situace a prevenci dalších problémů.
 - Spolupráce s odborníky: Učitel vyhledává pomoc mezi odborníky (ve škole i mimo školu) při řešení situací (např. speciální pedagog, psycholog, policie).
 - Upozornění: Učitel verbálně upozorňuje žáka na jeho rušivé chování.
 - Přemístění žáka: Řešení problému přemístěním žáka v rámci třídy, školy nebo systému (např. diagnostický ústav, psychiatrická léčebna).
 - Nerespektující komunikace: Učitel reaguje pouze skrze komunikační bloky jako je křik, vyhrožování nebo oplácení chování.
 - (Kázeňské) Tresty: Učitel uvaluje tresty, které nejsou přímým

- důsledkem chování žáka (např. práce navíc, domácí úkoly).
- Důsledky: Přirozené důsledky vyplývající z chování žáka, o kterých je informován (např. opakování ročníku, poznámka, snížená známka z chování).
 - Fyzické zakročení: Učitel fyzicky reaguje na chování žáka (např. facka, znehybnění).
 - Podpora: Individuální podpora žáka zaměřená na jeho motivaci, klidnění nebo zlepšení výkonu (např. odměny, pochvaly, doučování, IVP, asistent).
- Klasifikuj řešení pouze podle uvedených kategorií, jiné kategorie nejsou povoleny.
 - Zkontroluj, zda jsi identifikoval hlavní kategorii řešení správně, pokud ne, tak ji oprav.
 - Zkontroluj, zda má řešení kazuistiky `###{solution}###` odpovídat více kategoriím, a pokud ano, tak zdůvodni proč a dopiš další návrhy kategorií.
 - `###` Zde jsou Few-Shot Examples pro zlepšení identifikace jednotlivých kategorií: [příklady řešení ke každé kategorii vynechány]
 - Dodržuj tento formát výstupu:
 - Hlavní kategorie řešení: (uveď pouze label kategorie)
 - Zdůvodnění:
 - Zhodnocení, zda k řešení patří další kategorie:
 - Případně navrhní další kategorie a odděl je čárkou:
 - Zdůvodnění ostatních kategorií:
 - Ověření:

Poznámka. Klasifikační notebook první etapy, varianta s příklady v instrukci. Pro srovnání s lidmi se z výstupu přebírá hlavní kategorie (kapitola 6).

Mezi příklady klasifikačního promptu jsou i celá řešení ze srovnávacího souboru Studie 2;

kapitola 6 proto uvádí shodu i po vyloučení těchto případů.

B.4 Generace srovnávacího řešení pro druhou etapu

Ke každé kazuistice druhé etapy vygeneroval model gpt-4o alternativní řešení; instrukce požadovala paritu délky, gramatického rodu a minulého času (kapitola 7). Znění instrukce uvedené v souvisejícím článku (anglicky):

System:

- You are an expert on inappropriate and challenging student behavior and an expert in classroom behavior management.
- The user will provide a detailed description of a problem involving a student's (or students') challenging behavior; based on the case study and the student's developmental/educational history, you will propose the best possible behavior-management techniques.

User:

- Task: Based on the following problem description, produce a detailed account of the resolution of the problematic behavior.
- Write in the first person, chronologically, and you may include some key lines of dialogue with the student and concrete activities that took place during the resolution.
- Write the output as a single continuous paragraph and preserve as much of the original stylistic tone of the problem description as possible. The solution must address the specific situation.
- Begin directly with the description of the solution; do not begin with a phrase such as "In this situation, I...".
- Avoid using technical terms from the "Task:" section.
- Write the entire text in the past tense and maintain the grammatical

```
gender implied by the problem description.
- Child's age: {row["year"]} years.
- Length: approximately {row["length"]} sentences, {row["length_char"]}
  characters.
###
Here is the problem description: {row['description_cs']}
###
Output:
```

Poznámka. Rukopis souvisejícího článku o srovnání učitelských a generovaných řešení, do-
datek „Generation Prompt“.

Znění instrukce v návrhu studie (česky, s ukázkovými hodnotami věku a délky):

```
Zadání: Na základě následujícího popisu problému s náročným chováním
žáka zformuluj velmi praktické, co nejvíce pedagogicky a didakticky
vhodné řešení z pohledu efektivního classroom managementu. Řešení napiš
v jednom uceleném odstavci a zachovej co nejvíce původní styl
(stylistiku) z popisu problému. Jen pokud je to opravdu vhodné, tak
použij pro řešení přímou řeč. Uveď konkrétní rady a postupy, které budou
ihned použitelné v každodenní praxi. Tyto nejlepší možné rady a postupy
vychází z kontextu popisu problému. ###
- Začínej vygenerované řešení rovnou popisovaným řešením.
- Nepoužívej ve vygenerovaném řešení žádná odborná slova ze "Zadání:"
- Napiš řešení v minulém gramatickém čase celé řešení.
- A zohledni gender, který vyplývá z popisu problému.
- Věk dítěte: 9 let.
- Navržené řešení bude mít ideálně kolem 168 slov.
###
Zde je popis problému: {text}
```

Výstup:

Poznámka. Návrh designu druhé anotační etapy. Hlavní generaci předcházely přípravné generace deseti a poté patnácti řešení.

B.5 Rámcovaná řešení platformy Edustories

Platforma nabízí u kazuistik šest generovaných řešení v pojmenovaných rámcích (kapitola 3): proaktivní přístup, reaktivní přístup, sociálně-emoční učení, pozitivní behaviorální podpora, restorativní praxe a nenásilná komunikace. Analýzy knihy z těchto řešení nevycházejí; srovnávací řešení Studie 3 (oddíl B.4) s nimi nesouvisí (kapitola 7).

B.6 Predikce dopadu řešení jazykovým modelem

Při predikčním experimentu Studie 4 (oddíl 8.6) dostal model popis situace, anamnézu a řešení a měl odhadnout dopad v jedné ze tří kategorií třetí etapy; definice kategorií v instrukci jsou zkrácenou parafrází operacionalizace etapy (příloha A.3). Popis výsledku (v anotačních listech část „Výstup“) se do zadání záměrně nevkládal. Test úniku informace o výsledku hledal v zadání prvních šedesát znaků popisu výsledku (s rozlišením velkých a malých písmen; kratší popisy se netestovaly); dvě kazuistiky, u nichž test popis výsledku v zadání zachytil, byly vyřazeny. Parafráze, jiné části popisu výsledku ani významové nápovědy test zachytit nemohl. Predikce probíhala na modelech kimi-k3 a glm-5.3 (ve vývojové fázi i na dalších kandidátech) nasazených v národní akademické infrastruktuře e-INFRA CZ, s teplotou nula. Systémová instrukce zněla takto:

Jsi zkušená posuzovatelka pedagogických kazuistik. Dostaneš popis situace ve škole, anamnézu žáka a popis toho, jak učitel situaci řešil. Popis výsledku k dispozici nemáš; tvým úkolem je odhadnout,

jak řešení podle všeho dopadlo.

Vyber právě jednu z těchto tří možností:

NEU = neúspěch: problém trvá dál nebo se zhoršil, řešení nepomohlo.

KU = krátkodobý nebo částečný úspěch: situace se zklidnila
jen na čas nebo jen zčásti.

DU = dlouhodobý úspěch: problém se podařilo vyřešit a změna vydržela.

Odpověz jediným slovem: NEU, KU, nebo DU.

Uživatelská zpráva měla vždy týž tvar: POPIS SITUACE:, ANAMNÉZA: a ŘEŠENÍ:, každý oddíl s doslovným zněním příslušné části kazuistiky. V doplňkové variantě předcházelo šest příkladů z kazuistik mimo hodnocenou množinu, uvedených jako dvojice zpráva uživatele a odpověď modelu.

Poznámka. Skript predikčního experimentu v doprovodném repozitáři (Nehyba, 2026).

Příloha C. Doplňkové tabulky

Příloha uvádí registr anotačních událostí, přehled velikostí datových řezů a přehled měř shody. Počty kazuistik se vztahují k záznamům spárovaným na jednotný identifikátor; úplná tabulka pokrytí (kazuistika × událost) je v doprovodném repozitáři (Nehyba, 2026).

C.1 Registr anotačních událostí

Tabulka C.1. Anotační události tří etap (anotátorky pod projektovými pseudonymy A1–A12).

Číslo pseudonymu je projektové: táž osoba vystupuje pod týmž pseudonymem ve všech etapách a kapitolách.

Událost	Období	Kódující	Kazuistik	Poznámka
K1 zácvik (pokus)	prosinec 2023 – únor 2024	A1, A2, A5, A6	–	kvalitativní kódy bez matice; kódovací kniha, strom V1→V2
K1 test 1	březen 2024	pět souborů	10	
K1 test 2	duben 2024	A1–A4, A7	10	
K1 finál	od května 2024	A1–A4, A7	1 344	dělba práce s vestavěným překryvem
K1 sloučení do matice	srpen 2024	–	1 377	lidské kódy K1 sloučené do matice korpusu
K2 zácvik 1	duben 2025	A1, A2, A8–A11	10	
K2 zácvik 2	duben 2025	A1, A2, A8–A12	38	

Příloha C. Doplňkové tabulky

Událost	Období	Kódující	Kazuistik	Poznámka
K2 příprava	červen 2025	gpt-4o (strojově)	–	generace srovnávacích řešení, ne anotace
K2 finál	červen 2025	A1, A2, A8, A9, A10, A12	309	zaslepené hodnocení dvojic řešení (kapitola 7)
K3 zácvik	2026	A2, A12	6	
K3 pilot 1	2026	A2, A12	19	
K3 pilot 2	2026	A2, A12	49	
K3 finál	březen až září 2026	A2, A12	3 185	dělba práce s vestavěným překryvem; uzavřeno řezem ze září 2026 (kapitola 8)
zveřejnění datové sady (ne anotace)	podzim 2024	–	1 246	veřejná datová sada, výběr z řezu srpen 2024

Poznámka. Vlastní výpočet z tabulky pokrytí; kód a data v doprovodném repozitáři (Nehyba, 2026).

C.2 Přehled velikostí datových řezů

Tabulka C.2. Velikosti datových řezů a vzorků užívaných v knize.

Číslo	Co označuje	Zdroj
3 202	zveřejněné kazuistiky, řez květen 2026	kapitola 5
1 695	řez srpen 2024 (prostor identifikátorů druhé etapy)	kapitola 5

Číslo	Co označuje	Zdroj
1 492	veřejná datová sada (kurátorovaný výběr z řezu srpen 2024); záznamy výběru odpovídají 1 246 unikátním kazuistikám (Tabulka C.1)	kapitola 5, příloha C
1 359	reprodukovatelný řez K1 (deduplikované kazuistiky s lidskou anotací v matici)	kapitoly 4 a 5
324	vzorek druhé etapy (zaslepené dvojice)	kapitola 7

Poznámka. Hodnoty z tabulek výsledných hodnot analýz v uvedených kapitolách.

C.3 Přehled měř shody

Tabulka C.3. Shrnutí spolehlivosti napříč etapami (podrobnosti a intervaly spolehlivosti v kapitolách 6, 7 a 8).

Etapu	Veličina	Hodnota
K1 (lidé)	α typ chování	0,50
K1 (lidé)	α typ řešení	0,29
K1 (lidé)	α dopad	0,72
K1 (model $\times$ člověk)	přesná shoda	0,72
K1 (model $\times$ člověk)	rozšířená shoda	0,93
K1 (model $\times$ člověk)	Cohenovo κ	0,64

Etap	Veličina	Hodnota
K1 (model $\times$ člověk), bez kazuistik z příkladů instrukce	Cohenovo κ	0,61
K2	α vhodnost	0,53
K2	α reaktivní–proaktivní	0,65
K2	α humanistická–behaviorální	0,53
K2	α systémová–situační	0,56
K2	Jaccard rámce	0,73
K3	κ kvalita (bez sporných kazuistik)	0,77
K3	κ kvalita (plný překryv)	0,67
K3	AC1 kvalita	0,98
K3	κ dopad	0,69
K3	vážená κ vhodnost	0,52
K1 $\times$ K3	κ dopad mezi etapami	0,59
K3 predikce	vyvážená přesnost modelu glm-5.3 vůči shodě anotátorek	0,60
K3 predikce	vyvážená přesnost modelu kimi-k3 vůči shodě anotátorek	0,60
K3 čtení výstupu	vyvážená shoda A2 vůči A12 (jiná úloha než predikce)	0,80
K3 predikce	vyvážená přesnost hádání většinové třídy	0,33

Poznámka. Vlastní analýza; tabulky výsledných hodnot studií 2 až 4 v doprovodném repozitáři (Nehyba, 2026; kapitoly 6 až 8).

Příloha D. Etické náležitosti

Příloha uvádí etické náležitosti sběru a zpracování dat: postavení přispěvatelů kazuistik, znění informovaného prohlášení, anonymizační protokol a schválení etickou komisí; pravidla nakládání s nezveřejněnými kazuistikami shrnuje oddíl 4.7. Sběr probíhá na pracovišti řešitele od roku 2021, od září 2023 v rámci projektu TA ČR TQ01000030. Identifikační údaje přispěvatelů, které eviduje informační systém, do analyzovaných dat nevstupují; texty jsou anonymizované. Databáze kazuistik je anonymizovaným výstupem tohoto projektu.

Zastřešující projekt (schválený etickou komisí, oddíl D.2) má dvě datové větve; tato kniha čerpá výhradně z větve kazuistik. Podle posouzení Etické komise pro výzkum MU autoři kazuistik (studující učitelství a učitelé) nejsou účastníky výzkumu ve vlastním smyslu, neboť nejsou předmětem zkoumání: neposkytují data o sobě, nýbrž odevzdávají anonymizovaný text případu, z něhož se dále zpracovávají dílčí datové výřezy. Nástrojem proto není klasický souhlas subjektu výzkumu, ale informované prohlášení přispěvatele (níže).

D.1 Informované prohlášení přispěvatele

Prohlášení se uděluje jako odpovědník přímo v Informačním systému MU (ne jako samostatný podepsaný dokument). Přispěvatel se seznámí s informací o výzkumu a poté vyplní sedm položek: vlastní identifikaci (jméno, UČO a kontaktní e-mail) sloužící k evidenci prohlášení, a tři výslovné souhlasy (využití anonymizované kazuistiky, potvrzení o informování učitele-zdroje, závazek úplné anonymity). Prohlášení zní (vynechávky vyznačeny [...]):

1. Informace o výzkumném projektu

Název projektu: Využití umělé inteligence a virtuální reality v rozvoji kompetencí k řešení výchovných situací ve škole

Hlavní výzkumník: Jan Nehyba, Katedra pedagogiky, Pedagogická fakulta MU

O čem je tento výzkumný projekt?

Významným problémem studentů učitelství a učitelů je udržení kázně ve školách. Cílem projektu je vytvořit rozsáhlou webovou platformu inteligentní databáze případů (kazuistik) nazvanou Edupříběhy, které dokládají osvědčené postupy úspěšného i neúspěšného řešení problémového chování žáků. Tato platforma umožní studentům učitelství, učitelům a dalším uživatelům vyhledávat podobné případy [...] a trénovat si specifické respektující dovednosti. [...]

Jak bude účast probíhat?

Vypracujete anonymizovanou kazuistiku, která bude zařazena do souboru výzkumných dat. [...] Získaná data v anonymní podobě mohou být použita v dalších výzkumných projektech v této oblasti. Souhlasem vyjadřuji porozumění tomu, co znamená vytvořit anonymizovanou kazuistiku [...].

Je účast v tomto výzkumu povinná?

Účast ve výzkumu je dobrovolná, pokud se rozhodnete neudělit souhlas s výzkumem, nebude Vaše odevzdaná kazuistika do souboru výzkumných dat zařazena. Udělení či neudělení souhlasu nijak neovlivňuje Váš průchod předmětem.

Kde se dozvědět více o projektu?

O projektu Edustories se můžete dozvědět zde: <https://edustories.cz/>

Kdykoliv je možné tento souhlas odvolat, a to formou odeslání e-mailu

s prosbou o odvolání na adresu nehyba@ped.muni.cz.

Podpsaný souhlas si můžete vytisknout z informačního systému.

[Potvrzení:] Potvrzuji, že jsem si přečetl/a informace o výzkumném projektu.

2. Souhlas s účastí ve výzkumu

Vypište prosím celé své jméno a příjmení

3. Uveďte své UČO

4. Uveďte kontaktní email

5. Souhlas s účastí ve výzkumu

Na základě výše uvedených informací, já (viz jméno a příjmení) souhlasím s poskytnutím anonymizované kazuistiky pro výzkumné a aplikační účely veřejné databáze kazuistik.

[Ano, souhlasím / Ne, nesouhlasím]

6. Potvrzuji, že informuji osobu, od které kazuistiku získám, že kazuistika bude použita ve výše uvedeném projektu a navazujících projektech.

[Ano, souhlasím]

7. Zavazuji se, že odevzdaná kazuistika bude zcela anonymní a nebude možné z ní identifikovat jakékoliv osoby.

[Ano, souhlasím]

Poznámka. Informované prohlášení přispěvatele, odpovědník v Informačním systému MU.

Ze znění prohlášení plynou tři důsledky. Za prvé, účast je dobrovolná, prohlášení lze odvolat e-mailem a jeho (ne)udělení nemá vliv na průchod předmětem. Za druhé, přispěvatelé se položkou 7 zavazují, že odevzdaný text bude zcela anonymní; to je první stupeň anonymizace, na který navazuje anonymizační protokol (oddíl D.2). Identifikace v položkách 2–4

slouží jen k evidenci prohlášení, ne jako výzkumná data. Za třetí, prohlášení výslovně počítá s užitím anonymních dat v navazujících projektech, o něž se opírá i tato kniha.

D.2 Anonymizační protokol

Anonymizace je trojstupňová s následnou ruční kontrolou. Metodologie projektu ji popisuje takto (slovensky):

Prvý stupeň zabezpečovali participanti, poučení o anonymizácii, tak, že nevkladali žiadne pravé, identifikačné údaje o školách, či osobách v kazuistikách. Druhý stupeň anonymizácie prebiehal vo výskumnom tíme prostredníctvom Czech Named Entity Corpus 2.0.. Tretí stupeň vykonával jazykový model (LLM). Následne boli kazuistiky manuálne skontrolované, či sa v nich nevyskytujú žiadne identifikačné údaje.

Poznámka. Metodologický dokument projektu, oddíl Etické aspekty.

Technicky je druhý stupeň realizován rozpoznáváním pojmenovaných entit (NameTag, Czech Named Entity Corpus 2.0) s náhradou osobních a místních jmen zástupným řetězcem a třetí stupeň instrukcemi jazykového modelu ve strukturním promptu; obojí uvádí příloha B. Tento postup platí pro všechna data analyzovaná v knize (řezy po květen 2026). Od července 2026 anonymizují databázi po lokálním rozpoznávání pojmenovaných entit nástrojem NameTag (ÚFAL, model CNEC) jazykové modely provozované v infrastruktuře e-INFRA CZ na serverech Masarykovy univerzity.

Schválení etickou komisí

Výzkumný projekt, v jehož rámci korpus vznikl („Využití umělé inteligence a virtuální reality v rozvoji kompetencí k řešení náročného chování ve třídě“, návrh č. 1599/2022), posoudila a schválila Etická komise pro výzkum Masarykovy univerzity pod jednacím číslem

EKV-2022-118: předběžné schválení 12. 12. 2022, finální schválení k řešení 27. 9. 2023 (předsedkyně prof. RNDr. Renata Veselská, Ph.D., M.Sc.). Finální schválení reprodukuje Obrázek D.1.



Jednací číslo	EKV-2022-118
Název projektu	Využití umělé inteligence a virtuální reality v rozvoji kompetencí k řešení náročného chování ve třídě
Návrh	1599/2022
Řešitel	Mgr. Bc. Jan Nehyba, Ph.D.
Pracoviště	Pedagogická fakulta

Etická komise pro výzkum Masarykovy univerzity posoudila návrh výzkumného projektu specifikovaného výše a dne 27.9.2023 jej **schválila k řešení**.

Komise očekává, že bude informována o všech změnách ve výzkumných protokolech a/nebo ve formuláři informovaného souhlasu a požaduje poskytnutí závěrečné zprávy.

Brno, 29.9.2023

prof. RNDr. Renata Veselská, Ph.D., M.Sc.
předsedkyně Etické komise pro výzkum Masarykovy univerzity

Digitally signed by
prof. RNDr. Renata Veselská, Ph.D., M.Sc.

30.09.2023

Etická komise pro výzkum Masarykovy univerzity
Žerotínovo nám. 617/9, 601 77 Brno, Česká republika
T: +420 549 49 6290, E: ekv@muni.cz, W: www.muni.cz

Obrázek D.1. Finální schválení Etické komise pro výzkum Masarykovy univerzity (reprodukce stanoviska; jednací číslo a data viz text).

Výběrové ověření anonymizace

Dodržení protokolu jsme ověřili na náhodném vzorku 50 zveřejněných kazuistik třemi vrstevami kontroly: heuristickým vyhledáváním zbytků strojového zpracování, nezávislým druhým průchodem týmž nástrojem rozpoznávání entit a ručním posouzením všech nálezů. Ve vzorku jsme nenašli žádné technické zbytky automatického zpracování, žádné plné jméno, příjmení, název školy ani obec vázanou na osobu. V textu knihy nevystupují žádné přímé identifikátory; anotátorky vystupují výhradně pod pseudonymy A1–A12 (kapitola 4). Doprovodný repozitář knihy (Nehyba, 2026) zpřístupňuje odvozené tabulky, nikoli úplné texty kazuistik. Pravidla nakládání s texty kazuistik popisuje kapitola 4.

