

Modul 4

Inženýrská odpovědnost a
zodpovědnost v
autonomních systémech



Co-funded by
the European Union



Ethical Engineer

www.ethicalengineer.eu

OBSAH

- 01 Rozdělení odpovědnosti a pravomocí
- 02 Návrh odolnosti a lidské kontroly
- 03 Osvojte si odpovědnost prostřednictvím případových studií
- 04 Vytvoření sledovatelného technického rozhodnutí



Co-funded by
the European Union

Tento projekt byl financován s podporou Evropské komise. Za tuto publikaci (sdělení) nese výlučnou odpovědnost autor a Komise nenese žádnou odpovědnost za jakékoli použití informací v ní obsažených. V souladu s novým rámcem GDPR upozorňujeme, že partnerství bude zpracovávat vaše osobní údaje výhradně v zájmu a za účelem projektu a bez jakéhokoli dotčení vašich práv.



Přehled kurzu „Etický inženýr“

Hlavní obsah a výstupy pro každého studenta v jednotlivých modulech

Modul 01	Modul 02	Modul 03	Modul 04	Modul 05	Modul 06
Etika a transparentnost Etické zásady a hodnoty EU Transparentnost a vysvětlitelnost Analýza případů z oblasti umělé inteligence a robotiky VÝSTUP: mapa etických otázek	Zaujatost, spravedlnost a inkluze Zaujatost v průběhu celého životního cyklu umělé inteligence Cíle spravedlnosti a důkazy Příklady inkluzivního inženýrství VÝSTUP: audit zaujatosti	Soukromí a ochrana osobních údajů Ochrana soukromí již od návrhu a GDPR Správa dat a PET Posuzování a napadnutelnost VÝSTUP: návrh s ohledem na ochranu soukromí	Odpovědnost a bezpečnost Role a odpovědnost Selhání, odolnost a předání Protokolování a poučení z incidentů VÝSTUP: záznam o sledovatelnosti	Návrh založený na hodnotách Požadavky na etický design Volba dat, modelů a rozhraní Kompromisy a agilní kontrolní body VÝSTUP: rozhodovací matice	Udržitelnost a dopad Energie, voda a hardware společenské dopady a dopady v průběhu životního cyklu Metriky, správa a přepracování VÝSTUP: plán udržitelnosti

ZÁVĚREČNÁ PRÁCE: zkombinujte šest výstupů do etického inženýrského spisu s využitím jednoho případu z WP2 nebo inženýrského projektu

Zaměření modulů

Selhání umělé inteligence může způsobit škody v reálném světě. Tento modul ukazuje, jak inženýři navrhují transparentnost, bezpečnost a odpovědnost ještě před zahájením provozu.

Účel

Vybavit studenty technických oborů schopností rozdělovat odpovědnost, analyzovat poruchy a obhajovat rozhodnutí týkající se bezpečnosti na základě dohledatelných důkazů.

Cílová skupina

Studenti vysokých škol v oborech inženýrství, informatiky, robotiky a oborech souvisejících s bezpečností.

Navazuje na

Moduly 1 až 3: etika, transparentnost, spravedlnost, ochrana soukromí, správa dat a záznamy o inženýrských rozhodnutích.

[Zdroje] Aplikace KA220-HED schválená organizací Ethical Engineer, WP3 Výuka a učení.

Zaměření modulu

Selhání umělé inteligence může způsobit škody v reálném světě. Tento modul ukazuje, jak inženýři navrhují transparentnost, bezpečnost a odpovědnost ještě před zahájením provozu.

Délka

Základní varianta 210 minut. Rozšířená varianta 270 minut.

Vzdělávací sekvence

Přidělte pravomoci, navrhnete odolnost, analyzujte případy a přijměte zodpovědné inženýrské rozhodnutí.

Výstupy účastníků

Mapa odpovědností, analýza selhání, argumentace pro zajištění kvality a záznam o odpovědnosti.

[Zdroje] Aplikace KA220-HED schválená organizací Ethical Engineer, WP3 Výuka a učení.

Výstupy učení

Na konci modulu budou studenti schopni:

1. Rozlišujte mezi odpovědností, zodpovědností a právní odpovědností.
2. Zmapujte odpovědnost za rozhodování a řízení v rámci celého životního cyklu autonomního systému.
3. Analyzujte režimy selhání dat, modelů, provozních, organizačních a lidských faktorů.
4. Specifikujte opatření pro odolnost, záložní řešení, předávání, protokolování, zajištění a řízení změn.
5. Vytvořte záznam o odpovědnosti propojující důkazy, vlastníky, zbytkové riziko a postupy přezkumu.

Odpovědnost musí být navržena ještě před selháním

- Autonomní systémy rozdělují rozhodování mezi dodavatele dat, modelové týmy, integrátory, operátory, manažery a regulační orgány. Pokud není jasné, kdo má pravomoc, může bezpečnostní problém zůstat nevyřešen.
- Odpovědnost vyžaduje, aby osoba nebo organizace vysvětlila a odůvodnila rozhodnutí, poskytla důkazy, podstoupila přezkum a podpořila nápravu nebo opravu.
- Inženýři to umožňují prostřednictvím rozhodovacích práv, sledovatelnosti, důkazů z testů, monitorování, postupů při incidentech a pravomoci zastavit provoz.

[Zdroje] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Evropské rámce propojují povinnosti s inženýrskými důkazy

Důvěryhodné pokyny pro AI identifikuje odpovědnost, lidský faktor a technickou robustnost jako vzájemně propojené požadavky.

Zákon EU o umělé inteligenci přisuzuje povinnosti definovaným subjektům a vyžaduje, aby příslušné systémy s vysokým rizikem podporovaly řízení rizik, protokolování, lidský dohled, přesnost, robustnost a kyberbezpečnost. Směrnice o odpovědnosti za výrobky se za stanovených podmínek zabývá odpovědností za vadné výrobky, včetně softwaru.

Rámcové dokumenty nerozdělují všechny úkoly v rámci projektu. V rámci technického řízení je nutné definovat, kdo rozhoduje, realizuje, ověřuje, schvaluje, monitoruje a reaguje.

01

Přidělte odpovědnost a pravomoci

Rozlišujte povinnosti,
mapujte rozhodovací
pravomoci a propojujte
etické požadavky s důkazy.

Odpovědnost, zodpovědnost a právní odpovědnost jsou odlišné pojmy

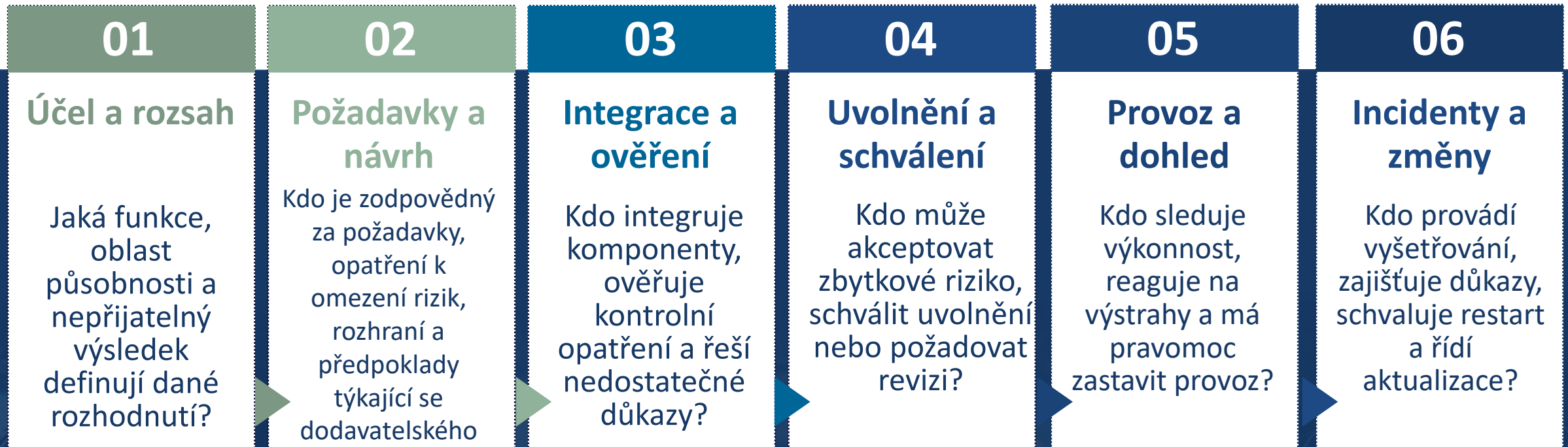
- Odpovědnost se týká přidělené povinnosti nebo činnosti a může být sdílená mezi různými rolemi. Jeden inženýr může provést test; jiný může ověřit výsledek.
- Účinnost odpovědnosti je povinnost vysvětlit a zdůvodnit rozhodnutí, poskytnout důkazy a podrobit se kontrole. Jmenovaný vlastník vydání musí být schopen obhájit výsledek.
- Právní odpovědnost se týká důsledků stanovených podle platných právních předpisů a skutečností. V nemocnici se na péči podílí mnoho zaměstnanců, ale pravomoci, záznamy o rozhodnutích a postupy přezkumu musí jasně ukazovat, kdo mohl danou činnost schválit, zastavit nebo opravit.

[Zdroje]

- <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>
- <https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng>

Zmapujte odpovědnost v rámci celého životního cyklu systému

Vyberte jedno rozhodnutí týkající se bezpečnosti. V každé fázi určete pravomoc, odpovědnost, důkazy, kontrolu a eskalaci.



VÝSTUP Mapa odpovědností propojující každou fázi s osobou odpovědnou za rozhodnutí, osobou odpovědnou za kontrolu, osobou odpovědnou za důkazy, osobou oprávněnou k zastavení a spouštěčem přezkumu.

Současný kontext odpovědnosti v EU vyžaduje přesná tvrzení

- Zákon EU o umělé inteligenci upravuje definovaná rizika a povinnosti subjektů. Nevytváří však ucelený kodex občanskoprávní odpovědnosti pro každou škodu související s umělou inteligencí.
- Směrnice (EU) 2024/2853 aktualizuje pravidla odpovědnosti za výrobky a zahrnuje software do rámce pro výrobky. Mohou se rovněž uplatnit vnitrostátní občanské právo, smluvní právo, předpisy o ochraně osobních údajů, ochrana spotřebitele, bezpečnostní předpisy a odvětvová legislativa.
- Navrhovaná směrnice o odpovědnosti za umělou inteligenci byla v roce 2025 stažena a nestala se právem EU. V konkrétních případech zajistěte důkazy a vyhledejte odbornou právní radu.

[Zdroje]

- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20250727>
- <https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng>
- https://eur-lex.europa.eu/procedure/EN/2022_303

Rozhodovací práva: Proměňte seznam rolí v systém řízení

- Definujte, kdo může stanovit provozní limit, přijmout riziko, schválit aktualizaci modelu, přepsat automatizaci, pozastavit používání a nahlásit incident.
- Oddělte přípravu, ověření a schválení tam, kde záleží na nezávislosti. Zaznamenávejte konzultace a nesouhlasné názory. Vyhněte se prisuzování odpovědnosti výboru bez jmenovaného vlastníka rozhodnutí.
- Otestujte schéma na realistické selhání a zeptejte se, kdo jedná v první minutě, hodině a dni.

[Zdroje]

- <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>

Poskytovatelé, provozovatelé, operátoři a dodavatelé vidí různé důkazy

- Poskytovatel má pod kontrolou důkazy týkající se návrhu a vývoje. Nasazovatel má pod kontrolou kontext použití, konfiguraci, personál a provozní procesy. Provozovatel jedná prostřednictvím rozhraní. Dodavatel může mít pod kontrolou důkazy týkající se modelu, softwaru nebo hardwaru.
- Technická dokumentace a smlouvy by měly propojit tyto pohledy prostřednictvím zamýšleného použití, omezení, testů, protokolů, aktualizací, incidentů a přístupu k revizím.
- Chybějící důkazní řetězec může znemožnit vyvození odpovědnosti, i když jsou role jasně stanoveny.

[Zdroje]

- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>
- <https://eur-lex.europa.eu/eli/dir/2024/2853/oj/eng>

Odpovědnost za riziko vyžaduje pravomoc, zdroje a eskalaci

- Vlastník rizika sleduje riziko a rozhoduje o tom, jak bude řízeno. Vlastník kontrolního opatření zavádí ochranné opatření. Vlastník důkazů vede záznamy potřebné pro ověření a přezkoumání.
- Přiřazení odpovědnosti bez rozpočtu, kompetence nebo pravomoci vede k pouhé formální kontrole.
- U každého významného rizika zaznamenejte vlastníka, kontrolní opatření, kritérium přijatelnosti, zbytkové riziko, prahovou hodnotu pro eskalaci a datum přezkumu.

[Zdroje]

- <https://www.nist.gov/itl/ai-risk-management-framework>
- <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Etické zásady se stávají požadavky a použitelnými důkazy

- Autonomie se stává lidskou autoritou, informacemi a dobou odezvy. Spravedlnost se stává testováním distribuce, spravedlivým procesem a nápravou. Transparentnost se stává důkazy specifickými pro dané publikum a sledovatelností. Neškodnost se stává kontrolami rizik, ověřováním a monitorováním.
- **Shromažďujte důkazy, které podporují rozhodnutí:** požadavky, analýza rizik, výsledky testů, modelové karty, datové listy, záznamy o schválení, protokoly, změny a incidenty.
- Recenzent by měl být schopen vysledovat ovlivněný výsledek až k verzi systému, datovým hranicím, podmínkám, ovládacím prvkům, lidskému zásahu a schválení.

[Zdroje]

- <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>

Příklad z praxe strojního inženýrství: Rozhodnutí o uvolnění čerpadla

Údržbářský tým navrhuje po výměně ložisek obnovit plný výkon odstředivého procesního čerpadla. Rozhodnutí podporuje monitorovací systém využívající umělou inteligenci, avšak pravomoc k uvolnění čerpadla do provozu zůstává v kompetenci inženýrů.

01

Kritické rozhodnutí

Uvolněte čerpadlo pro nepřetržitý provoz při 3 600 ot./min. Rozhodnutí o uvolnění je v kompetenci vedoucího údržby.

02

Odpovědnost za práci

Strojní inženýr stanoví limity vibrací a teploty. Technik zaznamená vyrovnaní, utahovací moment a montáž ložisek.

03

Nezávislé důkazy

Před schválením provádí samostatný inženýr pro spolehlivost kontrolu základních vibrací, teplotního vývoje, zprávy o vyrovnaní a zkušební vypnutí.

[Zdroje]

- Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.
- <https://www.nist.gov/itl/ai-risk-management-framework>

Příklad z praxe strojního inženýrství: Rozhodnutí o uvolnění čerpadla

Údržbářský tým navrhuje po výměně ložisek vrátit odstředivé procesní čerpadlo do provozu při plném zatížení. Monitor stavu s umělou inteligencí toto rozhodnutí podporuje, avšak pravomoc k uvolnění zůstává v kompetenci inženýrů.

04

Řídicí opatření

Použijte řízený nárůst otáček, automatické vypnutí při vysokých vibracích, chráněnou výlučnou zónu a lokální ruční vypnutí.

05

Oprávnění k zastavení

Čerpadlo může zastavit operátor z velínu a vedoucí bezpečnosti. Dodavatel nemůže přepsat oprávnění k odstavení na místě.

06

Rozhodnutí a přezkum

PROVEĎTE ÚPRAVY, dokud nebude úspěšně absolvován zkušební provoz a nezávislé posouzení. Schvalte 24hodinový monitorovaný provoz a poté před plným uvolněním zkontrolujte trendy.

[Zdroje]

- Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.
- <https://www.nist.gov/itl/ai-risk-management-framework>

Příklad z praxe strojního inženýrství: Rozhodnutí o uvolnění čerpadla

Údržbářský tým navrhuje po výměně ložisek vrátit odstředivé procesní čerpadlo do provozu při plném zatížení. Monitor stavu s umělou inteligencí toto rozhodnutí podporuje, avšak pravomoc k uvolnění do provozu zůstává v kompetenci inženýrů.

Zodpovědnost změnila uvolnění

Skóre modelu se stalo jednou z položek v řetězci důkazů s verzemi a nezávislým ověřením, nazvaném „oprávnění k zastavení“ a „stupňové schvalování“.

[Zdroje]

- Ethical Engineer schválil aplikaci KA220-HED, WP3 Výuka a učení.
- <https://www.nist.gov/itl/ai-risk-management-framework>

02

Návrh odolnosti a lidské kontroly

Specifikujte odolnost proti poruchám, záložní řešení, předávání, omezenou pravomoc, protokolování a řízení změn.

Odolnost proti poruchám začíná definovanou poruchou

Odolnost proti poruchám umožňuje systému pokračovat v provozu v bezpečných mezích i po výskytu definované poruchy. Začněte tím, že pojmenujete relevantní poruchy senzorů, akčních členů, dat, softwaru, komunikace a lidských činností.

Technická dokumentace musí prokázat, která kritická funkce je chráněna, jak je porucha detekována a jaký přechodový stav zajišťuje bezpečnost systému.

[Zdroje]

- Aplikace KA220-HED schválená organizací Ethical Engineer, WP3 Výuka a učení.
- <https://www.nist.gov/itl/ai-risk-management-framework>

Odolnost proti poruchám začíná definovanou poruchou

01

Porucha Porucha

Definujte věrohodné poruchy jednotlivé, kombinované a způsobené společnou příčinou, jejich trvání a bezpečnostní funkci, na kterou mohou mít vliv.

02

Řízení architektury

Použijte nezávislou redundanci, diverzitu, monitorování, izolaci a bezpečný stav odpovídající danému nebezpečí.

03

Ověření

Otestujte pokrytí detekce, dobu přechodu, provoz v omezeném režimu, reakci obsluhy a zbývající jednotlivé body selhání.

Návrh je odolný vůči poruchám pouze v případě poruch a provozních podmínek podložených důkazy. Neznámé závislosti a netestované kombinované poruchy představují zbytkové riziko.

Redundance vyžaduje nezávislost a rozhodovací pravidlo

- Dvě součásti nabízejí slabou ochranu, pokud sdílejí napájení, data, softwarové chyby nebo zranitelnost vůči okolnímu prostředí. Definujte, jak se zjišťují a řeší neshody.
- V případech, kdy má selhání způsobené společnou příčinou význam, použijte různorodé senzory nebo metody. Ověřte pokrytí, latenci a přechod po neshodě.
- Zdokumentujte chráněnou funkci, předpokládanou nezávislost, důkazy a zbývající jednotlivé body selhání.

[Zdroje]

- <https://www.nist.gov/itl/ai-risk-management-framework>

Detekce, omezení a zotavení tvoří jeden kontrolní řetězec

- Detekce identifikuje abnormální chování. Diagnostika rozlišuje pravděpodobné příčiny. Omezení zabraňuje šíření. Záložní řešení přechází na bezpečnější funkci nebo stav. Obnova obnovuje službu za kontrolovaných podmínek.
- Každá fáze vyžaduje časový limit, pozorovatelný signál a odpovědnou osobu. Varování bez proveditelné reakce představuje slabou ochranu. Záložní řešení, které se spoléhá na selhanou komponentu, není nezávislé.
- Otestujte celý řetězec za realistického zatížení a zhoršených komunikačních podmínek.

[Zdroje]

- - <https://www.nist.gov/itl/ai-risk-management-framework>

Plynulé omezování funkcí zachovává nejdůležitější funkce

[Zdroje] - <https://www.nist.gov/itl/ai-risk-management-framework>

- Omezte funkčnost kontrolovaným způsobem, když dojde ke zhoršení spolehlivosti, snímání, komunikace nebo výpočetního výkonu. Definujte, které funkce zůstanou zachovány, které se zastaví a jak uživatelé tuto změnu pochopí.
- Mezi příklady patří nižší rychlost, omezená automatizace, konzervativní prahové hodnoty, lokální ovládání nebo bezpečné vypnutí.
- Ověřte, zda je režim s omezenou funkčností stabilní, sledovatelný a vhodný pro dané prostředí. Zaznamenejte jeho spouštěcí podmínky a maximální povolenou dobu trvání.

Lidský dohled je časově omezená systémová funkce

- Smysluplný dohled vyžaduje, aby osoba rozpoznala situaci, pochopila stav systému, rozhodla se a jednala dříve, než dojde ke škodě. Doba odezvy zahrnuje vnímání, interpretaci a fyzickou akci při realistické pracovní zátěži.
- Navrhněte výstrahy, zobrazení nejistoty, oprávnění, přepsání, záložní řešení a školení jako jednu funkci. Nepřetržité sledování vzácných událostí je bez podpory nespolehlivé.
- Testujte s reprezentativními operátory, podmínkami a scénáři selhání.

[Zdroje] <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Předání musí specifikovat stav, čas a pravomoc

- Předání mění řízení mezi automatizací a osobou nebo jiným systémem. Převzetí vyžaduje aktuální stav, důvod, dostupné akce, zbývající čas a potvrzení, že řízení bylo převedeno.
- Předběžná upozornění mohou operátora připravit. Pokud nedojde k žádné reakci, může být nutný manévr s minimálním rizikem. Rozhraní by měla zabránit nejednoznačnému sdílení pravomocí.
- Zaznamenejte spouštěcí událost, zobrazené informace, reakci a výsledný stav.

[Zdroje]

- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>
- <https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-issue-2>

Agentní systémy potřebují ohraničený rozsah akcí

- Systémy, které plánují, volají nástroje nebo provádějí akce napříč propojenými službami, mohou vytvářet dlouhé řetězce akcí. Před nasazením definujte povolené cíle, data, nástroje, fyzické stavy, limity výdajů nebo zdrojů a zakázané akce.
- Pro akce kritické z hlediska bezpečnosti nebo nevratné akce vyžadujte schvalovací brány. Sledujte volání nástrojů, změny stavů a nesplněná omezení. Zajistěte možnost rychlého zrušení, bezpečný záložní režim a přehledný náhled pro operátora.
- Důkazy o odpovědnosti musí ukazovat, kdo rozsah autorizoval, která verze byla provedena a jak byla akce zastavena nebo zvrácena.

[Zdroje]

- <https://www.nist.gov/itl/ai-risk-management-framework>
- <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>

Protokoly umožňují rekonstrukci a poučení

- Užitečné protokoly propojují čas, verzi systému, konfiguraci, vstupy, výstupy, míru spolehlivosti, alarmy, lidské zásahy, přepsání a změny. Měly by podporovat reakci na incidenty, aniž by shromažďovaly zbytečné osobní údaje.
- Definujte strukturovaný formát, synchronizaci času, identifikátory, integritu, přístup, uchovávání, rotaci, export a zálohování. Otestujte, zda revizor dokáže rekonstruovat kritické rozhodnutí.
- Protokol se stává prostředkem kontroly odpovědnosti, pokud jej někdo sleduje, anomálie vyvolávají akci a záznam může posloužit jako podklad pro přezkum po incidentu.

[Zdroje]

- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>
- <https://www.nist.gov/itl/ai-risk-management-framework>

Reakce na incident uzavírá smyčku odpovědnosti

- Definujte, jak mají zaměstnanci identifikovat, omezit, nahlásit, vyšetřit a poučit se z incidentu nebo téměř nehody. Zachovejte důkazy a chraňte dotčené osoby. Pověřte odpovědné osoby pravomocemi k pozastavení provozu, externímu oznámení, komunikaci a obnovení provozu.
- Prověřte technické, lidské, organizační a dodavatelské příčiny. Nápravná opatření vyžadují odpovědné osoby, termíny a ověření.
- Sdílejte získané poznatky napříč podobnými systémy a aktualizujte posouzení rizik.

[Zdroje]

- <https://www.nist.gov/itl/ai-risk-management-framework>
- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>

Zajišťovací případ propojuje tvrzení o bezpečnosti s důkazy

- Zajišťovací případ uvádí tvrzení, odůvodnění, které jej podporuje, a důkazy, které lze přezkoumat. Příklad: robot zůstává v bezpečných rozestupových mezích, protože spolupracují snímání, řízení rychlosti, ochranné zastavení a ověřovací důkazy.
- Zaznamenejte předpoklady, provozní podmínky, míru spolehlivosti, chybějící důkazy a protidůkazy. Každou kontrolu propojte s měřitelným kritériem přijatelnosti.
- Případ musí uvádět verzi systému a datum schválení. Slouží jako podklad pro přezkoumání a nemůže nahradit testování ani nezávislé ověření.

[Zdroje]

- <https://www.nist.gov/itl/ai-risk-management-framework>
- <https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-issue-2>

Kontrola změn zabraňuje tomu, aby se důkazy o zajištění staly zastaralými

- Vytvořte referenční stav schváleného modelu, softwaru, hardwaru, dat, konfigurace a provozních podmínek. Klasifikujte každou změnu a posuďte její vliv na rizika, rozhraní, lidskou práci a předchozí důkazy.
- Vyžadujte cílené regresní testy, nezávislé přezkoumání změn relevantních pro bezpečnost, postupné uvolňování, monitorování a vrácení do původního stavu. Informujte aktéry, kteří se spoléhají na změněné důkazy.
- Podstatná úprava může změnit regulační povinnosti. Zkontrolujte AI Act a příslušná pravidla pro daný produkt nebo odvětví.

[Zdroje]

- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>
- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02023R1230-20260727>
- <https://unece.org/transport/documents/2021/03/standards/un-regulation-no-156-software-update-and-software-update>

03

Naučte se zodpovědnosti prostřednictvím případových studií

Analyzujte projektové a oficiální případy z technické, lidské i organizační roviny.

Projektová případová studie: Autonomní taxíky s umělou inteligencí

- Případ „Ethical Engineer“ zkoumá autonomní taxíky jako službu, která může zlepšit mobilitu a efektivitu, ale zároveň vyvolává otázky týkající se bezpečnosti, odpovědnosti, zaměstnanosti a důvěry veřejnosti.
- Inženýrské zaměření: provozní doména, limity vnímání, záložní řešení, vzdálená podpora, informace pro cestující, odpovědnost za incidenty a sociální dopad.
- Otevřete případ „Ethical Engineer“ o autonomních taxících. Zjistěte, kdo může povolit provoz, kdo může zastavit vozový park a jaké důkazy k tomu potřebuje.

[Zdroje]

- <https://ethicalengineer.eu/download/223/>
- <https://ethicalengineer.eu/ethical-engineer-case-studies/>

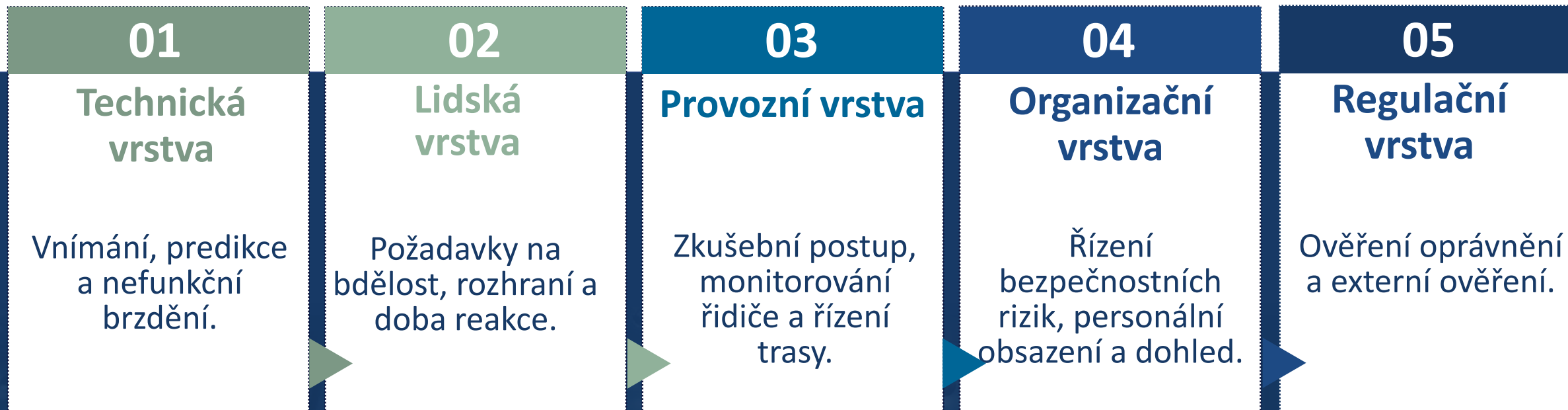
Oficiální případ: Nehoda autonomního vozidla Uberu

- V roce 2018 srazilo a zabilo testovací vozidlo Uberu chodce v Tempe v Arizoně. NTSB zjistila nedostatečné posouzení bezpečnostních rizik, neúčinný dohled a slabou bezpečnostní kulturu. Funkce varování před kolizí a automatického nouzového brzdění byly vypnuty.
- Tento případ ukazuje, že lidský dohled nemůže kompenzovat nedostatečné systémové a organizační řešení.
- Vysledujte selhání přes snímání, klasifikaci, brzdění, roli operátora, dohled a řízení.

[Zdroje] <https://www.nts.gov/investigations/Pages/HWY18MH010.aspx>

Aplikujte vrstvenou analýzu selhání na případ Uberu

Důkladný audit zaujatosti promění důkazy v argument, za který lze nést odpovědnost



Metodu pěti „proč“ používejte opatrně. Pokračujte za hranice chyby operátora až k předchozímu rozhodnutí,
nedostatku důkazů a kontrole, které mohly událost zabránit nebo zmírnit.

Oficiální případ: Boeing 737 MAX

- Vyšetřování výboru Sněmovny reprezentantů Spojených států prozkoumalo návrh, vývoj a certifikaci letounu Boeing 737 MAX po dvou haváriích, při nichž zahynulo 346 lidí. Zpráva se zabývala návrhem systému MCAS, závislostí na senzorech, informacemi pro piloty, certifikací a organizačními tlaky.
- Technické zaměření: selhání v jediném bodě, pravomoc automatizace, předpoklady reakce, komunikace změn a nezávislé přezkoumání.
- Zjistěte, jaké záznamy a rozhodnutí byly vyžadovány v jednotlivých schvalovacích fázích.

[Zdroje] <https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation>

Odpovědnost v případě Boeing přesahuje hranice organizací

- Vysvětlení na úrovni jednotlivých komponent nestačí. Sledujte rozhodnutí napříč konstrukcí letadla, softwarem, posuzováním bezpečnosti, výcvikem, certifikací, řízením a provozem letecké společnosti.
- Zjistěte, kdo vytvořil důkazy o rizicích, kdo je posoudil, kdo kontroloval informace a kdo schválil výsledný systém. Zjistěte, zda tlaky související s výrobou a harmonogramem ovlivnily rozhodnutí týkající se bezpečnosti.
- Vytvořte řetězec sledovatelnosti, který zviditelní předpoklady, změny a nesouhlasné názory před certifikací.

[Zdroje] <https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation>

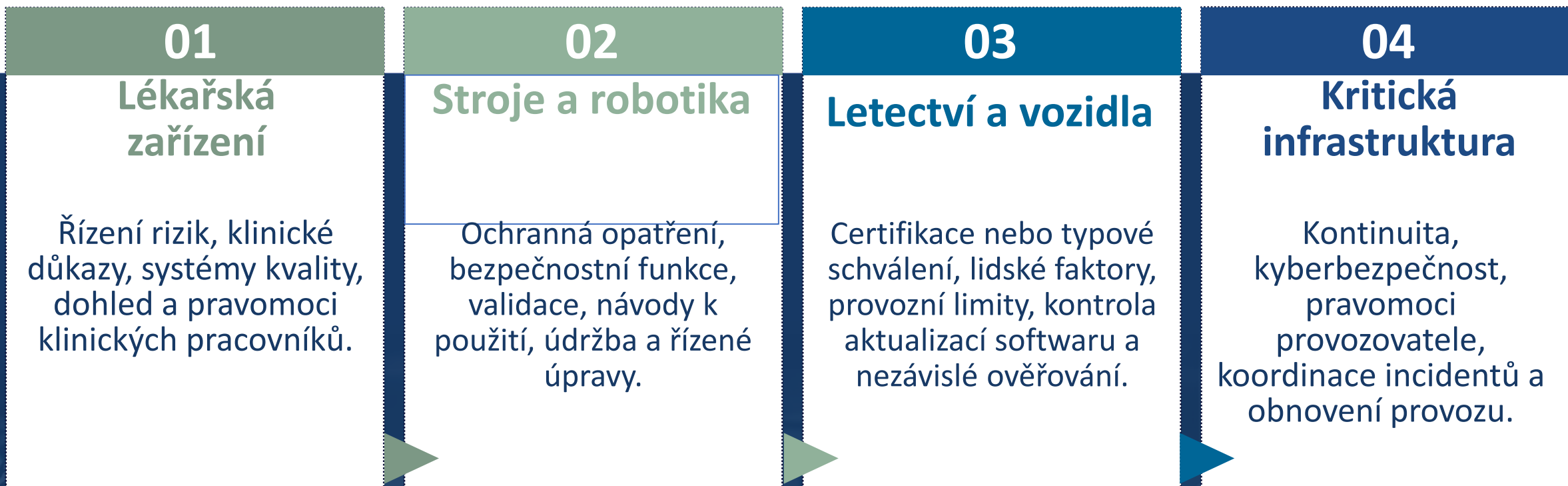
Případová studie: Umělá inteligence v pulzních srdečních pumpách

- Případová studie z časopisu „The Ethical Engineer“ zkoumá napětí mezi standardizovaným designem zdravotnických zařízení a personalizovanou péčí. Standardizace může usnadnit výzkum, klinické hodnocení a schvalovací procesy, zatímco různí pacienti mohou vyžadovat odlišné provozní parametry nebo způsob monitorování.
- Technické zaměření: cílová populace, fyziologické rozdíly, bezpečné limity, pravomoc klinického lékaře, alarmy, důkazy a odpovědnost za úpravy.
- Otevřete případ pulzní srdeční pumpy. Určete, kdo může změnit nastavení a kdo je zodpovědný za jeho ověření.

[Zdroje]

- <https://ethicalengineer.eu/download/235/>
- <https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng>

Odpovědnost za bezpečnost využívá důkazy specifické pro danou oblast



Mapu odpovědnosti lze použít opakovaně. Důkazy o přijatelnosti a příslušné orgány se liší podle dané oblasti.

[Zdroje] <https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng>.
<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02023R1230-20260727>
<https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-issue-2>
<https://unece.org/transport/documents/2021/03/standards/un-regulation-no-156-software-update-and-software-update>

Porovnejte odpovědnost ve všech třech případech

- Případ taxislužby zdůrazňuje provozní oblast a správu vozového parku. Vyšetřování společnosti Uber přidává bezpečnostní kulturu a dohled nad testováním. Případ společnosti Boeing ukazuje certifikaci, informace a organizační tlak. Případ srdeční pumpy zdůrazňuje klinickou autoritu a individuální rozdíly.
- U každého z nich identifikujte klíčové rozhodnutí, odpovědnou osobu, požadované důkazy, pravomoc k zastavení a nápravné opatření.
- Oddělte opakovaně použitelné kontrolní mechanismy odpovědnosti od důkazů z dané oblasti, které jsou nutné pro uvolnění produktu.

[Zdroje]

- <https://ethicalengineer.eu/download/223/>
- <https://www.nts.gov/investigations/Pages/HWY18MH010.aspx>
- <https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation>
- <https://ethicalengineer.eu/download/235/>

Praktický úkol: Proved'te strukturovanou analýzu poruchy

Vyberte si případ
Uberu, Boeingu
nebo srdeční
pumpy.

1. Definujte nežádoucí událost a hranice systému.
2. Vytvořte časovou osu rozhodnutí a podmínek.
3. Identifikujte technické, lidské, provozní a organizační faktory, které k události přispěly.
4. Vyhledejte chybějící nebo přehlédnuté důkazy.
5. Navrhněte opatření pro prevenci, detekci, zmírnění dopadů a obnovu.
6. Určete odpovědné osoby a důkazy pro ověření.

Výstup:

Jednostránková mapa selhání a dvouminutové prohlášení o odpovědnosti.

[Zdroje]

- Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.
- <https://www.nts.gov/investigations/Pages/HWY18MH010.aspx>
- <https://democrats-transportation.house.gov/committee-activity/boeing-737-max-investigation>
- <https://ethicalengineer.eu/download/235/>

Dilema: Uvedení na trh po nevysvětlitelné téměř nehodě

- Dodavatel hlásí, že aktualizace týkající se bezpečnosti splňuje standardní benchmark. Následný reprezentativní test v terénu však vykazuje jeden nevysvětlitelný případ téměř nehody. Vedení chce uvolnění produktu, aby bylo dodrženo smluvní datum.
- Rozhodněte se pro POKRAČOVAT, UPRAVIT nebo ZASTAVIT. Uveďte, které důkazy chybí, kdo může uvedení na trh odložit a zda je ohraničený test obhajitelný. Definujte jeho provozní limity, monitorování, spouštěč zastavení a vrácení do původního stavu.
- Zaznamenejte nejvýraznější nesouhlasný názor a důkazy, které by mohly změnit rozhodnutí.

[Zdroje]

- Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.
- <https://www.nist.gov/itl/ai-risk-management-framework>

04

Vypracujte sledovatelné inženýrské rozhodnutí

Vyplňte záznam o
odpovědnosti, provádějte revizi
vydání, kontrolu učení a
přispějte do dokumentace.

Případová studie

A: Autonomní skladový robot

- Mobilní robot přepravuje palety ve smíšeném skladu. Aktualizace softwaru zlepšuje efektivitu trasy, ale zvyšuje počet těsných průjezdů v blízkosti pracovníků. Dodavatel poskytuje metriky modelu, ale neprovádí bezpečnostní testy na úrovni systému. Místní manažeři požadují okamžité nasazení před sezónním špičkovým obdobím.
- Rozhodněte se pro POKRAČOVAT, UPRAVIT nebo ZASTAVIT. Definujte provozní hranice, detekci, řízení rychlosti a odstupu, schvalování aktualizací, protokoly, pravomoc k zastavení a důkazy o restartu.
- Určete osobu odpovědnou za každé rozhodnutí.

[Zdroje]

- Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.
- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02023R1230-20260727>

Vyplněný záznam o odpovědnosti: Aktualizace skladového robota

Hranice

Smíšená chodba pro chodce č. 3, vyznačené přechody, suchá podlaha, denní směna pod dohledem a maximální rychlost 1,2 m/s.

Rozhodnutí: Revize

Plné nasazení je pozastaveno. Pilotní projekt v omezeném rozsahu mimo špičku může pokračovat.

Odpovědné osoby

Za uvolnění odpovídá provozní ředitel. Za kontroly rozestupů odpovídá bezpečnostní inženýr. Za aktualizaci důkazů odpovídá dodavatel. Vedoucí směny má pravomoc k zastavení.

[Zdroje]

- Inženýr pro etiku schválil aplikaci KA220-HED, WP3 Výuka a učení.
- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02023R1230-20260727>

Vyplněný záznam o odpovědnosti: Aktualizace skladového robota

Důkazy

Záznam o verzi, posouzení rizik, 200 zkušebních jízd, výsledky minimální vzdálenosti, test ochranného zastavení, prohlídka pracovníkem a test zpětného chodu.

Ovládací prvky

Rychlostní omezení, přejezdy s geofencem, zvukový signál při přiblížení, nezávislý ochranný skener, protokol incidentů a okamžitý zpětný chod.

Přezkoumání

Schválit pouze v případě, že jsou úspěšně absolvovány všechny ochranné testy, nedojde k žádnému nevysvětlitelnému těsnému průjezdu a jsou vyřešeny připomínky pracovníků. Provedte nové posouzení po 50 provozních hodinách.

[Zdroje]

- Etický inženýr schválil žádost KA220-HED, WP3 Výuka a učení.
- <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02023R1230-20260727>

Projektový případ B: Stanovení priority lékařských výstrah

- Služba umělé inteligence řadí výstrahy pacientů pro tým vzdáleného monitorování. Aktualizace modelu celkově snižuje počet falešných poplachů, avšak její vliv na vzácná srdeční onemocnění je nejistý. Personál vidí skóre, ale nemůže identifikovat verzi modelu ani důvod zařazení do žebříčku.
- Definujte kontroly validace, záložních řešení, rozhraní, protokolování, eskalace a incidentů, které jsou nutné před vydáním.
- Uveďte, kdo akceptuje zbytkové riziko a který výsledek vyvolá okamžité vrácení do původního stavu.

[Zdroje]

- Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.
- <https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng>

Týmová aktivita: Vytvořte řetězec odpovědnosti

Vyberte si případ skladového robota nebo lékařského poplachu. Přiřad'te role vlastníka systému, vedoucího bezpečnosti, dodavatele, operátora, revizora a zapisovatele.

1. Definujte kritické rozhodnutí a hranice systému.
2. zmapujte požadavky, implementaci, ověření, schválení, provoz a reakci na incidenty.
3. V každé fázi přidejte vlastníka důkazů, vlastníka kontroly a osobu s pravomocí zastavit proces.
4. Dohodněte se na závěru POKRAČOVAT, UPRAVIT nebo ZASTAVIT.
5. Vypracujte dvouminutovou obhajobu založenou na řetězci sledovatelnosti.

[Zdroje]. Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.

Důkladná analýza odpovědnosti propojuje pravomoc s důkazy

[Zdroje]

- Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.
- <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

- Důkladná analýza definuje hranice systému a nežádoucí událost. Rozlišuje přidělené povinnosti, rozhodnutí, za která nese odpovědnost, a právní kontext. Analýza selhání zahrnuje technické, lidské a organizační faktory.
- Kontrolní opatření se zaměřují na prevenci, detekci, zmírnění dopadů a obnovu. Každé kontrolní opatření má kritérium přijatelnosti a odpovědnou osobu. Záznam zachycuje předpoklady, nesouhlasné názory, zbytkové riziko, podmínky monitorování a podmínky pro obnovení provozu.
- Alternativní rozhodnutí mohou být platná, jsou-li jasně stanoveny jejich důkazy, pravomoc a podmínky přezkumu.

Pracovní list: Záznam o odpovědnosti a sledovatelnosti

System, zamýšlené použití a provozní hranice:

Kritické rozhodnutí a nežádoucí událost:

Odpovědné role a osoba zodpovědná za rozhodnutí:.....

Technické, lidské a organizační faktory:

Bezpečnostní kontroly a akceptační testy:

Důkazy o zajištění, protokoly a záznamy o změnách:

[Zdroje] Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.

Pokračování na dalším snímku

Rozhodnutí: POKRAČOVAT REVIDOVAT ZASTAVIT

Zbytkové riziko, oprávnění k zastavení a spouštěč přezkumu:

Incident, náprava a postup pro obnovení:

Nesouhlas nebo alternativní názor:

[Zdroje] Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.

Role Play: Přezkum uvolnění pod tlakem termínu

- Aktualizace dodavatele splňuje kritéria, zatímco terénní test odhalil jeden nevysvětlitelný případ téměř nehody. Vedení chce, aby vydání proběhlo v souladu se smluvním termínem.
- Role: technický manažer, bezpečnostní inženýr, vlastník produktu, dodavatel, zástupce provozovatele a nezávislý recenzent.
- Rozhodněte, jaké důkazy jsou nutné, kdo může vydání odložit, jak se zaznamenává nesouhlas a za jakých podmínek je možné provést omezený test. Uveďte jeden nepřijatelný argument a jeden platný důvod pro schválení.

[Zdroje]

- Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.
- <https://www.nist.gov/itl/ai-risk-management-framework>

Individuální reflexe & Přenos

Napište krátkou
technickou
argumentaci.

1. Které rozhodnutí vyžadovalo jasnou odpovědnost?
2. které kontrolní mechanismy selhaly nebo zcela chyběly?
3. Který důkaz měl změnit rozhodnutí?
4. Kdo měl pravomoc zastavit postup?
5. Dokázal tým identifikovat, vysvětlit a napravit selhání?
6. Jak byste do svého příštího projektu začlenili odpovědnost?

Výstup:

Přeneste toto uvažování do jiné disciplíny a uveďte jednu změnu v rolích, načasování nebo důkazech.

[Zdroje] Ethical Engineer schválila žádost KA220-HED, WP3 Výuka a učení.

Kvíz

1. Jak se liší odpovědnost od povinnosti?
2. Proč mohou dva senzory selhat jako jeden ovládací prvek?
3. V čem spočívá smysl lidského dohledu?
4. Co patří do případu zajištění?
5. Jaké důkazy podporují rekonstrukci incidentu?
6. Proč by aktualizace týkající se bezpečnosti měla vyvolat nové posouzení?
7. Kdo by měl akceptovat zbytkové riziko?

Odpovědi na kvíz a zdůvodnění výuky

1. Odpovědnost zahrnuje odůvodnění, důkazy, přezkoumání a nápravu.
2. Senzory mohou sdílet společnou příčinu, data nebo závislost.
3. Smysluplný dohled vyžaduje autoritu, informace, kompetence a rychlou reakci.
4. Tvrzení, podpůrný argument, důkazy, předpoklady, omezení a verze.
5. Protokoly s verzemi, konfigurace, vstupy, výstupy, alarmy, akce a změny.
6. Změna může zneplatnit předpoklady o rizicích, kontrolní opatření nebo dřívější důkazy.
7. Jmenovaná role s pravomocí, kompetencí a organizačním mandátem.

Kontrola výstupů učení

Než skončíte, ujistěte se, že umíte:

1. Rozlišujte mezi odpovědností, zodpovědností a právní odpovědností.
2. Zmapujte vlastnictví rozhodnutí a kontroly v průběhu celého životního cyklu.
3. Analyzujte příčiny selhání na různých úrovních.
4. Specifikujte opatření pro odolnost, záložní řešení, předání, protokolování, zajištění a řízení změn.
5. Vypracujte sledovatelné rozhodnutí s vyhodnocením zbytkového rizika a revizí.

Důkazy:

Mapa odpovědností, analýza selhání, argumenty pro zajištění kvality a záznam o odpovědnosti.

[Zdroje] Ethical Engineer schválil žádost KA220-HED, WP3 Výuka a učení.



Oznámení o financování a využití ve vzdělávání

Tento otevřený vzdělávací zdroj slouží k podpoře výuky a diskuse. Nepředstavuje právní, regulační, bezpečnostní ani odborné inženýrské poradenství. Při aplikaci jeho obsahu se řiďte platnými zákony, normami, institucionálními zásadami a oborově specifickými pokyny.

Spolufinancováno programem Erasmus+ Evropské unie.





Ethical Engineer

www.ethicalengineer.eu

Děkujeme

Dokončili jste Modul 4



Co-funded by
the European Union