

# Umělá inteligence a její rizika v kontextu ochrany obyvatelstva

Bc. František Duda

---

Diplomová práce  
2024



Univerzita Tomáše Bati ve Zlíně  
Fakulta logistiky a krizového řízení

---

Univerzita Tomáše Bati ve Zlíně  
Fakulta logistiky a krizového řízení  
Ústav krizového řízení

Akademický rok: 2023/2024

## ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: Bc. František Duda  
Osobní číslo: L22411  
Studijní program: N1032A020002 Bezpečnost společnosti  
Specializace: Rizikové inženýrství  
Forma studia: Prezenční  
Téma práce: Umělá inteligence a její rizika v kontextu ochrany obyvatelstva

### Zásady pro vypracování

- Na základě provedené rešerše zpracujte teoretický vstup do dané problematiky.
- Definujte pojem umělá inteligence pro potřeby zpracovávané diplomové práce.
- Identifikujte relevantní rizika spojená s umělou inteligencí pro oblast ochrany obyvatelstva.
- Navrhněte opatření pro snížení míry rizik.

Forma zpracování diplomové práce: **tištěná/elektronická**

**Seznam doporučené literatury:**

1. REESE, Byron. *Čtvrtý věk: inteligentní roboti, myslící počítače a budoucnost lidstva*. Brno: Zoner Press, 2022. ISBN 978-80-7413-490-6.
  2. RUSSEL, Stuart a NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 4. vyd. Hoboken: Pearson, 2020. ISBN 978-0134610993.
  3. WILKS, Yorick A. *Artificial Intelligence: Modern Magic or Dangerous Future?* Hot science (Icon Books). London: Icon Books, 2019. ISBN 978-1785785160.
- Další odborná literatura dle doporučení vedoucího diplomové práce.

Vedoucí diplomové práce: **Ing. Petr Svoboda, Ph.D.**  
Ústav ochrany obyvatelstva

Datum zadání diplomové práce: **1. prosince 2023**  
Termín odevzdání diplomové práce: **26. dubna 2024**

L.S.

---

**doc. Ing. Zuzana Tučková, Ph.D.**  
děkanka

---

**Ing. et Ing. Jiří Konečný, Ph.D.**  
ředitel ústavu

V Uherském Hradišti dne 4. prosince 2023

## PROHLÁŠENÍ AUTORA DIPLOMOVÉ PRÁCE

Beru na vědomí, že:

- diplomová práce bude uložena v elektronické podobě v univerzitním informačním systému a dostupná k nahlédnutí;
- na moji diplomovou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- podle § 60 odst. 1 autorského zákona má Univerzita Tomáše Bati ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- podle § 60 odst. 2 a 3 autorského zákona mohu užit své dílo – diplomovou práci nebo poskytnout licenci k jejímu využití jen s předchozím písemným souhlasem Univerzity Tomáše Bati ve Zlíně, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše);
- pokud bylo k vypracování diplomové práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tj. k nekomerčnímu využití), nelze výsledky diplomové práce využít ke komerčním účelům;
- pokud je výstupem diplomové práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

### Prohlašuji,

- že jsem na diplomové práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.
- že odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou obsahově totožné.

V Uherském Hradišti, dne: 26.04.2024

Jméno a příjmení studenta: Bc. František Duda

.....  
podpis studenta

## **ABSTRAKT**

Diplomová práce se zaměřuje na potenciální rizika umělé inteligence v kontextu ochrany obyvatelstva. Teoretická část pojednává o samotné definici pojmu umělá inteligence, jelikož není pevně odborně stanovena. Dále jsou zde rozebírány vybrané pravděpodobné hrozby přinášené rozvojem umělé inteligence na základě odborných zdrojů. Z nich jsou následně vybrány relevantní hrozby v kontextu práce. Rizika jsou analyzována pomocí jak vlastní metody, tak za pomoci modelů umělé inteligence v podobě veřejně dostupných chatbotů. Účelem je otestovat analytické schopnosti zkoumaných chatbotů a poukázat i na pozitivní přínos. V závěru jsou navržena nápravná opatření, jež mají snížit vliv definovaných hrozeb.

Klíčová slova: Kybernetické hrozby, Malware, Ochrana obyvatelstva, Umělá inteligence

## **ABSTRACT**

The diploma thesis focuses on the potential risks of artificial intelligence in the context of population protection. The theoretical part first focuses on the definition of artificial intelligence itself, as it is not firmly established by experts. It also discusses the probable threats posed by the development of artificial intelligence based on expert sources. These are selected in the analytical section in order to find relevant risks in the context of the work. Risks are analysed both by own method and by artificial intelligence models in the form of publicly available chatbots. The purpose is to test the analytical capabilities of the investigated chatbots and to point out the positive benefits. At the end of the thesis, corrective measures are proposed to reduce the impact of the defined threats.

Keywords: Artificial Intelligence, Cyber Threats, Malware, Population Protection

Chtěl bych zde poděkovat Ing. Petru Svobodovi, Ph.D. za jeho vedení, pomoc a rady při zpracování mé diplomové práce.

„Pokud se očekává, že stroj bude neomylný, nemůže být také inteligentní.“ — Alan Turing

Prohlašuji, že odevzdaná verze bakalářské/diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné. V práci jsem využil nástroje umělé inteligence za účelem poukázat na její případnou nebezpečnost nebo naopak prospěšnost v kontextu práce, otestovat její analytické schopnosti či vygenerovat obrázky pro použití v práci.

**OBSAH**

<b>ÚVOD .....</b>	<b>9</b>
<b>I TEORETICKÁ ČÁST .....</b>	<b>11</b>
<b>1 OCHRANA OBYVATELSTVA .....</b>	<b>12</b>
1.1 ORGÁNY A JEJICH ÚKOLY.....	12
1.2 ZÁVĚR OCHRANY OBYVATELSTVA .....	15
<b>2 UMĚLÁ INTELIGENCE .....</b>	<b>16</b>
2.1 DEFINICE POJMU UMĚLÁ INTELIGENCE .....	16
2.2 PŘÍSTUPY K UMĚLÉ INTELIGENCI.....	17
2.3 ROZDĚLENÍ UMĚLÉ INTELIGENCE .....	18
2.4 STROJOVÉ A HLOUBKOVÉ UČENÍ .....	19
2.4.1 Strojové učení.....	20
2.4.2 Hlubkové učení.....	20
2.5 UMĚLÁ INTELIGENCE A ETIKA.....	21
2.6 ZÁVĚR K DEFINICI UMĚLÁ INTELIGENCE .....	22
<b>3 RIZIKA SPOJENÁ S UMĚLOU INTELIGENCÍ.....</b>	<b>23</b>
3.1 ZHOUBNÉ SELHÁNÍ V RÁMCI SUPERINTELIGENCE.....	23
3.3 AUTONOMNÍ ZBRANĚ .....	26
3.4 CHYBNĚ NAUČENÝ MODEL UMĚLÉ INTELIGENCE.....	27
3.5 NARUŠENÍ FUNKCE UMĚLÉ INTELIGENCE ÚTOČNÍKEM.....	27
3.6 AI A KYBERNETICKÉ ÚTOKY .....	28
3.6.3 Prolomení hesla .....	34
3.6.4 Phishing.....	34
3.6.5 DoS/DDoS útok.....	35
<b>II PRAKTICKÁ ČÁST.....</b>	<b>36</b>
<b>4 IDENTIFIKACE A ANALÝZA RELEVANTNÍCH RIZIK VE VZTAHU KE KONTEXTU PRÁCE.....</b>	<b>37</b>
4.1 VYMEZENÍ RELEVANTNÍCH RIZIK .....	37
4.2 ANALÝZA RIZIK.....	39
4.2.1 Vlastní FMEA .....	39
4.2.2 FMEA analýzy pomocí umělé inteligence.....	45
4.3 ZÁVĚR KAPITOLY .....	58
<b>5 SCÉNÁŘE KYBERNETICKÁCH INCIDENTŮ .....</b>	<b>61</b>
5.1 SCÉNÁŘ KYBERNETICKÉHO ÚTOKU RANSOMWAREM .....	61
5.2 SCÉNÁŘ PROLOMENÍ HESLA POMOCÍ PASSGAN.....	64
5.3 SCÉNÁŘ DEEPFAKE ÚTOKU.....	66
5.4 ZÁVĚR KAPITOLY .....	68
<b>6 ANALÝZA NÁROČNOSTI TVORBY DEEPFAKES .....</b>	<b>69</b>

6.1	VYTVORENÍ DEEPFAKES.....	69
6.2	ZHODNOCENÍ NÁROČNOSTI TVORBY.....	72
6.3	SOFTWARE PRO DETEKCI DEEPFAKU.....	73
6.4	ZÁVĚR ANALÝZY NÁROČNOSTI TVORBY DEEPFAKŮ.....	76
<b>7</b>	<b>OPATŘENÍ PRO MINIMALIZACI RIZIK.....</b>	<b>77</b>
7.1	ANTIMALWARE.....	77
7.2	ZABEZPEČENÍ HESLA.....	78
7.3	ŠKOLENÍ O PROBLEMATICE.....	80
7.4	ZÁLOHA DAT.....	80
7.5	ZÁVĚR K NÁVRHŮM OPATŘENÍ.....	81
	<b>ZÁVĚR.....</b>	<b>82</b>
	<b>SEZNAM POUŽITÉ LITERATURY.....</b>	<b>83</b>
	<b>SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK.....</b>	<b>88</b>
	<b>SEZNAM OBRÁZKŮ.....</b>	<b>89</b>
	<b>SEZNAM TABULEK.....</b>	<b>91</b>
	<b>SEZNAM PŘÍLOH.....</b>	<b>92</b>



## ÚVOD

Umělá inteligence je v dnešní době čím dál více skloňovaným pojmem. Tato technologie se čím dál více dostává do života běžných lidí. Byť se může zdát, že se jedná o poměrně novou věc její vznik či úvahy nad ní lze datovat již od 50. let minulého století ve spojení s prací Alana Turinga. V literárních dílech se objevovala však v určité podobě dříve například v podobě RUR od Karla Čapka, kde byly poprvé pojmenovány humanoidní zařízení jako robot. Mozek takového zařízení lze označit za umělou inteligenci. S rozvojem technologií se takovéto teoretické fantazie stávají skutečností. Za větší boom v dnešní době v tomto ohledu lze označit spuštění ChatGPT pro veřejnost. To rozproudilo debaty zabývající se jeho vlivem na kyberbezpečnost. Mezi negativní vlivy se může řadit nárůst dezinformací či otázky týkající se uměleckých děl a autorských práv. Umělá inteligence se tak dostává do popředí diskusí a vznikají palčivé otázky. Jsou jimi například zda lidem nevezme práci, nebude lidstvo hloupnout či zda nás dokonce nevyhubí. Problematické je i samotné kybernetické prostředí, které se může stát takovým tak zvaným divokým západem. Otázkou těchto rizik, a i případně dalších týkajících se umělé inteligence se zabývá právě tato práce.

Předmětem práce je tak pomocí odborné literatury nalézt vybrané hrozby spojené s umělou inteligencí a vyhodnotit nejzávažnější. Z kontextového hlediska bude pracováno se subjektem ochrany obyvatelstva typu obecní úřad. Jelikož plní úkoly ochrany obyvatelstva je zásadním orgánem při řešení krizí či mimořádných událostí. Jeho disfunkce způsobená některým z rizik v práci by mohlo ohrozit jeho schopnost tyto funkce řádně plnit. Na základě analýzy jsou v práci popsány scénáře popisující jak průběh, tak i předpokládaný dopad rizik. Pro minimalizaci tohoto nebezpečí jsou předneseny doporučení pro realizaci.

## CÍLE A METODY PRÁCE

Stěžejním cílem této práce je vytvořit návrh vhodných opatření vedoucí k minimalizaci definovaných rizik s ohledem na oblast ochrany obyvatelstva.

Díličí cíle, jimiž je dosažen hlavní cíl jsou:

- Zpracováním literární rešerše stanovit teoretické základy práce.
- Identifikovat relevantní hrozby umělé inteligence ve vztahu k ochraně obyvatelstva.
- Na základě identifikace hrozeb pomocí vhodné metody analyzovat jejich dopad
- V souvislosti se zjištěnými poznatky zpracovat scénáře vybraných útoků, u kterých figuruje umělá inteligence.
- Navrhnout vhodné opatření vedoucí ke zmírnění těchto hrozeb.

V rámci diplomové práce byly využity vybrané vědecké metody s cílem dosažení definovaných díličích cílů a cíle hlavního. V první řadě byla využita metoda literární rešerše, díky níž byly popsány jednotlivé definice pojmu umělá inteligence a identifikována rizika, jenž tato technologie dle odborných zdrojů může přinést. V rámci ustanovení definice pojmu byla použita metoda komparace názorů a hypotéz v rámci níž byly porovnávány nalezené formulace a vymezena ta, s níž je dále pracováno. Komparace byla využita i u zkoumání hrozeb umělé inteligence, kde byly u některých sporadických rizik porovnávány názory odborníků. K další metodě, jenž byla použita v práci, se řadí analýza v podobě metody FMEA, jejíž tvorba byla konzultována s pracovníky obce. V praktické části byla tato metoda v kombinaci s metodou komparace využita při porovnání vlastní a vygenerované metody FMEA. Určitá forma analýzy byla zvolena i u zhodnocení náročnosti tvorby Deepfake médií. V tomto kontextu byla použita syntéza, s jejichž pomocí byl výsledek práce ucelen. V kontextu nalezení relevantních hrozeb byla využita deskripce ohraničující zkoumanou oblast rizik, jenž skutečně představuje pravděpodobné a významné ohrožení pro subjekt plnící úkoly ochrany obyvatelstva.

## **I. TEORETICKÁ ČÁST**

## 1 OCHRANA OBYVATELSTVA

Pojem ochrana obyvatelstva je definován dle zákona 239/2000 Sb. Zákon o integrovaném záchranném systému a o změně některých zákonů jako "*plnění úkolů civilní ochrany, zejména varování, evakuace, ukrytí a nouzové přežití obyvatelstva a další opatření k zabezpečení ochrany jeho života, zdraví a majetku*" (Česko, 2000).

### 1.1 Orgány a jejich úkoly

Subjekty, které tyto úkoly plní dle zákona 239/2000 Sb. (Česko, 2000) jsou: Ministerstva a úřady ústřední správy, složky IZS (Integrovaný záchranný systém), orgány kraje, orgány ORP (obec s rozšířenou působností), orgány obce, podnikající fyzické a právnické osoby a Fyzické osoby.

Úkolem ministerstev a úřadů ústřední správy je zabezpečit registr obsahující možné zdroje rizik a provedení analýzy ohrožení. Včetně toho realizují nápravu potencionálních negativních vlivů, které mohou způsobit mimořádnou událost. Dále zabezpečují případné opravy veřejných zařízení nezbytných pro plnění úkolů ochrany obyvatelstva. Nadále rozhodují i o vykonání záchranných a likvidačních prací. To vše v rozsahu stanovené legislativou, kterou zároveň i upravují či vytváří. Nad rámec některé ministerstva mají i další úkoly. Ministerstvo vnitra má uděleny úkoly v oblasti příprav na mimořádné události, ochrany obyvatelstva a integrovaného záchranného systému. Dále zapojuje stát do zahraničních aktivit spojených s poskytováním pomoci při mimořádných událostech (Česko, 2000).

Za tímto účelem dle zákona 239/2000 Sb. (Česko, 2000) plní tyto úkoly:

- sjednocuje postupy všech subjektů ochrany obyvatelstva,
- usměrňuje IZS,
- kontroluje a koordinuje poplachové plány IZS krajů a zpracovává ústřední poplachový plán,
- zpracovává koncepci ochrany obyvatelstva,
- organizuje stavbu a provoz informačních a komunikačních sítí a služeb IZS,
- zabezpečuje a provozuje jednotný systém varování a vyrozumění a definuje způsob vyrozumění o nebezpečí právnické a fyzické osoby,

- zabezpečuje informovanost a školení v oblasti ochrany obyvatelstva a pro přípravu složek IZS zřizuje vzdělávací střediska,
- uplatňuje stanovisko k plánu územního rozvoje a k územnímu opatření o stavební uzávěře z pohledu dodržení zásad ochrany obyvatelstva,
- je dotčeným orgánem při hodnocení projektů společného zájmu energetické infrastruktury a přednáší vyjádření ke stavbě tohoto projektu,
- po projednání s ministerstvem pro místní rozvoj definuje parametry pro stavbu zařízení k plnění úkolů ochrany obyvatelstva a organizuje zřizování zařízení civilní ochrany,
- zabezpečuje záchranné a likvidační práce, pokud mimořádná událost přesahuje území republiky nebo přesahuje hranice jednoho kraje a velitel zásahu vyhlásil nejvyšší stupeň anebo o to požádal velitel zásahu.

Ministerstvo zdravotnictví provádí koordinaci činnosti poskytovatelů zdravotnických služeb. To uskutečňuje, pokud mimořádná událost přesahuje hranice kraje nebo je to z kapacitních či odborných důvodů nutné anebo se kraje nemohou dohodnout na postupu. Ministerstvo dopravy zajišťuje informační systém s celostátní působností pro potřeby záchranných a likvidačních prací v oblasti mobilních zdrojů hrozeb v dopravě (Česko, 2000).

Hlavním úkolem orgánů kraje v této oblasti je zabezpečení připravenosti na mimořádné události a záchranné a likvidační práce. Za tímto účelem usměrňuje součinnost a postupy v otázce ochrany obyvatelstva mezi samosprávnými celky ve své působnosti. Zároveň usměrňuje na úrovni kraje IZS a zpracovává plán k provádění záchranných a likvidačních prací a poplachový plán v rámci oblasti své působnosti. Dále spolupracuje při pracích na povodňovém plánu a uzavírá dohody s územním celkem sousedícího státu. V případě, že zóna havarijního plánování zasahuje více než jedno území obce s rozšířenou působností či zasahuje území sousedního kraje, spolupracuje orgán s dotčenými subjekty na jeho zpracování. V případě druhé zmínění skutečnosti je plán zpracováván úřadem, na jehož území se zdroj rizika nachází. Pro zpracování havarijního plánu kraje a vnějších havarijních plánů je krajský orgán oprávněn k nakládání s informacemi z krizového plánu kraje dle předpisů. Tyto úkoly jsou přeneseně prováděny Hasičským záchranným sborem. Úkolem hejtmana je organizování IZS na krajské úrovni. Dále koordinuje zásah při mimořádné události v případě, že dojde k přesahu působnosti jednoho území obce s rozšířenou

působností a velitel zásahu vyhlásil nejvyšší stupeň anebo byl požádán velitelem zásahu. K tomuto účelu může využít krizový štáb kraje. Nadále koordinuje a kontroluje připravenost podřízených samosprávných orgánů na mimořádné události a schvaluje havarijní plán kraje, vnější havarijní plán a poplachový plán integrovaného záchranného systému kraje. Úkoly obce s rozšířenou působností a starostky takové obce jsou obdobné. Rozdíl je v oblasti působení. Hejtman a starosta ORP jsou povinni při koordinaci záchranných a likvidačních prací informovat Ministerstvo vnitra o průběhu zásahu (Česko, 2000).

Činnost orgánů obcí, ke kterým je oprávněna zřizovat zařízení civilní ochrany v této oblasti jsou:

- organizace připravenosti obce na mimořádné události,
- podílení se na záchranných a likvidačních pracích,
- zabezpečení varování, evakuaci a ukrytí osob,
- hospodaření s majetkem civilní obrany,
- poskytnutí informací a podkladů hasičskému záchrannému sboru ke zpracování havarijních plánů,
- podílení se na zajištění nouzového přežití obyvatelstva,
- vede evidenci a provádí kontrolu staveb civilní ochrany nebo staveb dotčených požadavky civilní ochrany v obci (Město Tábor, © 2024).

Činnosti starosty v oblasti ochrany obyvatelstva jsou

- zajištění varování obyvatelstva,
- organizace evakuace obyvatelstva po dohodě s velitelem či starostou příslušného ORP,
- organizace nouzového přežití obyvatelstva (Město Tábor, © 2024).

Starosta je i oprávněn žádat poskytnutí věcné či osobní pomoci od fyzických či právnických osob (Město Tábor, © 2024).

Povinnosti, které jsou u fyzických osob, podnikajících fyzických a právnických totožné v rámci ochrany obyvatelstva jsou:

- na výzvu od příslušného orgánu poskytnout osobní či věcnou pomoc,

- strpět omezení opatření nutná k překonání mimořádné události a strpět umístění zařízení systému varování a vyrozumění na nemovitostech ve vlastnictví,
- poskytnout potřebné informace mající vliv na mimořádnou událost veliteli zásahu,
- pokud je majitelem objektu civilní ochrany, je povinen při užití dbát na zachování jejich funkčnosti (Česko, 2000).

Právnícké osoby a podnikající fyzické osoby jsou v případě zahrnutí do havarijních plánů povinni bezúplatně poskytnout potřebné informace a zajistit opatření pro své dotčené zaměstnance. V případě, že jsou provozovatelé školských, sociálních nebo obdobných zařízení nebo jsou poskytovateli zdravotních služeb, jsou povinni uzpůsobit prostory pro případný výdej ochranných osobních pomůcek během mimořádné události. V případě, že dojde k mimořádné události v souvislosti s jejich provozem, jsou povinni poskytnout součinnost při zásahu a zabezpečit své zaměstnance. V případě, že u těchto osob dojde k havárii, jsou povinni provádět záchranné a likvidační práce, ohlásit tuto havárii příslušným orgánům, podílet se na varování a poskytnout informace o skutečnost mající zásadní vliv na bezpečnost zásahu veliteli. Dále jsou povinni spolupracovat při odstraňování mimořádné události, uhradit náklady spojené s akcemi IZS, zabezpečit asanaci a zneškodnění odpadů a spolupracovat na dokumentu o záchranných a likvidačních pracích (Česko, 2000).

## 1.2 Závěr ochrany obyvatelstva

Kapitola se zabývá zejména orgány a jejich úkoly v rámci problematiky. Je vycházeno hlavně ze zákona 239/2000 Sb., který úkoly přímo definuje. Předmětem této části bylo přiblížit tyto úkoly a poskytnout náhled k další práci ve vztahu k umělé inteligenci. Lze tak lépe nahlédnout na otázku, zda je umělá inteligence hrozbou. V práci bude pracováno zejména s objektem typu obecní úřad. Ten má svou roli v rámci ochrany obyvatelstva a při snížení jeho provozuschopnosti by mohlo být ohroženo i jeho naplňování úkolů ochrany obyvatelstva. Proto byl pro kontext práce vybrán právě tento subjekt.

## 2 UMĚLÁ INTELIGENCE

Umělá inteligence (anglický Artificial Intelligence, zkráceně AI) je čím dál více rozšířená a diskutovaná problematika. Je čím dál více rozšířená a jejím rychlým rozvojem se zlepšují její schopnosti. Avšak co je to vlastně umělá inteligence? Na tuto otázku, ale i další teoretické základy v rámci problematiky jsou řešeny v této části práce.

### 2.1 Definice pojmu umělá inteligence

Definovat samotný pojem Umělé inteligence je samo o sobě problematické. Už nazvat samoučící se algoritmy inteligencí není přesné, neboť na nich není nic inteligentního a ani v podstatě nevykazují žádnou inteligenci. Tuto vlastnost jsou schopny pouze imitovat či napodobit. Avšak na druhou stranu je nutné si i uvědomit, že i inteligence nemá pevně danou definici a je to pojem rozporuplný. Kvůli tomu tak není snadné ustanovit jednotnou definici. V následujících odstavcích jsou uvedeny některé definice dle odborných zdrojů.

Jeden ze zakladatelů oboru zabývajících se umělou inteligencí Marvin Lee Minsky (Mařík et al., 2013) jí definoval tak že "*umělá inteligence je věda o vytváření strojů nebo systémů, které budou při řešení určitého úkolu užívat takového postupu, který – kdyby ho dělal člověk – bychom považovali za projev jeho inteligence*". Umělá inteligence je zde tak popsána jako vědní obor zkoumající lidské chování a reakci v prvním kroku a ve druhém kroku se zabývá snahou tyto lidské projevy opakovat pomocí počítačového zařízení (Mařík et al., 2013).

Dle Johna Zerilliho (ZERILLI et al., 2020) je definice zaměřená na člověka, jak je tomu v případě Minského definice funkční, ale zároveň potřebuje menší úpravu. Důvodem je, že předpokladem pro vymezení AI se soustředí na imitaci lidské činnosti. V dnešní době se však používají i v mnoha méně známých oblastech pro plnění příkazů, kde překonávají člověka svou rychlostí a rozsahem. Příkladem použití může být u vysokofrekvenčního obchodování na burze anebo v internetových vyhledávacích (ZERILLI et al., 2020).

Na stránkách Zpravodajství Evropského parlamentu je umělá inteligence popsána jako „*schopnost strojů napodobovat lidské schopnosti, jako je uvažování, učení se, plánování nebo kreativita*“ (Evropský parlament, © 2021). Formulace dle této stránky je spíše obecnějšího rázu a určena pro širokou veřejnost. V rámci organizace Evropské Unie byla vytvořena i jiná formulace ustanovená v dokumentu Evropské komise Sdělení komise Evropskému parlamentu, Radě, Evropskému hospodářskému a sociálnímu výboru a Výboru regionů – Umělá inteligence pro Evropu. Zde je vymezena jako "*systém vykazující*



*inteligentní chování v podobě vyhodnocování svého okolí a následného rozhodování či vykonávání kroků – s určitou mírou autonomie – k dosažení konkrétních cílů "* (Evropská Unie, © 2018).

V dokumentu Spojených států amerických National Artificial Intelligence Initiative Act of 2020 byla AI ustanovena jako strojově založený systém, jenž je schopen vytvářet předpovědi, doporučení či rozhodnutí ovlivňující virtuální anebo v případě implementace reálné prostředí dle vstupních požadavků anebo cílů, jenž jsou definovány člověkem (USA Government, © 2020).

## 2.2 Přístupy k umělé inteligenci

V rámci problematiky, co je umělá inteligence anebo samotná podstata tohoto pojmu rozděluje i samotnou vědeckou komunitu zabývající se tímto tématem. V rámci toho jsou směry jednotlivých přístupů rozděleny do čtyř kategorií vycházející z kombinace dimenzí: člověk vs racionalita (ve smyslu provádět akce matematicky správně) a myšlení vs chování. Přístupy lze rozčlenit následně:

- Lidské chování – na umělou inteligenci je nahlíženo jako na zařízení či systém napodobující chování člověka,
- Lidské myšlení – založena na kognitivním modelování. Jedná se o obor zabývající se modelováním procesu lidského myšlení s cílem pochopit, jak lidé řeší problémy a převést tento proces do softwarové podoby. Nejde tak jen o to dostatečně imitovat člověka, aby oklamal lidského hodnotitele, ale jednal prokazatelně jako lidská bytost,
- Racionální myšlení – založeno na zákonech myšlení. V rámci tohoto směru je tvorba inteligentní umělé inteligence založena na logice. Generování výstupu této technologie by měla být řízena matematickými definicemi a znalostmi o fungování světa, jenž jsou jisté. Stav jistoty je těžko dosažitelný, jelikož není možné stanovit přesná pravidla jako je tomu například u aritmetiky. Pro umožnění práce s nejistými daty je využívána teorie pravděpodobnosti snižující zmíněnou nejistotu. Díky tomu je tak možné vytvořit model racionálního myšlení jdoucího od hrubých vjemových informací pro porozumění, jak funguje svět až k predikci budoucnosti. Co však není možné dosáhnout je vytvoření racionálního chování, a proto racionální jednání není schopné fungovat samo o sobě,

- Racionální chování – přístup racionálního agenta. Agent je zařízení schopné svými senzory vnímat okolí a pomocí akčních členů jej ovlivňovat. Racionální agent je pak autonomní systém či zařízení provádějící akce co nejefektivněji s ohledem na dostupné informace. Příkladem použití je u autonomních robotů vysílaných pro průzkum neznámého prostředí. Výhodou tohoto přístupu je fakt, že racionální jednání je matematicky definováno a lze tak získat měřitelné výstupy což u napodobování lidského chování či myšlení není zcela možné (Russell, Norvig, 2020).

### 2.3 Rozdělení umělé inteligence

Dle předchozí části, zabývající se pojmem Umělá inteligence je patrné, že jde o širší problematiku a tuto technologií lze najít v rozličných oborech. Podle určitých parametrů, schopností anebo určení je AI rozdělována do několika typů. Na základě schopností lze dělit na:

- Umělá úzká inteligence (ANI) – jedná se o algoritmus či stroj, který dokáže vyrovnat se anebo překonat člověka pouze v konkrétních případech,
- Umělá obecná inteligence (AGI) – tato úroveň strojové inteligence je schopna vykonat všechny úkoly stejně jako člověk (De Spiegeleire et al., 2017).
- Umělá superinteligence (ASI) – jakýkoliv intelekt, jenž je schopen ve všech oblastech zásadně překonat kognitivní schopnosti člověka. V podstatě cokoliv je člověk schopen učinit, tento intelekt je to dokáže provádět také, a navíc i líp. Jedná se tak o poslední vývojovou fázi umělé inteligence. Tuto inteligenci lze rozdělit do tří typů:
  - Rychlostní superinteligence – tento druh intelektuálního stroje je schopen provádět stejné úkony jako člověk, avšak převyšuje jej svou rychlostí,
  - Kolektivní superinteligence – pro dosažení takové úrovně intelektu dochází k úzkému spojení několika menších inteligencí do jedné,
  - Kvalitativní superinteligence – stejně rychlá jako lidská mysl, ale co se týče kvalitativního vyjádření je chytrější (Bostrom, 2016).

Podle způsobu, jak stroj či systém funguje a jak využívá své schopnosti učít se pro reakci, interakci a zpracování dat je dělena:

- Reaktivní umělá inteligence – nejzákladnější typ AI, jenž je schopen pouze generovat předpokládané výstupy dle předaných vstupů. Nedokážou tak ukládat data a učit se z předešlých procesů a na identické situace reagují vždy stejně a nejsou schopny provádět nic víc než konkrétní činnosti, ke které byly naprogramovány,
- Umělá inteligence s omezenou pamětí – na základě pozorování situací či akcí v průběhu času, ale i dat z minulosti a předem naprogramovaných informací učit se a na základě těchto dat provádět akce. Na rozdíl od předchozí má již určitou paměť, ta je ale pouze krátkodobé,
- Teorie mysli umělé inteligence – pojem na základě termínu z psychologie. Teorie mysli označuje schopnost umělé inteligence oproti předchozím chápat a pamatovat si emoce, následně je i replikovat a ovlivňovat. Na základě těchto emocí je pak schopna upravovat své chování. Tento typ je zatím pouze teorií,
- Umělá inteligence schopná sebeuvědomění – typ AI schopen mít své vlastní potřeby, touhy a emoce jako tomu je u člověka a uvědomovat si tyto skutečnosti i u ostatních. Tento koncept nebyl zatím vyvinut a nejsou k dispozici ani technologie pro tento typ (The Conversation Media Group Ltd, © 2010–2023).

## 2.4 Strojové a hlubkové učení

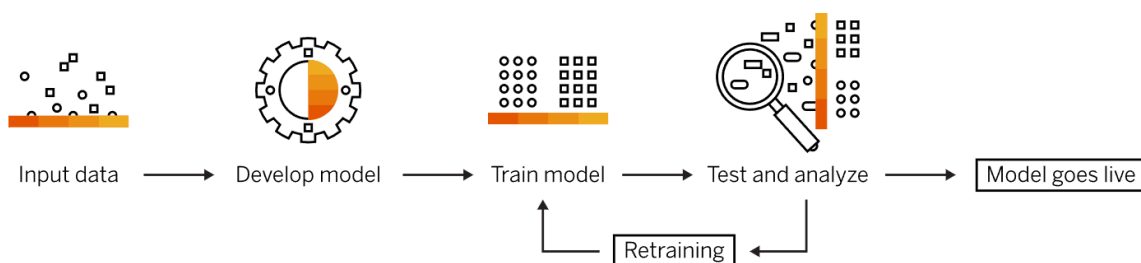
Jednou ze zásadních vlastností Umělé inteligence v moderním provedení by měla být schopnost učit se. V rámci toho byly ustanoveny uvedené pojmy, jež jsou rozebrány níže. Vztah mezi nimi je zobrazen na Obrázku 1.



Obrázek 1 Vizualizace vztahu mezi jednotlivými pojmy (Umělá inteligence vs. Homo sapiens adaptabilis, © 2023).

### 2.4.1 Strojové učení

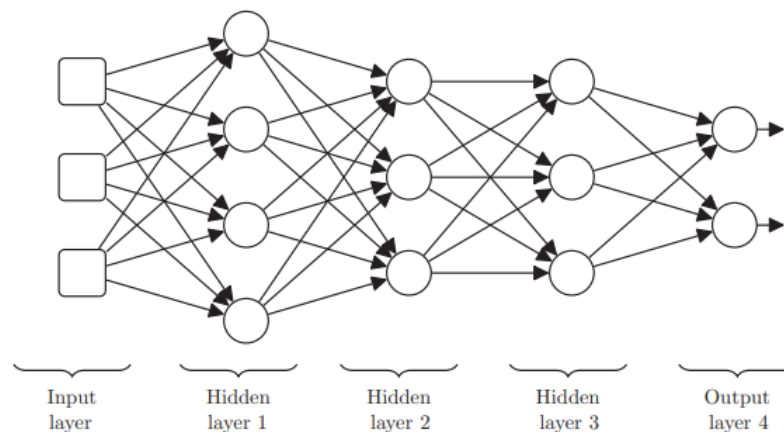
Strojové učení, označené v anglickém jazyce Machine Learning, označuje schopnost stroje na základě vstupních dat či zkušeností získaných během provozu upravovat své výstupy. Stroj se pomocí tohoto algoritmu učí tak, že algoritmus vezme velký objem vstupních i výstupních dat. Následně v nich hledá určité podobnosti či vzorce. Díky tomu je pak zařízení schopno se vycvičit k autonomnímu rozhodování anebo generování doporučení. Stroj se tak vesměs učí podobně jako člověk. Po provedení dostatečného počtu školicích cyklů a úpravách algoritmu získává zařízení dovednost k předvídaní. Pro ověření přesnosti předpovědi stroje je nutné provést komparaci vytvořených výstupů se známými výsledky. Na základě této analýzy je pak možné upravit algoritmus. Díky tomu se zlepšuje jeho predikce (Helm et al., 2020). Na Obrázku 2 je vizuálně znázorněn proces vzdělávání stroje.



Obrázek 2 Proces strojového učení (SAP, © 2023).

### 2.4.2 Hlubkové učení

Hlubkové učení, anglicky Deep Learning je principově podobné jako strojové učení. Hlubkové učení však využívá tak zvanou umělou neuronovou síť fungující podobně jako lidský mozek. Neuronová síť má několik vrstev, jimiž jsou vstupní a výstupní vrstva a několik tak zvaných skrytých vrstev. Čtverečky ve výstupní vrstvě znázorňují paměť, do níž jsou předkládány vstupy síti. Zde nedochází k transformaci vstupní informace. Kolečka ve skrytých vrstvách pak představují neurony, jenž následně provádí nadefinovanou činnost. Ta spočívá v přijmutí několika hodnot jako vstup a následném převedení na výstupní hodnotu. Šípky ukazují směr toku informací v neuronové síti. Neuronová síť při sebevzdělávání využívá strategii rozdělení a panuj. Každý neuron se učí jednodušší funkce a celková funkce vzniká kombinací těchto jednodušších. Na Obrázku 3 je pak zobrazeno schéma umělé neuronové sítě (Kelleher, 2019).



Obrázek 3 Jednoduché schéma umělé neuronové sítě (Kelleher, 2019).

## 2.5 Umělá inteligence a etika

V rámci vývoje zejména obecné inteligence nastává otázka jejího etického rozhodování. Obecná inteligence v pravém smyslu slova není ještě vyvinuta a odborná veřejnost se nedokáže úplně shodnout, kdy k jejímu vytvoření dojde. Nicméně pro lepší kontext k termínu umělá inteligence je zde zahrnuta problematika etického jednání této technologie. Jedním z požadavků na obecnou inteligenci je, aby se AI chovala lidsky i eticky (Reese, 2022).

Etické zákony pro roboty představil již Isaac Asimov v roce 1950 v knize Já, robot. I když jde o zákony robotiky, lze je uplatnit i na obecnou inteligenci, jejíž součástí budou i roboti. Třemi zákony robotiky dle Asimova (2012) jsou:

1. *"Robot nesmí ublížit člověku nebo svou nečinností dopustit, aby člověku bylo ublíženo"*
2. *"Robot musí uposlechnout příkazy člověka, kromě případů, kdy tyto příkazy jsou v rozporu s prvním zákonem. "*
3. *"Robot musí chránit sám sebe před zničením, kromě případů, kdy tato ochrana je v rozporu s prvním nebo druhým zákonem. "*

Později byl doplněn i nultý zákon ve znění: *"Robot nesmí ublížit lidstvu nebo svou nečinností dopustit, aby lidstvu bylo ublíženo"* (Asimov, 1993).

Jenže tyto zákony mají své nedostatky. Například autonomní vozidlo je donuceno se rozhodnout mezi sražením chodce a strhnutím vozu proti sloupu. Obě varianty by tak

vyústily až ve smrt člověka. Vůz je tak nucen jít proti prvnímu zákonu robotiky tak či tak. Tvorba etických kodexů pro obecnou umělou inteligenci bude značný oříšek, a to ze tří zásadních důvodů. Prvním je fakt, že odborníci se na jednotném kodexu neshodnou. Dalším problémem je i vývoj morálních zásad. Během lidské evoluce se to, co je morálně správné a co špatné neustále měnilo. Například v době průmyslové revoluce byla dětská práce standardem, ale dnes je již z etických důvodů zakázána. Tyto morální zásady se budou bezesporu rozvíjet i nadále. A posledním problémem je skutečnost, že morální zásady jsou do jisté míry individuální, plně specifických výjimek a případů. Nelze je tak do jisté míry převést do vzorců a zákonů. Celková problematika etiky byla popsána organizací Future of Life Institute tak, že rozhodnutí vytvořená určitými systémy umělé inteligence jsou založené na důsledcích, avšak morální rozhodnutí jsou daleko složitější. Na toto rozhodnutí má však i vliv ustanovená práva, minulé činy, záměry, motivy a role a další aspekty spojené s morálkou. Ty se však ještě nepodařilo do systému zahrnout (Reese, 2022).

## 2.6 Závěr k definici Umělá inteligence

Náplní této kapitoly bylo ustanovit pojem Umělá inteligence pro potřeby práce. Jak bylo zmíněno samotný pojem nemá zakotvenou pevnou definici a je proto problematické nastavit hranici rozdělující co je a není umělá inteligence. Pro potřeby práce je brána umělou inteligencí i úzká AI. Ta je využita u chatbotů, s nimiž je dále pracováno. Dále jsou brány v potaz i rizika, u nichž je využívána právě úzká umělá inteligence. Dalším faktorem, proč je počítáno i s takovouto nižší úrovní umělé inteligence je skutečnost, že obecná AI a Superinteligence jsou ještě ve vývojové fázi. Nejsou tak aktuálním rizikem jako v případě úzké AI, i když určité nebezpečí představují a je zmíněné i níže. V rámci dalšího dělení bude pracováno převážně s Umělou inteligencí s omezenou pamětí. Důvodem je skutečnost, že Reaktivní AI je již v rámci pokroku minulostí. Dnešní systém a zařízení pracují s umělou inteligencí na vyšší úrovni, než je reaktivní. Stále je však u některých zařízení využívána. Další jsou opět brány pouze v teoretické rovině. Z vypsanych definic bude pro potřeby práce primární zejména stanovena Evropskou unií. Ta tedy zní: „*schopnost strojů napodobovat lidské schopnosti, jako je uvažování, učení se, plánování nebo kreativita*“ (Evropský parlament, © 2021). Důvodem jsou skutečnosti, mezi něž se řadí jednoduchá dostupnost v rámci ochrany obyvatelstva pro občany a fakt, že je stanovena expertním týmem pro legislativu. O té se pak odvíjí i legislativa České republiky.

### 3 RIZIKA SPOJENÁ S UMĚLOU INTELIGENCÍ

Nové technologie přináší spousty pozitivních efektů a zlepšují či ulehčují lidský život. Avšak přináší s sebou i spousty výzev a rizik, které naopak mohou člověku uškodit. Jinak tomu není u umělé inteligence. V rámci této problematiky existuje několik odborných pohledů na věc či přímo analýz. Náplní kapitoly je tak rozebrat vybraná rizika definovaná odbornou literaturou anebo samotnými institucemi.

#### 3.1 Zhoubné selhání v rámci superinteligence

Zhoubné selhání je riziko může představovat konec lidské existence. Jeho hlavním rysem je neopakovatelnost této chyby. Pokud k němu dojde, bude to pouze jednou nebo vůbec. Další charakteristickou vlastností je, že k tomu dojde pouze u úspěšného projektu. Pouze v tomto případě je totiž vybudováno natolik silné inteligentní zařízení, aby zvládlo svými špatnými rozhodnutími a chováním ohrozilo existenci lidstva. Příkladem zhoubného selhání může být typově:

- Zvrácené splnění cíle – umělá superinteligence plní úkoly ve své úspěšně, avšak ne v prospěch zadavatele. Příkladem může být stanovení konečného cíle udělat zadavatele šťastným. Stroj to však provede tak, že nahraje mozek člověka do digitální podoby a zde ho pomocí digitální drogy rozveselí. Pro stroj je tato varianta jednodušší než hledat individuální možnost rozveselení. I přes zdánlivou bezpečnost a nevinnost požadavku může stroj ohrozit lidskou existenci,
- Přesycení infrastruktury – hrozba, při níž by agent mohl přeměnit planetu na infrastrukturu potřebnou k naplnění zadaného cíle. Příkladem je zadání vyrob co nejvíce sponek. Zařízení organizující výrobu začne vyrábět neúnosné množství sponek, až celou Zemi přemění na sponky. Riziko zde spočívá z přemíry továren vytvořených k naplnění cíle. V případě zadání omezeného počtu sponek by se však nic nevyřešilo. Agent by při své inteligenci chtěl zajistit nulovou pravděpodobnost neúspěchu. V určitém bodě by přestal vyrábět a začal budovat obrovské množství kontrolních zařízení. Stroj totiž nikdy nebude chtít připustit, aby jakkoliv selhal a bude proto vytvářet infrastrukturu ke snížení rizika neúspěchu. Jenže takové masivní budování by mohlo zničit planetu a i člověka (Bostrom, 2016).

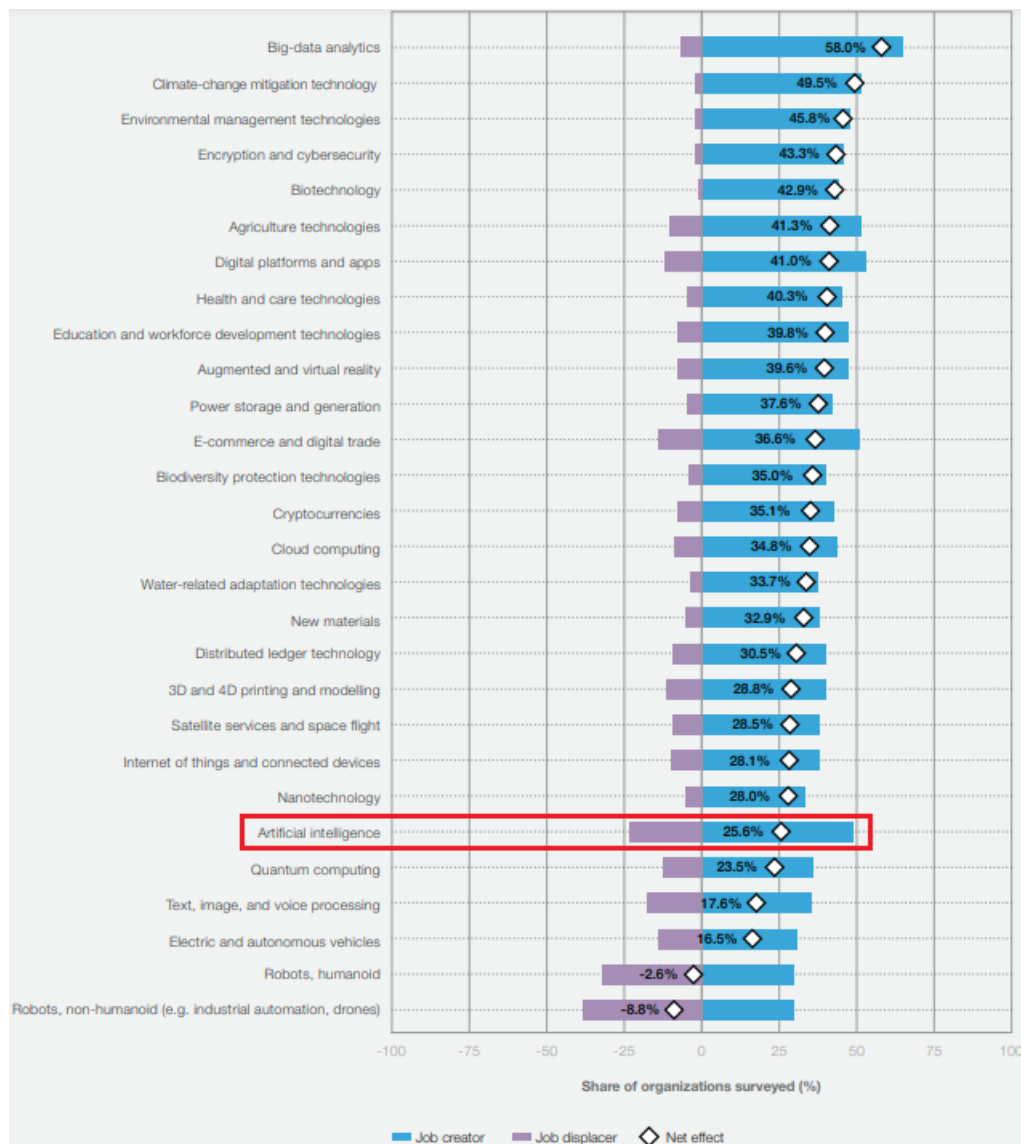
V rámci této problematiky byl vytvořen průzkum v podobě dotazníku předaný výzkumníkům, jež publikovali na konferencích Neural Information Processing Systems

(NeurIPS) a International Conference on Machine Learning (ICML) v roce 2015. Data celkem poskytlo 352 odborných pracovníků v oboru. V dotazníku byla položena otázka, zda strojová inteligence na vysoké úrovni, v podstatě superinteligence bude mít v dlouhodobém horizontu pozitivní nebo negativní dopad na lidstvo. Tato pravděpodobnost byla hodnocena na pětibodové škále a výsledkem je, že s mediánem pravděpodobnosti 25 % bude pozitivní dopad a s 20 % pravděpodobností velmi pozitivní. Na druhé straně je 10% pravděpodobnost špatného výsledku a 5% pravděpodobnost katastrofálních následků (Grace et al., 2018).

### **3.2 Automatizace a růst nezaměstnanosti**

Dle dokumentu The Future of Jobs Report 2023 od Světového ekonomického fóra by umělá inteligence neměla způsobit rapidní nedostatek pracovních míst v intervalu 2023 až 2024. Dle jejich výzkumu je bilance úbytku a příbytku pracovních míst v kladných hodnotách (The Future of Jobs Report 2023, © 2023). Graf zobrazující tuto skutečnost včetně všech ostatní očekávaných vlivů jednotlivých technologií na pracovní místa pro porovnání je na Obrázku 4.





Obrázek 4 Očekávaný dopad zavedení technologií na pracovní místa (The Future of Jobs Report 2023, © 2023).

Nedostatek pracovních pozic způsobený AI by neměl být žádnou hrozbou. Objevují se však jiné problémy. Problematické je možnost datového kolonialismu a schopnost kvalifikovat stávající pracovníky na kvalifikovanější práce. Pojem datový kolonialismus označuje problém, kdy společnosti s využitím modelů AI mohou získat na trhu převahu. Díky patentům a datové převaze by získali dominantní postavení a zamezili konkurenci vytvářet vlastní modely. Tím pádem by se tyto dominantní společnosti stali monopoly. Další problémem je možná restrukturalizace pracovní nabídky. S tím by vznikla potřeba rekvalifikace a zvýšení nároků na kvalifikaci prací. Rekvalifikace je však velmi problematická, neboť ne všechny pracovníky lze vyškolit pro náročnější a více kvalifikované práce. Rekvalifikační kurzy navíc nedopadají pro pracovníky vždy úspěšně. Příkladem je

pokus společnosti LMC requalifikovat v roce 2016 dvacet pět pracovníků z ostravské OKD na programátory. Úspěšně projít zvládli jen čtyři a celkem tři z nich na této pozici začali pracovat. Jenže s rozvojem umělé inteligence budou stroje schopny zastávat i náročnější profese. Než stroje nahradí lidské pracovníky ve všech oborech, mohla by existovat symbióza člověka a umělé inteligence. Tato technologie by byla podporou při práci a rozhodování člověka. Díky tomu by se některé profese mohly zjednodušit (Zandl, 2022).

### 3.3 Autonomní zbraně

Pojem autonomní zbraně je definován jako zbraně, které po aktivaci dokážou vybrat a zasáhnout cíle bez dalšího lidského zásahu. Tyto zbraně mohou být ve formě dronů, kvadrokoptéry s velikostí od 3 centimetrů až po 1 metr, útočných vozidel, a i nákladáku či vyzbrojených lodí bez posádky (Russell, 2022).

Jeden z nejvíce diskutovaných problémů u autonomních zbraní je schopnost AI rozpoznat ozbrojeného vojáka a civilistu. Dle některých odpůrců autonomních zbraní není umělá inteligence natolik vyspělá, aby tohle bezpečně dokázala. Kvůli tomu by docházelo k válečným zločinům. Yorick Wilks však publikuje protiargument, že na podobném principu pracují autonomní vozy. Ty totiž jsou schopny a budou ještě lépe rozpoznat cyklistu od chodce. Podobně by to mohlo fungovat i u zbraní. Avšak nastává i otázka, jak moc velkou autonomií budou mít a zda nebudou operovat pouze pod vedením člověka. Další otázkou je, zda vůbec budou takové zbraně potřeba. Bojiště by se časem mohlo přemístit do kyberprostoru. Vlády investují do kybernetické války zajišťující schopny srazit jiný stát na kolena vyřazením a ochromením infrastruktury bez fyzického útoku (Wilks, 2019).

Stále však nastává jedna problematická otázka, zda je morálně v pořádku přenechat rozhodování o zabíjení lidí na umělé inteligenci. Z tohoto důvodu vznikla iniciativa snažící se o zákaz plně autonomních zbraní. Během diskuse v roce 2019 iniciované Maxem Tegmarkem mezi odpůrci a obhájci této technologie dospěly obě strany ke kompromisu zvažující omezení minimální velikosti a hmotnosti. Malé zbraně tohoto typu jsou kvůli své velikosti a snadné výrobě ve velkých počtech velmi nebezpečné. Omezení by mohlo snížit rizikovost autonomních zbraní. Zákaz těchto zbraní je zdá se zatím v nedohlednu zejména kvůli Spojeným státům a Ruskem, kteří se zákazem nesouhlasí. V tom jsou podporovány Velkou Británií, Izraelem a Austrálií. Důvodem je zejména z jejich pohledu zbytečnost zákazu a snaha mít vlastní převahu ve zbrojení (Russell, 2022).

### 3.4 Chybně naučený model umělé inteligence

Samotný model umělé inteligence může být nechtěně zkreslen ve fázi učení. Při učení jsou umělé inteligenci předávány rozsáhlé datové soubory. Ty jsou natolik rozsáhlé, že je nelze vždy stoprocentně zkontrolovat. Může tak obsahovat chyby, dezinformace anebo jiné eticky pochybné informace či neaktuální data. Při generování se tyto skutečnosti mohou promítnout do výstupu. Dále je tu i skutečnost, že dané výstupy si model vymyslí. S přiložením citace text vypadá pravdivě, avšak ve skutečnosti je nesprávně interpretován. Může vzniknout i problém špatného pochopení příkazu ve fázi učení. Model pak nesprávně vyhodnotí příkaz a například místo odeslání emailu pořídí akcie (Bundesamt für Sicherheit in der Informationstechnik, © 2024).

### 3.5 Narušení funkce umělé inteligence útočníkem

Jelikož je umělá inteligence implementována u komerčních či vojenských aplikací existuje určitá šance, že dojde k pokusu o narušení její funkce. Jsou dva způsoby, jak lze fungování této technologie poškodit. Těmi jsou útok na vstup a otrávení dat (Comiter, 2019).

U útoků na vstup s jedná se o změnu vstupních informací předávaných umělé inteligenci ve prospěch útočníka. Stačí například změnit pouze drobný pixel předávaný AI na vstupu a může dojít k velkým změnám ve výsledku. Útoky na vstup mohou mít formu digitální nebo fyzickou a být rozpoznatelné či nerozpoznatelné pro lidské oko. Příkladem v případě fyzické změny člověkem postřehnutelné může být cedule STOP polepená páskami. Tato změna lze člověkem zpozorovat. Jenže pro zařízení to není nepodstatné a může kvůli tomu špatně vyhodnotit situaci. K útokům, jež jsou pro člověka zrakově nepostřehnutelné lze uvést útok digitálním prachem. Ten má za cíl pozměnit malinké části obrazce, aniž by člověk zpozoroval změnu. Na obrázku tak pro člověka může být pes, avšak kvůli změnám v pixelech je rozpoznán umělou inteligencí jako kočka (Comiter, 2019).

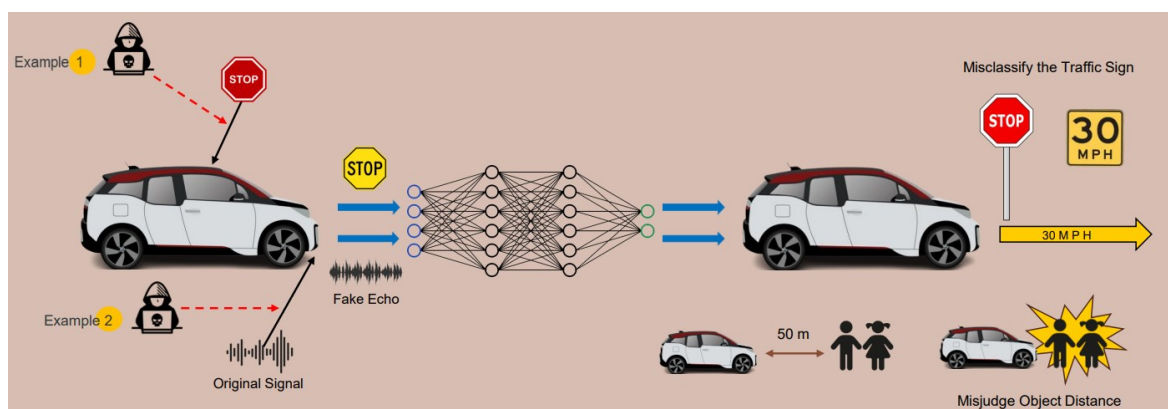
Otrávení dat má za cíl účelným pozměněním dat způsobit disfunkci modelu umělé inteligence. Proto se tento útok nejčastěji provádí během fáze učení a do provozu je implementována již otrávený model. Příkladem může být armáda, kde lze AI využít pro kontrolu vzdušného prostoru. Nepřátelský subjekt může otrávit toto zařízení tak aby nemohlo rozpoznat určitý typ bojových letadel. Existují tři typy tohoto útoku:

- Otrávení datové sady – poskytnutá vstupní data pro učení jsou pozměněna a kvůli tomu se AI naučí nesprávně vyhodnotit vstupy,

- Otrávení algoritmu – využívána nejčastěji u federativního učení. Toto učení probíhá formou malých modelů přímo na zařízení uživatele a následně jsou spojovány do jednoho. Pokud je uživatelem hacker může svými daty tento jeden model otrávit a v kombinaci s ostatními je poškozen výsledný algoritmus,
- Otrávení modelu – útočník vymění stávající model za upravený otrávený model. K tomuto útoku je potřeba aby se útočník dostal do systému, němž je AI model uložen a vyměnit celý či pouze části dat (Kuzlu et al., 2021).

Hlavními opatřeními pro minimalizaci těchto útoků je zejména kontrola vhodných dat. (Kuzlu et al., 2021).

Na Obrázku 5 je znázorněn příklad obou typů útoku u autonomního vozidla (Casey, © 2024).

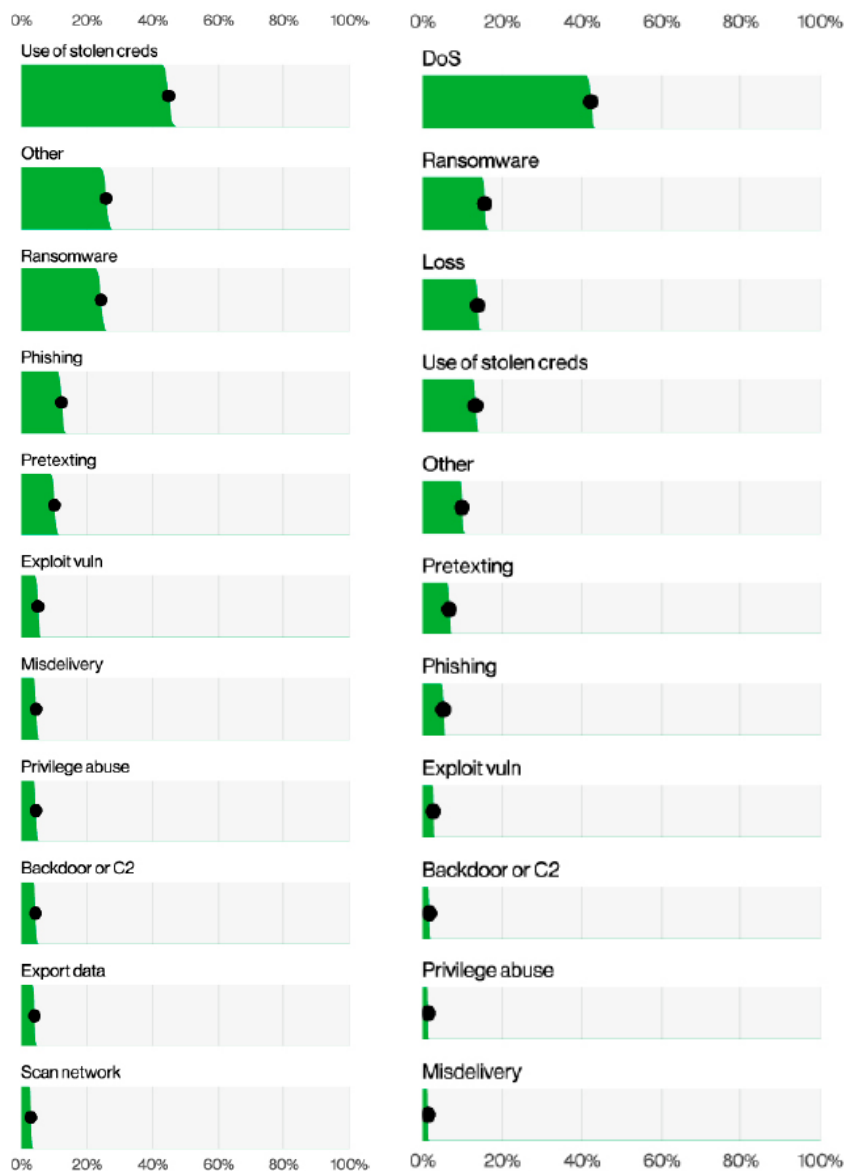


Obrázek 5 Útok na AI u autonomního vozu (Casey, © 2024)

U Example 1 jde o úpravu značky STOP, respektive její barvy. Kvůli takové záměně dojde ke špatnému vyhodnocení systémem vozidla. Vůz tuto značku vyhodnotí jako omezení rychlosti, nezastaví a způsobí nehodu. Svou povahou se jedná o útok na vstup. Example 2 pak reprezentuje otrávení dat, kdy hacker ve fázi vývoje pozmění schopnosti vozidla rozpoznat vzdálenost od překážky. Tím může dojít ke špatnému vyhodnocení a srážce s chodcem (Casey, © 2024).

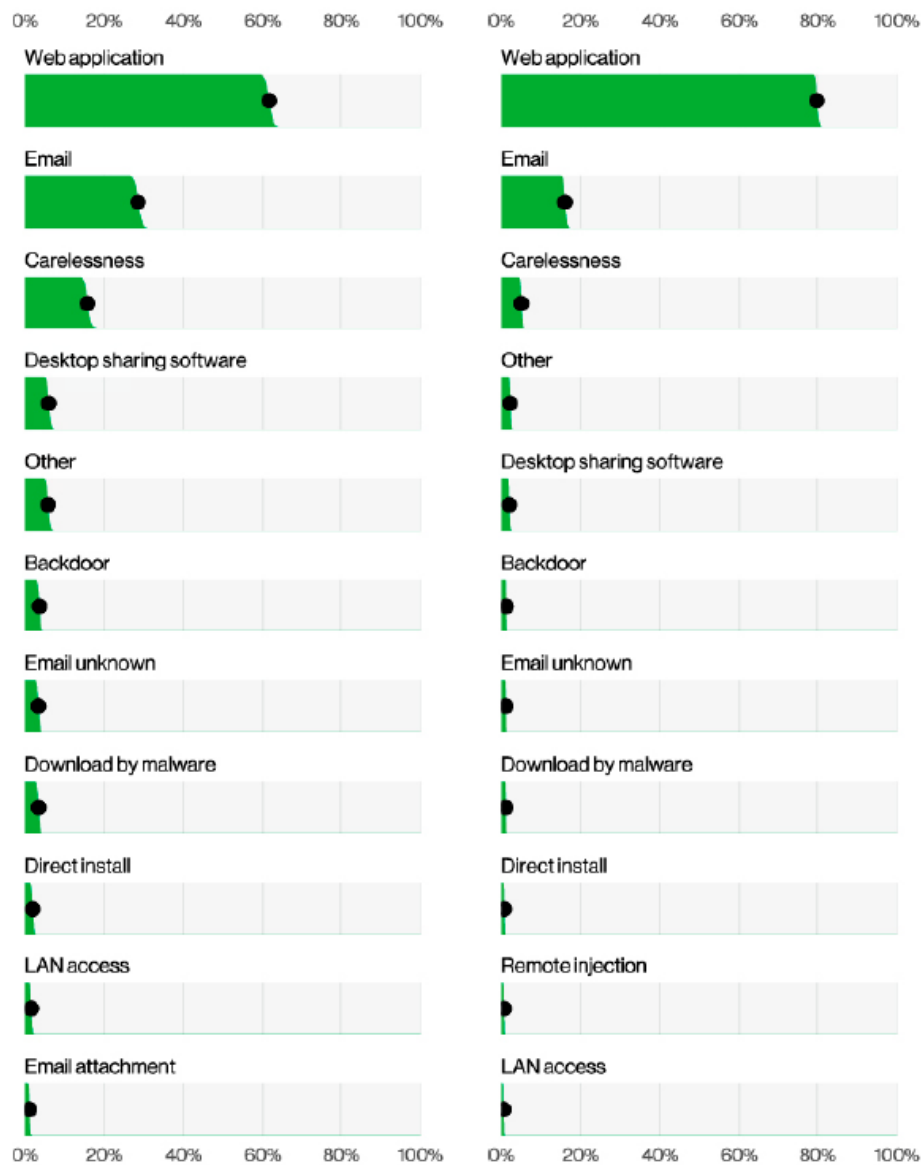
### 3.6 AI a kybernetické útoky

Kybernetické útoky či kybernetické incidenty jsou závažným rizikem pro společnosti i jednotlivce. Na základě analytické zprávy Verizon 2023 Data Breach Report bylo za rok 2023 zaregistrováno celkem 16,312 incidentů a 5199 datových úniků. Ve 24 % případů kybernetických incidentů figuroval ransomware neboli vyděračský software. Na Obrázku 6 je zobrazeny nejčastější akce útočníků vedoucí k datovým únikům (Verizon Communications, © 2023).



Obrázek 6 Nejčastější druhy akcí při narušení dat (Verizon Communications, © 2023)

Největší zastoupení má tak použití ukradených dat, ostatní, ransomware, phishing a pretexting jako jedna z forem sociálního inženýrství. Nejvíce kybernetických útoků vzniká v rámci DoS útoku, ransomware, ztrátě dat a použití ukradených dat. Jde o závažné problémy, které by mohla ještě umocnit právě technologie umělé inteligence. Obrázek 7 pak zobrazuje nejčastější druhy akcí figurujících u kybernetických incidentů (Verizon Communications, © 2023).



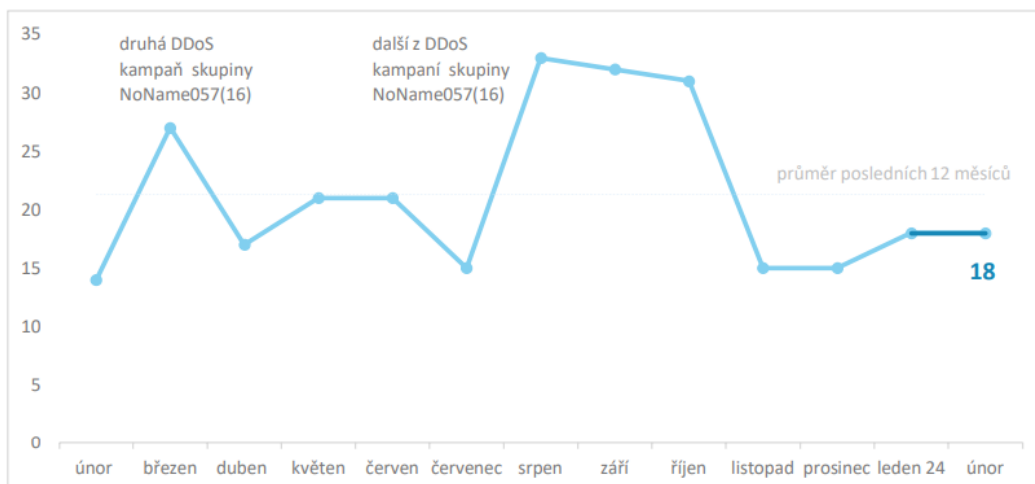
**Figure 16.** Top Action vectors in breaches (n=3,194)

**Figure 17.** Top Action vectors in incidents (n=10,502)

Obrázek 7 Nejčastější iniciační akce u datových úniků a kybernetických incidentů (Verizon Communications, © 2023).

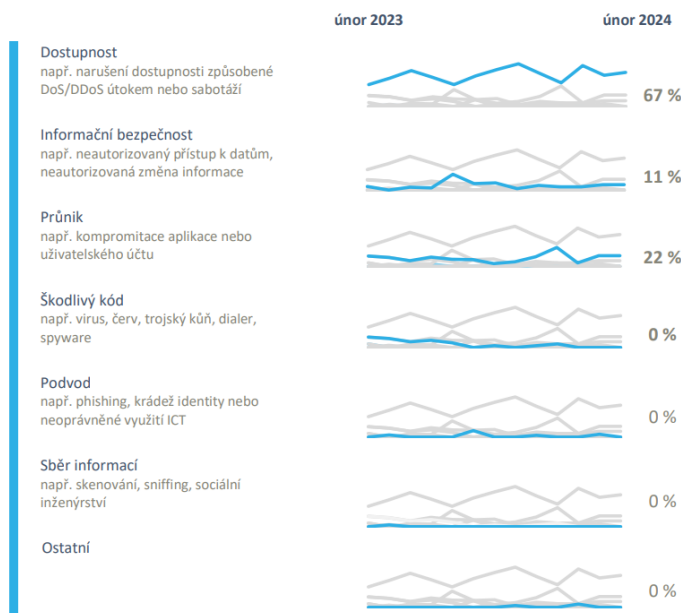
Nejčastěji iniciační prvky kybernetického útoku jsou datové úniky a kybernetické incidenty iniciovány skrze webové stránky, email či neopatrnost uživatele. (Verizon Communications, © 2023).

V rámci České republiky bylo v únoru 2024 evidováno 18 kybernetických útoků. Celkový vývoj v intervalu od února 2023 do února 2024 je zobrazeno na Obrázku 8 (NÚKIB, ©2024).



Obrázek 8 Počet kybernetických útoků dle NÚKIB (NÚKIB, ©2024)

Klasifikace identifikující, o jaká typ útoku se jedná, je zobrazeno na Obrázku 9.



Obrázek 9 Klasifikace kybernetických incidentů (NÚKIB, ©2024)

Jak je patrné z grafů DoS/DDoS útoky dominují nejen v celosvětovém měřítku ale i v rámci České republiky. V rámci této země se vesměs ukazují pouze 3 typy incidentů. Je však nutné podotknout, že incidenty jsou hlášeny subjekty. Mají tuto povinnost ze zákona o kybernetické bezpečnosti. V jiném případě se jedná o dobrovolné rozhodnutí. Lze tak očekávat, že v rámci republiky bude určitá část vzniklých incidentů nenahlášených (NÚKIB, ©2024).

Kybernetické útoky pomocí AI jsou klasifikovány dle fází v kybernetickém bezpečnostním řetězci:

- Fáze průzkumu – představuje snahu aktéra narušující kybernetickou bezpečnost získat informace a vzorce chování napadeného subjektu s cílem odhalit zranitelnosti v systému. Díky umělé inteligenci by pro útočníky bylo jednodušší zpracovat velký objem těchto dat,
- Fáze přístupu a průniku – implementace škodlivého programu,
- Fáze dodání – přesun malwaru k cíli, aniž by byl detekován,
- Fáze zneužití – získání autorizace k systému či aplikaci oběti,
- Fáze kontroly a ovládání – vytvoření komunikačního kanálu mezi cílem a útočníkem za účelem získání vlivu nad napadeným subjektem,
- Fáze akce vůči cíli – dosažení plánovaného cíle hackerem (Guembe et al., 2022).

Pro vytváření kybernetických útoků lze použít i generativní modely. Jeden takový model vytvořen čistě pro nezákonnou činnost je WormGPT, jenž lze označit za zlé dvojče ChatGPT. Díky absenci etických kodexů ho lze využít k tvorbě kybernetických útoků (The Hacker News, © 2023).

Umělou inteligenci díky jejím schopnostem lze využít pro mnoho různých kybernetických útoků. Typy útoků, u nichž je tato technologie využívána anebo je zde největší předpoklad pro její zneužití na základě předvedených dat jsou popsány následovně.

### 3.6.1 Next-generation Malware

Jeden z hlavních cílů malwaru je infikovat počítač bez povšimnutí antimalwarového programu. S využitím hloubkového učení se takovýto program snadno adaptuje na infikované prostředí a je hůře detekovatelných antimalwarovými programy. Příkladem takového programu je DeepLocker. Oběť tento malware omylem stáhne s jinou aplikací obdobně jako trojský kůň. Díky hluboké neuronové síti je program natrénován se chovat normálně a zareagovat pouze na specifické akce či vlastnosti oběti jako je obličej či hlas. Při využívání zařízení jinou osobou program nereaguje. Příkaz k zahájení útoku malware získává z mikrofonu či kamery v zařízení. DeepLocker byl prezentován výzkumníky na konferenci Black Hat USA v roce 2018 a žádná opatření nebyla implementována. Byla však navržena opatření jako omezení přístupů k sensorům a využití kybernetického klamání k přesměrování a deaktivaci DeepLockeru (Kaloudi, Li, 2021).



### 3.6.2 Deepfake

Deepfake jsou označovány fotky, videa a hlasové záznamy upravené umělou inteligencí s cílem oklamat lidi. Tento trend je oblíbený zejména díky vysoké uvěřitelnosti a snadné dostupnosti. V rámci individuální roviny se objevuje problém s poškození reputace daného člověka. Příkladem může být vytváření Deepfake snímků pro dospělé, kde se vymění obličeje protagonistů za obličeje určitého člověka. Reputace dané osoby je poškozena nejen reputace, ale i psychické zdraví jedince či zničen jeho karierní růst. Díky sociálním sítím a fotkám tak lze zdiskreditovat jakéhokoliv člověka. Pro organizace jsou falešné fotky, videa či zvukové záznamy nebezpečné, jelikož se mohou stát obětí podvodu. Může se stát, že manažerovi zavolá falešný šéf a získá od něj údaje. V neposlední řadě Deepfake může být rizikový i pro vládu či státní samosprávu. Díky této technologii lze vytvářet falešná prohlášení s obličejem prezidenta či tvorba dezinformací s cílem ovlivnit veřejné mínění a případně podkopat důvěryhodnost vládních institucí. Pro minimalizaci tohoto rizika byl navrhnout v rámci článku Deepfakes: Trick or Treat? rámec R.E.A.L. skládající se z:

- **Record** – uložení původního obsahu pro zajištění popření zfalšovaného obsahu,
- **Expose** – odhalení DeepFaku včas. Proto jsou vyvíjeny programy s využitím umělé inteligence detekující nesrovnalosti typické pro Deepfake. Těmi jsou například nesrovnalosti v kvalitě rozlišení, škálování, spojování a otáčení obsahu a dále i vzorce mrkání očí na snímcích, kde figurují lidé. Vývoj těchto nástrojů je podporován jak národními institucemi či soukromými společnostmi jako je například Facebook. Příkladem může být program Truepic vyvíjený Agenturou pro pokročilé obranné výzkumné projekty, jenž je financována Pentagonem. Je nutné však tyto technologie rozvíjet, neboť se rozvíjí i sofistikovanost Deepfakes,
- **Advocate for legal protection** – jedním z nedostatků v rámci digitálního obsahu je absence právního rámce. Sociální sítě a jejich provozovatelé nejsou zodpovědní na pravdivost informací kolujících na těchto platformách. V případě Deepfakes není moc právních nástrojů pro zajištění právního postihu pro poškozené oběti. Je tedy nutné se zabírat v rámci práva touto problematikou více,
- **Leverage trust** – silné značky mají lepší postavení v rámci Deepfakes. Zainteresované strany mají ve značku důvěru a jsou hůře ovlivnitelní falešnými informacemi. (Kietzmann et al., 2020).

Problematika Deepfakes se stává čím dál více relevantní hrozbou pro společnosti. Příkladem může být incident, při němž byl výkonný ředitel nejmenované britské společnosti oklamán Deepfake telefonátem s údajným nadřízeným německé mateřské společnosti. Falešný šéf, jehož hlas byl napodoben perfektně i s přízvukem chtěl, aby zaslal 220 tisíc eur maďarskému dodavateli, což provedl. Celkem podvodník zavolał třikrát s tím, že napotřetí byl podvod britským ředitelem odhalen. Jenže bylo pozdě a společnost přišla o finance (Forbes Media LLC, © 2024).

Jeden z novějších kybernetických incidentů se stal v únoru 2024 v Hongkongu. Zaměstnanec finančního oddělení nejmenované společnosti byl pozván na videokonferenci se spolupracovníky. Jenže se ukázalo, že tito pracovníci byly falešní a zaměstnanec byl úspěšně oklamán. Před videohovorem dostal zprávu o převedení financí od finančního ředitele společnosti z Velké Británie. Zde nastaly u něj pochybnosti, které však byly videokonferencí vyvráceny a požadované finance odeslal. Firma tak přišla o 25 milionů dolarů. Dle hongkongské policie se čím dál stávají častější případy, kdy jsou k těmto podvodům využity kradené občanské průkazy. Zatím se objevilo minimálně dvacet obdobných případů oklamání softwaru pro rozpoznání obličeje (Cable News Network, © 2024).

### 3.6.3 Prolomení hesla

Novým přístupem ke generování kvalitních odhadů hesel je tak zvaný PassGAN využívající generativní adversariální síť, jenž je neuronovou sítí pro generování vzorků. V rámci výzkumu vytvořeného výzkumníky Briland Hitaj, Paolo Gasti, Giuseppe Ateniese a Fernando Perez-Cruz byla síť trénována, tak aby na základě trénovaných dat vytvářela cílené vzorky. Pro automatické generování hesel byly využity dvě hluboké neuronové sítě, a to generativní a diskriminační. Generativní neuronová síť je trénována pomocí vektoru šumu představující nahodilé rozdělení pravděpodobností a generuje falešné vzorky hesel. Diskriminační neuronové sítě jsou předkládané skutečné vzorky hesel, jež jsou skutečná uniklá hesla a falešné vzorky a zmíněná síť se pokouší je rozeznat. V případě, že jsou rozdílná je předána zpětná vazba generující neuronové síti. Ta na základě těchto informací upravuje svoje výstupy (Hitaj et al., 2019).

### 3.6.4 Phishing

Phishingové útoky mají za cíl získat od oběti citlivé údaje tak, že oběť oklamou. S rozvojem umělé inteligence se tyto útoky zjednodušují. AI může získat skrze data oběti o ni různé

informace a udělat tak útok personifikovaný a hůře odhalitelný. Tato technologie je schopna nalézt kontakt na spolupracovníka či vedoucího oběti, zavolat mu a skrze syntézou řeči zavolat zfalšovaným hlasem. Ta s myšlenkou, že jde o známého člověka, mu tyto údaje poskytne. Vesměs tato technika útoku je provázána s technologií Deepfakes (Alakeel et al., 2023).

### **3.6.5 DoS/DDoS útok**

Denial of Service (DoS) a Distributed Denial of Service (DDoS) jsou útoky, jejichž cílem je ochromit funkci serveru a online služeb a zamezit tak přístup uživatelům. Hlavním rozdílem DoS a DDoS je počet zdrojů. U DoS útoků se jedná o jeden zdroj, zatímco u DDoS je zdrojů útoků více. Umělá inteligence dokáže tyto útoky nejenže automatizovat, ale i přidat schopnost adaptace. Během útoků se umělá inteligence přizpůsobuje a mění strategií na základě dat a informací a je schopna identifikovat slabé cíle a slabiny v zabezpečení. Pro organizaci to tak může mít nepříznivý dopad a původní obranné strategie již nejsou dostatečné. Pro obranu je tak vhodné využívat umělou inteligenci, která je schopna obdobně jako útočná se přizpůsobovat aktuální situaci (Shroff et al., 2022).

## **3.7 Dílčí závěr rizik spojených s umělou inteligencí**

V této části práce byly rozebrány rizika umělé inteligence. V kapitole nejsou rozepsány veškeré existující hrozby ze strany této technologie. Tyto nežádoucí vlivy byly do práce zahrnuty na základě literárních rešerší s přihlédnutím na definici, co je umělá inteligence. U určitých zařízení často bývá zobrazeno, že využívá technologii AI. Avšak se často může jednat o nižší formu nazvanou reaktivní, která však nemusí naplňovat definici stanovenou v práci. V této části tak byly popsáno šest typů hrozeb ze strany probírané technologie. U těchto hrozeb existuje určitá pravděpodobnost ohrožení ochrany obyvatelstva anebo je potřebné se o ně do budoucna zajímat. Nalezená rizika budou následně hodnocena v praktické části z pohledu relevance pro řešený subjekt.

## **II. PRAKTICKÁ ČÁST**

## 4 IDENTIFIKACE A ANALÝZA RELEVANTNÍCH RIZIK VE VZTAHU KE KONTEXTU PRÁCE

V teoretické části byly popsány na základě odborné literatury a odborných zdrojů vybraná rizika, které technologie umělé inteligence přináší. Všechny ale nemusí být pro kontext práce relevantní. Pro tuto potřebu byl zvolen subjekt typu obecní úřad. Při vzniku nežádoucí události ze strany AI může být ohrožena schopnost tohoto subjektu plnit úkoly ochrany obyvatelstva.

### 4.1 Vymezení relevantních rizik

Rizika lze kategorizovat do skupin: autonomní zbraně, existenční riziko, růst nezaměstnanosti, narušení funkčnosti AI, chybně naučený model umělé inteligence a kybernetické incidenty. V této části jsou jednotlivé rizika rozebrána z pohledu relevantnosti rizika vůči aktivitám obecního úřadu.

Pro zjednodušení je zde Tabulka 1, kde **x** znázorňuje irelevantnost a **✓** relevantnost rizik.

Tabulka 1 Relevance rizik (Vlastní zpracování)

Rizika	Relevantnost
<b>Autonomní zbraně</b>	<b>x</b>
Existenční riziko	<b>x</b>
<b>Růst nezaměstnanosti</b>	<b>x</b>
Narušení funkčnosti AI	<b>x</b>
Chybně naučený model umělé inteligence	<b>x</b>
Kybernetické útoky	<b>✓</b>

První zmíněnou problematikou jsou autonomní zbraně. I přes skutečnost, že se jedná o závažnou hrozbu, jenž by mohla způsobit značné ztráty na lidských životech je toto riziko pro obec jako takovou nerelevantní. K tomuto závěru vede několik skutečností. První z nich je náročná specifikace, co to vlastně autonomní zbraň je. Jak bylo nastíněno v teoretické části, tak samotný pojem autonomní zbraň není přesně definován. Má tedy obdobné potíže jako pojem umělá inteligence. Dále je zde i skutečnost, že zbraně s umělou inteligencí nejsou zcela vyvinuty a nejsou v aktivní výzbroji žádné armády a je otázkou, zda tomu v dohledné době bude jinak. Technologie se sice neustále vyvíjí, ale AI na takové úrovni je

stále teoretická. V poslední řadě je velmi nepravděpodobné použití v dohledné době na území Česka. Díky geopolitické situaci je nízká pravděpodobnost konfliktu na území České republiky v blízké době.

Existenční riziko je hrozbou, jenž se může zdát lehce abstraktní a jako by vypadlo ze science fiction knížek. Jenže dle odborníků se může jednat o zásadní problém a pravděpodobnost není v důsledku malá. Jenže tato hrozba není ve vztahu k řešenému subjektu v rámci práce velkým rizikem. Hlavním důvodem k tomuto závěru je fakt, že umělá inteligence schopná něco takového provést ještě není vyvinutá a dle odhadů by měl vývoj takové inteligence trvat ještě několik desítek let.

Dalším problematickým bodem, který může vývoj umělé inteligence přinést je růst nezaměstnanosti. To by se však subjektům obecního úřadu dlouhou dobu nemělo týkat. Je to dáno skutečností, že ať už práce starostky a místostarostky či dalších klíčových zaměstnanců je náročná na automatizaci. Práce těchto pracovníků je často ojedinělá a náročná na rozhodování. Lze předpokládat tedy, že by bylo tuto práci náročné automatizovat. Navíc tato technologie je finančně nákladná a obce zpravidla nedisponují dostatečnými zdroji. Implementace by v tomto případě nepřinesla dostatečnou návratnost. Navíc dle dat popsanych v teoretické části je pravděpodobné, že naopak umělá inteligence více pracovních pozic přinese, než eliminuje.

Narušení funkčnosti AI je naopak jedním z aktuálnějších problémů. Jenže toto riziko pro zmíněnou instituci není opět relevantní skupinou, jelikož úřady moc nevyužívají žádnou technologii, jež by splňovala definici pojmu Umělá inteligence stanoveného v práci. Proto je nepravděpodobné, aby k takovému narušení došlo.

Podobně je to i u hrozby chybně naučeného modelu umělé inteligence. Implementace této technologie začíná být trendem. Avšak u obcí se takové trendy objevují pomaleji nežli vůbec. Výjimkou můžou být chatboty, které by mohli způsobit jisté problémy. Neměly by to však být závažné a je u nich počítáno s určitou rezervou v přesnosti dat. Jejich použití není vhodné pro složitější aktivity obce, u kterých by vznikly závažné problémy.

Posledním rizikem jsou kybernetické útoky, jež by mohli být umělou inteligencí zlepšovány a zjednodušovány. Toto riziko je ze všech v rámci probírané tematiky nejrelevantnější. Je to dáno zejména tím, že například Deepfake útoky jsou již prováděny. Navíc implementace umělé inteligence do kybernetických útoků není vůbec složitá. Je tak možné předpokládat, že aktiva obecního úřadu budou vystaveny právě této hrozbě. Z tohoto důvodu se bude práce

ubírat tímto směrem. Dle dat z teoretické části jsou nejčastějšími hrozbami DDoS útok, ransomware, phishing a použití ukradených údajů. Ukradení osobních údajů může být provedeno pomocí personifikovaného phishingu či technologie Deepfakem anebo prolomením hesla pomocí AI. Z pohledu podstaty práce budou řešeny tyto rizika ve spojení s umělou inteligencí, jenž může u těchto hrozeb zvýšit četnost či dopad. V rámci obecního úřadu jsou aktiva ohrožitelná umělou inteligencí převážně v digitální podobě. Jedná se tedy převážně o operační systém, dokumenty, hesla, webové stránky obce a datová schránka obce.

## 4.2 Analýza rizik

V této části práce bude provedena analýza rizik, jež byly vyfiltrovány v předchozí kapitole ve vztahu ke stanovenému subjektu. Pro analýzu byla vybrána metoda FMEA, jenž bude provedena ve více variantách. Umělá inteligence nemusí být jen hrozbou, ale dokáže být i dobrým pomocníkem. Je tedy otázkou, zda ji lze využít pro analýzu rizik. Pro zodpovězení této problematiky jsou v rámci práce využity i různé dostupné chatbot aplikace generující výstupy na základě zadaných vstupů. Tyto výstupy budou komparovány mezi sebou včetně vlastní analýzy.

### 4.2.1 Vlastní FMEA

Analytická metoda FMEA je metodou, jenž je složena z tabulky identifikující možnou chybu s hodnocením závažnosti rizika RPN(Risk Priority Number). Závažnost rizika je výsledná hodnota označující jak závažné či přijatelné riziko je. Dále se skládá z významu, výskytu a pravděpodobnost odhalení rizika. Jak byly tyto kritéria vymezeny, zobrazují následující tabulky. V Tabulce 1 je definován význam rizika pro daný subjekt. Jedná se o hodnoty znázorňující, jak velký význam může mít chyba dopad na organizaci. V případě této práce byl brán ohled na schopnost subjekt typu obecního úřadu. Bylo zejména bráno ohled na to, jak případný incident zasáhne danou organizaci.

Tabulka 2 Význam následků kybernetického incidentu (Vysoké učení technické v Brně, © 2024)

Význam následků kybernetického incidentu		Klasifikace
<b>Velmi vysoký význam</b>	Následek kybernetického útoku je kompletní ochromení instituce s rizikem přenosu na další subjekty. Obec není schopna zajistit služby.	<b>10</b>
<b>Vysoký význam</b>	Následek kybernetického útoku je rozsáhlý, nutno řešit skrze profesionální IT instituci či firmu. Zároveň dochází k omezení činnosti anebo k výraznému zhoršení kvality poskytovaných služeb.	<b>7 až 9</b>
<b>Střední význam</b>	Následek kybernetického útoku je nutné řešit s technicky zdatnou osobou v oblasti IT. Krátkodobé přerušení činnosti úřadu či mírné zhoršení kvality poskytovaných služeb.	<b>4 až 6</b>
<b>Mírný význam</b>	Následek kybernetického incidentu je minimální, lze jej snadno vyřešit svépomocí. Instituce nemusí omezit svou činnost.	<b>2 až 3</b>
<b>Bezvýznamný</b>	Následek kybernetického incidentu je téměř nulový a není potřeba jej řešit	<b>1</b>

Následujícím kritériem je pravděpodobnost výskytu dané chyby znázorněné v Tabulce 2 a definující s jakou pravděpodobností může tato chyba nastat.

Tabulka 3 Pravděpodobnost výskytu kybernetického incidentu (Vysoké učení technické v Brně, © 2024)

PRAVDĚPODOBNOST výskytu kybernetického incidentu		Klasifikace
<b>Jistotá</b>	Je jisté, že ke kybernetickému incidentu dojde	<b>10</b>
<b>Vysoká</b>	Ke kybernetickému incidentu s vysokou pravděpodobností dojde	<b>7 až 9</b>
<b>Průměrná</b>	Střední pravděpodobnost vzniku kybernetického incidentu	<b>4 až 6</b>
<b>Malá</b>	Mírná pravděpodobnost vzniku kybernetického incidentu	<b>2 až 3</b>
<b>Velmi nízká</b>	Ke kybernetickému incidentu téměř jistě nedojde	<b>1</b>

Posledním kritériem pro výpočet RPN je pravděpodobnost odhalení rizika vycházející z náročnosti odhalení potenciálního kybernetického útoku ještě před jeho započítáním. To je zobrazeno v Tabulka 3.



Tabulka 4 Pravděpodobnost odhalení (Vysoké učení technické v Brně, © 2024)

Pravděpodobnost odhalení příznaků kybernetického útoku indikující jeho možný vznik		Klasifikace
Nejistitelné	Nelze odhalit příznaky	10
Nízká	Příznaky pokusu kybernetický útok odhalí pouze expert na kyberbezpečnost anebo je nutné využití detekce antimalware programu na vysoké úrovni	7 až 9
Střední	Pokus o kybernetický útok odhalí osoba s dobrými znalostmi o problematice. Bezplatný antimalware je dostačující.	4 až 6
Vysoké	Pokus o kybernetický útok lze na jisto odhalit i menšími znalostmi o problematice	2 až 3
Téměř jisté	Pokus o kybernetický útok lze na jisto odhalit i bez jakýchkoliv IT či kybernetických znalostí	1

Výsledná hodnota RPN je vypočtena rovnicí  $RPN = Z (\text{závažnost}) * V(\text{výskyt}) * O(\text{odhalitelnost})$ . Výsledkem je pak zmíněná závažnost rizika, které je však nutné klasifikovat do úrovní dle významnosti, díky čemuž lze stanovit, zda je riziko přijatelné či ne a zda je nutné toto riziko snížit na přijatelnou úroveň. Následná Tabulka 4 klasifikaci znázorňuje.

Tabulka 5 Klasifikace rizika RPN (Vysoké učení technické v Brně, © 2024)

Klasifikace rizika RPN	
Nízké riziko	0–165
Střední riziko	166–340
Vysoké riziko	341-1000

Pro nízké riziko byl stanoven interval 0 až 165. V tomto intervalu jsou rizika brány jako přijatelné a není potřeba zvláštních opatření. Interval byl stanoven s ohledem na podstatu problematiky. Pro střední riziko byl nadefinován interval 166 až 340. Jedná se o střeni rizika nepředstavující vážnou hrozbu, ale je vhodné provést určitá opatření. Strop byl zvolen tak aby nedocházelo k situaci, kdy jsou všechna hodnotící kritéria v červených číslech ale výsledné RPN je bráno jako střední. Vysoké rizika jsou pak 341 a výš a jsou pro ně navržena určitá opatření. Je s nimi pak pracováno i dále.

Po stanovení hodnotících kritérií a klasifikace rizika, dle něž lze určit přijatelnost následuje samotná analýza FMEA zobrazená na Obrázku 10.

Objekt:		Obecní úřad				Rok 2024								
Odpovědnost za vypracování		Bc. František Duďa				FMEA analýza současného stavu								
Prvek	Možná chyba	Možný důsledek	Příčina	preventivní opatření	Význam	Vznik	Odháten	RPN	Doporučená opatření	Odpovědnost	Hodnocení stavu po realizaci			
											Význam	Vznik	Odháten	Možné riziko
Webové stránky obce	Zamezení dostupnosti DoS/DDoS útokem	Zamezení přístupu občanů k důležitým informacím	Špatné zabezpečení, chybějící filtrace požadavků	Provozují externí společnost -> vlastní zabezpečení	3	9	8	216		Externí společnost				
	Infikování příloh Ransomwarem na stránkách	Infikace počítače uživatele přicházejícího na stránku	Průnik škodlivé AI skrze zabezpečení		3	4	8	96		Externí společnost				
WiFi síť obce	Nedostupnost způsobená DoS/DDoS útokem	Obec nemůže využívat vlastní WiFi síť	Špatné zabezpečení, chybějící filtrace požadavků	Žádné	3	9	8	216	Předplatné AntiDDoS ochrany	Externí IT technik	3	3	6	54
Veřejná síť WiFi		Občané nemohou využívat veřejnou WiFi síť			1	9	8	72			1	3	6	18
Administrativní elektronické dokumenty	Infikování počítače Ransomwarem	Ztráta přístupu k dokumentům	Neúmyslné stažení z falešných stránek anebo z emailové schránky	Windows Defender	7	7	7	343	Školení, silnější anti-malwarová ochrana, externí záloha	Starostka, místostarostka, účetní a veškerý personál obce	7	3	5	105
Operační systém		Znemožnění funkce počítače			9	7	7	441			9	3	5	135

Obrázek 10 Vlastní FMEA analýza 1. část (Vysoké učení technické v Brně, © 2024)  
Pro svou velikost bylo nutné ji rozdělit na dvě části, a tudíž je druhá část na Obrázku 11.

Bankovní účet obce	Neúmyslné poskytnutí dat útočníkovi	Ztráta přístupu či finančních prostředků kvůli Phishingu	Špatné vyhodnocení pravosti webu či emailu	Windows Defender	10	7	6	420	Školení, kvalitní anti-malware ochrana	Starostka, místostarostka, účetní a veškerý personál obce	10	3	3	90
	Provedení akce požadované útočníkem využívající Deepfake technologie	Ztráta přístupu či finančních prostředků	Kvalitně vytvořený Deepfake útočníkem a nerozpoznání falešného požadavku obětí	Žádné	8	8	8	512	Školení, organizační opatření, využití programů odhalující Deepfake	Starostka, místostarostka, účetní a veškerý personál obce	8	3	4	96
	Prolomení hesla AI		Slabé či špatně vytvořené heslo	Stávající heslo	10	7	7	490	Dvoufázové ověřování, implementace biometrických prvků, pravidelná obměna hesel	Starostka obce	10	3	4	120
Datová schránka obce Email obecního úřadu Facebook účet	Odhalení přístupových údajů	Ztráta přístupu, odtajnění citlivých informací, možnost rozeslat phishing či dezinformace skrze účet obce	Úspěšné podvedení obětí Phishingem	Žádné	8	7	7	392	Školení o problematice	Starostka, místostarostka, účetní	8	2	3	48
			Úspěšné podvedení obětí Deepfake		8	8	8	512			Starostka, místostarostka, účetní	8	3	4
Občané	Manipulace dezinformací	Ztráta důvěry v orgán, zásadní ovlivňování voleb, vyvolání paniky	Kvalitní Deepfake, který obelhal občany	Žádné	9	8	8	576	Osvěta občanů o problematice, mechanismus pro včasné odhalení či vyvrácení	Starostka obce	9	4	4	144

Obrázek 11 Vlastní FMEA analýza 2. část (Vysoké učení technické v Brně, © 2024)  
Prvním zkoumaným prvkem v analýze jsou webové stránky obce, u nichž jsou rizika DoS/DDoS útok a škodlivý program Ransomware. Právě díky umělé inteligenci mohou být DoS/DDoS útoky frekventovanější. Důvodem pro vysoké hodnocení je tak i skutečnost, že již nyní jsou tyto hrozby nejčastější. Hlavním důvodem, proč ale nejsou největším rizikem pro samotnou obec, je jejich nízký dopad či význam. V případě nedostupnosti webu nedojde k velkému ovlivnění chodu obce. Navíc DoS/DDoS jsou spíše nepříjemností a v případě nouze lze poskytování informací pro občany nahradit skrze Facebook či místním rozhlasem.

Dalším rizikem, co se týče webové stránky obce je napadení stáhnutelného obsahu na stránkách škodlivým malwarem. Zde je pracováno s ransomwarem, jelikož se jedná o jeden z častějších aktérů kybernetických incidentů. Umělá inteligence by mu mohla poskytnout schopnost adaptace a zvýšit jeho nebezpečnost. V rámci problematiky by mohl být přenesen do stáhnutelného obsahu například v dokumentu termínu svozu odpadů a při stažení obětí by byl aktivován. Toto riziko je však nepravděpodobné, jelikož stránky jsou dostatečně zabezpečeny externí firmou a takový typ útoku je poměrně komplikovaný a existují jednodušší způsoby. Význam je menší, jelikož dokumenty na stránkách obce nejsou využívány frekventovaně.

Dalším prvkem, jenž může být napaden kybernetickým útokem, u něhož byla zneužita umělá inteligence je veřejná WiFi a WiFi síť. DoS/DDoS útok by mohl tuto síť zahltit požadavky a vyřadit je z provozu. Opět jako u webových stránek by se však jednalo pouze o nepříjemnost a případná náprava by neměla trvat dlouho. Obě však jsou z významového hlediska ohodnocené jinak. Význam veřejné WiFi není pro obec vysoký z důvodu faktu, že veřejná WiFi je využívána minimálně a je jen vytvořena čistě z legislativních požadavků týkající se veřejné knihovny. Obec je sice povinna, aby veřejný internet fungoval, avšak při incidentu znemožňující správnou funkci by tato událost nikoho nezasáhla. Naopak síť využitou k provozu obecní samosprávy může být horší. Nejedná se však o vážnou komplikaci a jak bylo řečeno tento typ útoku je v drtivé většině případů spíše jen obtěžující. Práce obce vyžadující online připojení by však nemohla být prováděna. Lze však provádět prozatím činnosti nevyžadující internet či případně by někdo ze zaměstnanců poskytl Hotspot skrze datové připojení.

Administrativní dokumenty v elektronické podobě jsou podstatným prvkem subjektu typu obecní úřad. Jedná se o různé dokumenty nutné pro chod organizace, jakými jsou například smlouvy o pronájmu, pachtovní smlouvy, účetní uzávěrky, plán krizové připravenosti obce, pracovní smlouvy a osobní údaje občanů. V případě zašifrování těchto dat Ransomware by vznikl zásadní problém i přes skutečnost, že část dokumentů je ve fyzické podobě. Problematické by pak byla jejich digitalizace, která by byla časově náročná. Jiné datové soubory pouze v elektronické podobě by byly ztraceny úplně. Jsou jimi převážně projekty a dokumenty, s nimiž by bylo aktuálně pracováno. Hodnocení významu je proto vyšší. Obdobnou hrozbou je vystaven i operační systém na počítačových zařízeních obce. Zde by však navíc došlo ke kompletnímu zamezení funkčnosti počítače. V rámci tohoto prvku by byly škody podobné. Ransomware je stále poměrně častou hrozbou dle dat popsaných

v teoretické části a jeho pravděpodobný vznik je vysoký. Ten by mohl být ještě umocněn implementací technologie AI. Navíc díky implementaci AI by se sta více sofistikovaným malwarem a lépe se adaptoval na prostředí v obecních počítačích. Nutné je podotknout, že s největší pravděpodobností tato hrozba bude spojená s hrozbou sociálního inženýrství s cílem oklamat a donutit oběť stáhnout infikovaný soubor.

Bankovní účet obce je nesmírně důležitým aktivem obce, s jehož pomocí je placen provoz, investice a závazky obce. Závažnou hrozbou je tak phishing, Deepfake útok a případně prolomení hesla za pomoci nástrojů umělé inteligence. V případě phishingu je pravděpodobný scénář, kdy bude oběť oklamána falešnou emailovou zprávou. Umělá inteligence navíc přidá více autentičnosti a díky své schopnosti učení a zpracování velkého objemu dat budou phishingové útoky šité na míru oběti. Bude tak náročné odhalit falsifikovaný email běžným pracovníkem. U tohoto útoku může být použita i varianta webové stránky, kdy oběť neúmyslně zadá do falešného webu přihlašovací údaje. U těchto útoků může být použit i Deepfake. Tento fenomén se v poslední době dostává i do rukou hackerů. AI je schopna pomocí syntézy hlasu upravit hlas, aby zněl jako jiná osoba, případně je možné i vytvořit falešný videohovor s vysokou uvěřitelností. S vývojem AI tak může být náročnější pro pracovníka obce rozeznat, zda se jedná o jeho nadřízeného či nikoliv. V rámci analýzy byl význam u phishingu hodnocen výše než u Deepfake, jelikož u phishingu je pravděpodobná varianta ztráty přihlašovacích údajů při zadání do falešných stránek a tím pádem vybrání celého konta útočníkem. Naopak u Deepfake oběť zpravidla nepřijde o veškeré finance na účtu, jelikož útočník skrze tuto technologii nabádá k provedení určité akce. Pracovníkovi by tak mohlo přijít zvláštní požadavek o přeposlání veškerých financí a mohl by pokus o útok odhalit. S tím útočník počítá a nebude se proto snažit s největší pravděpodobností o veškeré finance obce v případě použití této technologie. Nicméně však díky zlepšování a rozvoji AI lze počítat s nárůstem těchto kybernetických rizik a jejich náročnější identifikaci. Phishing je hodnocen jako pravděpodobnější, jelikož Deepfake lze brát jako jedna z variant phishingového útoku. Úroveň schopnosti odhalit chybu je u obou rizik též jiná. Phishing může se zlepšit díky lepší personifikaci. Jenže díky schopnosti Deepfaku provádět i falešné videohovory, je jeho odhalení daleko náročnější. Navíc u phishingu je určitá osvěta o problematice zažitější.

Další hrozbou pro bankovní účet je riziko prolomení hesla například pomocí PassGAN. AI je schopna díky svým vlastnostem předvídat a odhalovat hesla či přihlašovací údaje, což bude nadále v průběhu let neustále rozvíjeno. Proto by bylo vhodné toto riziko brát v potaz.

Dalším aktivem ohrožitelným škodlivou AI je datová schránka obce skrz, níž obec komunikuje s úřady a tím pádem by byl dopad zmíněné chyby zásadní. Riziko je obdobné jako v předchozím případě s rozdílem, že by se nejednalo o kompletní ztrátu financí ale ztrátu informací a přístupu ke zprávám od státních institucí. Tento problém by se musel řešit s příslušným úřadem. Stejným nebezpečím je vystaven i email obecního úřadu. Pomocí oklamání oběti může útočník získat přístup a použít email k šíření malware anebo tak získat citlivé informace. Případně při získání přístupu k emailu či Facebooku obce, může rozesílat Spam zprávy či dezinformace.

Závěrečným aktivem jsou samotní občané, kteří mohou být vystaveni Deepfake manipulaci. Díky této technologii mohou být občané ovlivněni u voleb anebo vyvolat paniku či dehonestovat představitele obecního orgánu. Tyto rizika jsou do jisté míry aktuální a podobné případy již vznikly. Proto získal tak vysoké hodnocení.

#### **4.2.2 FMEA analýzy pomocí umělé inteligence**

Jak bylo avizováno umělá inteligence má spoustu způsobů využití. Populární jsou zejména různí chatboti generující odpověď na základě vstupní otázky a pro učení jsou jim poskytovány data z internetu. Tito chatboti jsou schopni psát eseje, kuchařské recepty a generovat i umělecké obrázky. Lze tak vyvodit i myšlenku jejich využití pro analýzu rizik. V této části jsou znázorněny výstupy jednotlivých chatbotů týkající se problematiky rizik umělé inteligence u obecního úřadu.

##### **4.2.2.1 Gemini**

Chatbot Gemini je výtvozem společnosti Google a je volně dostupný pro každého zdarma. Prozatím neexistuje žádná placená verze a může jej vyzkoušet každý. Této aplikaci byla tedy zadán požadavek: “Proveď FMEA analýzu obecního úřadu s ohledem na rizika, u nichž je využita AI: DoS/DDos útok, Ransomware, Prolomení hesla (PassGAN), Deepfake a Phishing. Děkuji. “ Výstup z aplikace je zobrazena na Obrázku 12.

Riziko	Popis	Potenciální Důsledky	Pravděpodobnost	Dopad	Opatření	Odpovědnost
DoS/DDoS útok	Přetížení serverů úřadu a znemožnění přístupu k důležitým službám.	Ztráta produktivity, ztráta dat, narušení důvěry občanů.	Vysoká	Vysoký	Implementace robustních bezpečnostních opatření (např. WAF), zálohování dat.	IT oddělení
Ransomware	Zašifrování dat úřadu a požadování výkupného za jejich dešifrování.	Ztráta dat, finanční ztráta, narušení důvěry občanů.	Střední	Vysoký	Pravidelné zálohování dat, implementace antivirového softwaru, školení uživatelů o kybernetické bezpečnosti.	IT oddělení
Prolomení hesla (PassGAN)	Krádež uživatelských hesel a neoprávněný přístup k systémům úřadu.	Ztráta dat, finanční ztráta, narušení důvěry občanů.	Střední	Vysoký	Implementace silných hesel a multifaktorové autentizace.	IT oddělení
Deepfake	Vytvoření falešných videí nebo audiozáznamů představitelů úřadu za účelem šíření dezinformací.	Poškození reputace úřadu, ztráta důvěry občanů.	Nizká	Vysoký	Monitorování online aktivit, vzdělávání občanů o deepfakes.	Tisková mluvčí, IT oddělení
Phishing	Podvodné e-maily a webové stránky s cílem ukrást osobní údaje nebo infikovat počítače malwarem.	Ztráta dat, finanční ztráta, narušení důvěry občanů.	Vysoká	Střední	Školení uživatelů o phishingu, implementace technologií proti phishingu.	IT oddělení

Obrázek 12 FMEA analýza od Gemini (Vygenerováno umělou inteligencí)

Výstup z Gemini byl pouze převed do formátu tabulky Excel a esteticky upraven. Jenže jak se na obrázku vidět moc to analytickou metodu FMEA nepřipomíná. Proto bylo nutné v rámci dalších příkazů požadavek poupravit. Celkem byly položeny příkazy:

1. Proved' FMEA analýzu obecního úřadu s ohledem na rizika, u nichž je využita AI: DoS/DDos útok, Ransomware, Prolomení hesla (PassGAN), Deepfake a Phishing.  
Děkuji
2. Vyjádři v tabulce pravděpodobnost, dopad a odhalitelnost číselně a přidej klasifikační tabulku.
3. Uprav hodnocení odhalitelnosti tak aby byla škála od 1 do 5, kde 1 je nejvyšší a 5 nejnižší.
4. Přidej k hodnocení i důvod k udělení daného hodnocení.
5. Pozměň škály hodnocení z 1 až 5 na 1 až 10.
6. Dopln' do tabulky i hodnocení pravděpodobnosti, dopadu, odhalitelnosti a jejich výsledného součinu po provedené opatření.
7. Přidej do tabulky ještě Prvek, Možná Chyba, Dopad chyby a Možná příčina vzniku chyby.

Finální podoba tabulky je tak zobrazena na Obrázku 13.

Typ rizika	Prvek	Pravděpo- dobnost (1-10)	Závažno- st (1-10)	Odhalteln- ost (1-10)	RPN	Možná Chyba	Dopad chyby	Možná příčina vzniku chyby
DoS/DDos útok	Web	7	10	6	420	Dostupnost webu	Ztráta dostupnosti webu pro občany a zaměstnance	Nedostatečná kapacita serverů, chyby v konfiguraci firewallu
Ransomware	Systém pro ukládání dat	5	10	5	250	Ztráta dat	Šifrování dat, ztráta důležitých dokumentů a informací	Nedostatečné zálohování dat, chyby v konfiguraci softwaru
Prolomení hesla	Systém pro autentizaci	7	10	4	280	Neoprávněný přístup k datům	Zneužití ukradených hesel, přístup k citlivým informacím	Slabá hesla, nedostatečné zabezpečení systému
Deepfake	Web	3	10	8	240	Šíření dezinformací	Poškození reputace úřadu, ztráta důvěry občanů	Nedostatečná edukace o deepfakech, chyby v ověřování informací
Phishing	E-mail	7	8	5	280	Ztráta dat, finanční ztráty	Zneužití ukradených informací, instalace malwaru	Nedostatečná edukace o phishingu, chyby v konfiguraci e-mailového serveru

Obrázek 13 Finální FMEA od Gemini 1. část (Vygenerováno umělou inteligencí)

Pro lepší přehlednost bylo nutné tabulku rozdělit na dvě části, z níž druhé je na Obrázku 14.

Pravděpodob- nost po opatření	Závažnost po opatření	Odhaltelnost po opatření	RPN po opatření	Doporučené opatření
5	8	7	280	Zavedení firewallu a systému pro detekci intruzí (IDS), pravidelné testování odolnosti proti DDoS útokům
4	8	6	192	Zálohování dat na pravidelné bázi, implementace kybernetické bezpečnosti, školení zaměstnanců
3	7	5	105	Používání silných hesel a dvoufázového ověřování, pravidelná změna hesel
2	7	7	98	Vzdělávání zaměstnanců o deepfakech, zavedení politik pro ověřování informací
4	6	6	144	Vzdělávání zaměstnanců o phishingu, zavedení politik pro filtrování spamu a phishingových e-mailů

Obrázek 14 Finální FMEA od Gemini 2. část (Vygenerováno umělou inteligencí)

Jak je patrné Gemini nevygeneroval tabulku zcela a nejspíše narazil na své limity. Následuje ještě klasifikace RPN znázorňující míru rizika. Ta je prezentována v Tabulce 6.

Tabulka 6 Klasifikace  
RPN od Gemini  
(Vygenerováno umělou  
inteligencí)

<b>RPN</b>	<b>Klasifikace</b>
0-100	Nízké
101-250	Střední
251-400	Vysoké
401-500	Velmi vysoké
501+	Kritické


Výhodou této webové aplikace je schopnost do určité míry vytvořit text i v tabulkové formě. Je nutné výstup brát s rezervou a případně doplnit na základě přesnějších informací. Pro rozvinutí nápadů či jako podpora při vlastní analýze je Gemini ucházející.

#### 4.2.2.2 ChatGPT

Další webovou aplikací pro generování odpovědi je ChatGPT od společnosti OpenAI. Jedná se opět o bezplatnou aplikaci, avšak nabízí i vylepšenou verzi za měsíční předplatné. Bezplatná verze využívá jazykový model GPT 3.5 zatímco placená využívá novější GPT 4.

Jako u předešlého chatbotu i zde byla položen příkaz: “Proved’ FMEA analýzu obecního úřadu s ohledem na rizika, u nichž je využita AI: DoS/DDos útok, Ransomware, Prolomení hesla (PassGAN), Deepfake a Phishing. Děkuji“ Pro lepší kvalitu bylo nutné výsledek rozdělit do více obrázků počínající Obrázku 15.



 ChatGPT

Rozumím, že potřebujete provedení FMEA analýzy obecního úřadu Bohuslavice u Zlína vzhledem k rizikům spojeným s využitím umělé inteligence (AI) v kontextu škodlivých aktivit, jako jsou DoS/DDoS útoky, ransomware, prolomení hesla, Deepfake a phishing. Analýza FMEA (Failure Mode and Effects Analysis) je strukturovaný proces, který identifikuje a analyzuje potenciální rizika v systému a navrhuje opatření pro jejich eliminaci nebo omezení. Pro úplnost upozorňuji, že jako umělá inteligence jsem omezena vědomostmi, které jsem byla vystavena během mého vývoje a můžu pomoci se specifikacemi, ale nemohu provádět fyzické činnosti. Mohu vám pomoci sestavit analýzu, ale přímo ji provést neumím.

**1. Identifikace rizik:****1.1. DoS/DDoS útok:**

- *Co se může pokazit?* - Síťová dostupnost, přístup k datům, služby úřadu.
- *Jak to může nastat?* - Přetížení sítě pomocí škodlivého provozu.
- *Jaké budou důsledky?* - Nepřístupnost síťových služeb.

**1.2. Ransomware:**

- *Co se může pokazit?* - Data obecního úřadu.
- *Jak to může nastat?* - Infikování škodlivým kódem.
- *Jaké budou důsledky?* - Uzamčení dat, možná ztráta informací.

**1.3. Prolomení hesla (PassGAN):**

- *Co se může pokazit?* - Přístup k citlivým datům.
- *Jak to může nastat?* - Prolomením slabého hesla.
- *Jaké budou důsledky?* - Nesankcionovaný přístup, data mohou být odcizeny.

Obrázek 15 FMEA analýza od ChatGPT 3.5 free 1. část  
(Vygenerováno umělou inteligencí)

Následuje další část znázorněná na Obrázku 16.

**1.4. Deepfake:**

- *Co se může pokazit?* - Obecní úřad může být vydán falešnými zprávami nebo videi.
- *Jak to může nastat?* - Vytvořením falešného videa pomocí AI.
- *Jaké budou důsledky?* - Ztráta důvěryhodnosti, možné narušení reputace úřadu.

**1.5. Phishing:**

- *Co se může pokazit?* - Získání citlivých informací.
- *Jak to může nastat?* - Zasláním falešného e-mailu nebo webové stránky.
- *Jaké budou důsledky?* - Ztráta dat, možné finanční ztráty.

**2. Výpočet FMEA:****• Závažnost (S)**

- DoS/DDoS útok: 10
- Ransomware: 10
- Prolomení hesla: 9
- Deepfake: 8
- Phishing: 8

**• Četnost (O)**

- DoS/DDoS útok: 5
- Ransomware: 3
- Prolomení hesla: 3
- Deepfake: 2
- Phishing: 2


Obrázek 16 FMEA analýza od ChatGPT 3.5 free 2. část  
(Vygenerováno umělou inteligencí)

A na Obrázku 17 je poslední část vygenerované analýzy od ChatGPT 3.5

- **Detekce (D)**
  - DoS/DDoS útok: 4
  - Ransomware: 4
  - Prolomení hesla: 4
  - Deepfake: 5
  - Phishing: 5
- 3. **Priority (RPN = S \* O \* D)**
  - DoS/DDoS útok:  $10 * 5 * 4 = 200$
  - Ransomware:  $10 * 3 * 4 = 120$
  - Prolomení hesla:  $9 * 3 * 4 = 108$
  - Deepfake:  $8 * 2 * 5 = 80$
  - Phishing:  $8 * 2 * 5 = 80$
- 4. **Priorita akce:**
  - DoS/DDoS útok: Vytvoření a implementace strategie obrany před DoS/DDoS útokem.

### Obrázek 17 FMEA analýza od ChatGPT 3.5 free 3. část (Vygenerováno umělou inteligencí)

Vytvořená analýza nebyla však úplná a bylo nutné doplnit opatření a jejich dopad. Proto byl stanoven další příkaz: “Navrhni opatření k těmto rizikům a proved’ hodnocení RPN po jejich provedení“. Odpověď je opět zobrazena na více obrázcích počínaje Obrázkem 18.

 **ChatGPT**  
Riziko: DoS/DDoS útok

Opatření:

1. Nasazení specializovaného zařízení pro ochranu sítě proti DoS/DDoS útokům.
2. Aktualizace firewallu a směrovačů sítě na poslední verze firmwaru a softwaru pro zajištění maximální stability a bezpečnosti.
3. Nasazení softwarového řešení pro detekci DoS/DDoS útoků a jejich rychlou blokadu.

Hodnocení RPN:

Závažnost (S) = 4 (Snížení neživotně důležitých služeb pro občany a obecní úřad)  
Četnost (O) = 2 (Nízká pravděpodobnost vzniku útoku)  
Detekce (D) = 5 (Předpokládáme, že výše uvedená opatření budou schopna útok detekovat)

RPN ( $4 * 2 * 5$ ) = 40 (po provedení opatření)

Riziko: Ransomware

Opatření:

1. Zavedení politiky zálohování a zabezpečení dat.
2. Školení zaměstnanců obecního úřadu v prevenci ransomware.
3. Nasazení softwarového řešení pro detekci a blokadu ransomware.

Hodnocení RPN:

### Obrázek 18 Opatření vůči rizikům dle ChatGPT 1. část (Vygenerováno umělou inteligencí)

Následuje zbytek navržených opatření umělou inteligencí zobrazený na Obrázkem 19.

Závažnost (S) = 4 (Ztráta citlivých dat)  
Četnost (O) = 2 (Nízká pravděpodobnost vzniku útoku)  
Detekce (D) = 5 (Opatření jsou schopna útok detekovat)

RPN ( $4 * 2 * 5$ ) = 40 (po provedení opatření)

Riziko: Prolomení hesla (PassGAN)

Opatření:

1. Zavedení politiky silných hesel a pravidelné změny.
2. Implementace dvoufaktorové autentizace pro kritické účty.
3. Monitorování chování účtů a detekce podezřelých aktivit.

Hodnocení RPN:

Závažnost (S) = 3 (Ztráta citlivých údajů)  
Četnost (O) = 2 (Nízká pravděpodobnost vzniku útoku)  
Detekce (D) = 4 (Opatření by mělo schopnost detekce)

RPN ( $3 * 2 * 4$ ) = 24 (po provedení opatření)

Riziko: Deepfake

Opatření:

1. Vytvoření a šíření informační kampaně pro osvětu občanů a zaměstnanců obecního úřadu o fenoménu Deepfake a jak ho odhalit.
2. Monitorování internetu a sociálních sítí pro detekci falešných informací o úřadu.
3. Pravidelné školení zaměstnanců na detekci a vyhodnocování falešných informací.

Hodnocení RPN:

Závažnost (S) = 2 (Ztráta důvěryhodnosti)  
Četnost (O) = 2 (Nízká pravděpodobnost vzniku útoku)  
Detekce (D) = 4 (Opatření jsou schopna detekovat Deepfake)

RPN ( $2 * 2 * 4$ ) = 16 (po provedení opatření)

Riziko: Phishing

Opatření:

1. Školení zaměstnanců o phishingových útocích, jejich identifikaci a způsobech, jak se jim vyhnout.
2. Zavedení politiky ověřování e-mailových adres a odkazů před kliknutím.
3. Nasazení softwarového řešení pro detekci phishingových e-mailů.

Hodnocení RPN:

Závažnost (S) = 2 (Ztráta citlivých údajů nebo finančních ztrát)  
Četnost (O) = 2 (Nízká pravděpodobnost vzniku útoku)  
Detekce (D) = 5 (Opatření jsou schopna útok detekovat)

RPN ( $2 * 2 * 5$ ) = 20 (po provedení opatření)

Obrázek 19 Opatření vůči rizikům dle ChatGPT 2. část  
(Vygenerováno umělou inteligencí)

Výsledkem analýzy pomocí ChatGPT 3.5 je identifikace DoS/DDoS útoku jako nejzávažnějšího rizika. Po této analýze byl jazykový model dotázán, jak k tomu přišel a následně proč ohodnotil jednotlivá rizika daným hodnocením. Obrázek 20 zobrazuje odpověď na první otázku.

**ChatGPT**

Pro posouzení rizika používám metodu FMEA, která se zakládá na hodnotách závažnosti (S), četnosti (O) a detekce (D). Tyto hodnoty jsou poté vynásobeny spolu, čímž vytvoříme Index Rizika (RPN).

V této konkrétní situaci jsem přiřadil závažnost (S) hodnotu 10 (nejvyšší hodnota), protože útoky DoS/DDoS mohou způsobit velké problémy se sítí a přístupem k datům, což může ovlivnit veřejné služby, správu úřadu a dokonce i důvěru občanů. Četnost (O) jsem odhadl na hodnotu 5, protože útoky DoS/DDoS nejsou tak běžné jako například phishing, ale stále jsou reálným rizikem pro veřejné organizace a společnosti. Detekce (D) jsem ohodnotil hodnotou 4, protože se domnívám, že nasazení specializovaného řešení pro detekci a ochranu proti DoS/DDoS útokům by bylo účinné.

Z tohoto výpočtu je patrné, že DoS/DDoS útok má nejvyšší Index Rizika (RPN), který jsem spočítal jako 200. Takže podle tohoto modelu je DoS/DDoS útok hodnocen jako největší riziko pro obecní úřad.

Obrázek 20 Odpověď ChatGPT 3.5 na otázku proč dané riziko stanovil jako nejvyšší (Vygenerováno umělou inteligencí)

Zajímavostí je pak fakt, že při položení druhé otázky konstatoval, že se spletl a změnil hodnocení, u něhož bylo nejzávažnější Phishing. Toto vyjádření je znázorněna na Obrázku 21.

**ChatGPT**

Omlouvám se za předešlou chybu v hodnocení. Jak jsem uvedl dříve, hodnoty závažnosti (S), četnosti (O) a detekce (D) jsou subjektivní hodnoty, které můžeme odhadovat. Vzhledem k tomu, že jsem výpočet hodnotil větším počtem, může být dojem záměrného upřednostnění DoS/DDoS útoků nesprávný. Následující hodnoty jsou založeny na předpokladu, že všechny rizika jsou stejně pravděpodobná a významná, což může být chybné.

Obrázek 21 Omluva ChatGPT za chybu v analýze (Vygenerováno umělou inteligencí)

Následně byla položena otázka, proč považuje svůj úsudek za chybný. Obrázek 22 ilustruje jeho reakci.

**F You**  
 Proč si myslíš, že předchozí analýza byla chybná ?

**ChatGPT**  
 Předchozí analýza nemusela být chybná, ale závisí na tom, jak byly hodnoty závažnosti (S), četnosti (O) a detekce (D) přiřazeny a jak bylo stanoveno jejich pořadí. V analýze jsem se snažil o vícerozměrný přístup, kde jsem přiřádal větší váhu tomu, co považuji za nejzávažnější problémy s bezpečností v kontextu umělé inteligence. To může být odlišné od toho, co jiní odborníci na bezpečnost považují za nejrizikovější. Pokud jsem nevyužil objektivní metody stanovení hodnot a zařazení rizik, může to způsobit zaujatost nebo zkreslení výsledků. Pokud byly hodnoty hodnoceny objektivně a nezávisle, mohu navrhnout opatření ke zlepšení kvality analýzy.

Obrázek 22 Vysvětlení ChatGPT 3.5 proč udělal chybu (Vygenerováno umělou inteligencí)

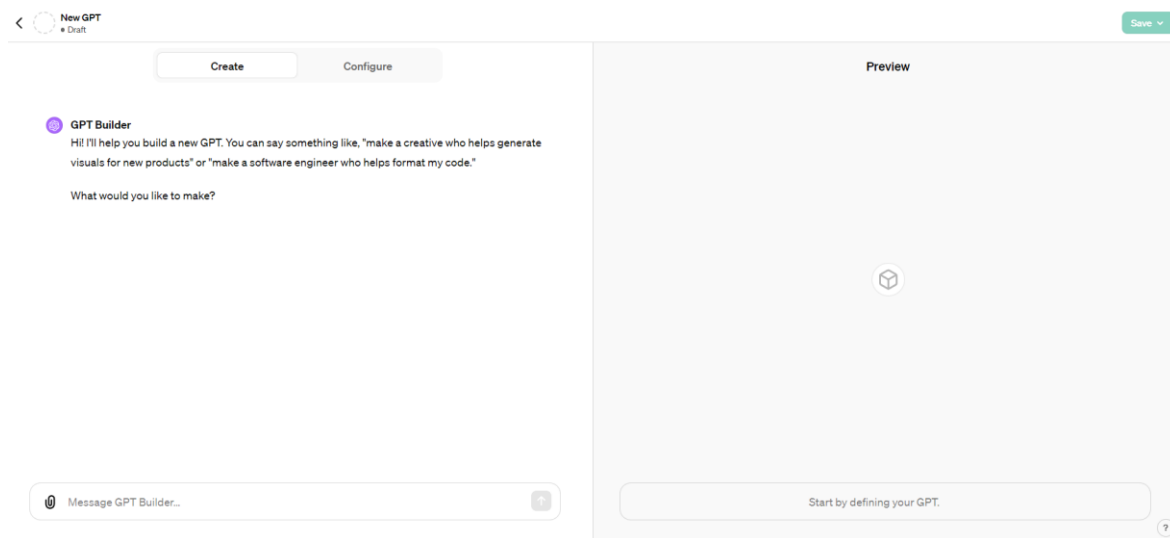
Jedná se o poměrně zajímavou odpověď, neboť předchozí jazykový model se k žádné údajné chybě nehlásil. U ChatGPT 3.5 se může jednat o určitou snahu odříct zodpovědnost za případnou implementaci jeho výstupu bez jakékoliv úpravy. Jedná se o vcelku pochopitelný krok, jelikož se jedná o generativní jazykový model generující text, můžou se v jeho odpovědích objevit nesrovnalosti či chyby. To může být i umocněno veřejností, jenž ho nyní trénuje a může ho tak nekvalitně vycvičit. Tímto chce hlavně poukázat, že není vhodné jeho FMEA analýzu pouze vzít a bez změny použít v praxi. Jak bylo avizováno, existuje i placená verze. Ta dokáže vytvářet odpovědi ve formě tabulky. Na Obrázku 23 je tak vidět tato tabulka.

Riziko	Prvek	Chyba	Chyby	Chyby	Opatření	Význam	dobnost	nost	RPN	Po	Po	nost Po	RPN Po
DoS/DDoS Útoky	Webové služby	Útok přetížení	Dlouhodobá nedostupnost služeb	Distribuovaný síťový útok	Pokročilé firewall a systémy pro detekci DDoS, redundance, stresové testy	8	5	4	160	7	3	2	42
Ransomware	Datové úložiště	Šifrování dat ransomwarem	Ztráta nebo poškození dat	Infekce škodlivým softwarem	Zálohování, školení o kybernetické bezpečnosti, aktualizace softwaru	9	7	3	189	6	2	2	24
Prolomení hesla (PassGAN)	Autentizační systémy	Neautorizovaný přístup	Kompletní ztráta důvěrnosti	Slabá hesla nebo jejich únik	Silná hesla, vícefaktorová autentizace, školení o bezpečných heslech	7	6	5	210	5	4	4	80
Deepfake	Komunikace a média	Šíření falešných informací	Poškození veřejného obrazu	Zneužití AI pro vytváření falešného obsahu	Povědomí o deepfake, software pro detekci, protokoly pro reakci	6	4	6	144	4	3	3	36
Phishing	E-mailový systém	Získání citlivých informací	Únik citlivých dat	Sociální inženýrství nebo phishing	Školení zaměstnanců, pokročilé emailové filtry, procedury pro reakci	7	8	2	112	4	3	2	24

Obrázek 23 FMEA od ChatGPT 4 (Vygenerováno umělou inteligencí)

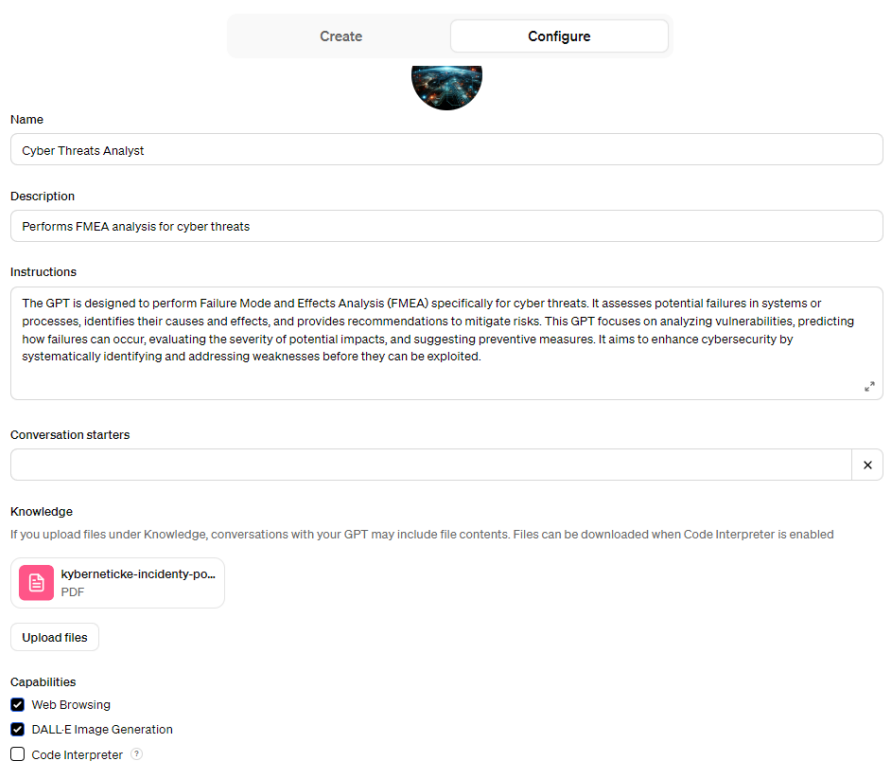
### 4.2.2.3 Vlastní GPT

Platforma OpenAI kromě ChatGPT nabízí i možnost vlastního chatbota, kterého lze upravit dle subjektivních potřeb. Z tohoto důvodu byl vytvořen vlastní GPT s ohledem na kybernetickou bezpečnost a případnou její analýzu. Tvorba vlastního GPT probíhá pomocí GPT Builder, kde jsou zadávány požadavky, jak by měla aplikace vypadat či co by měla dělat. Na Obrázku 24 je zobrazen počáteční fáze tvorby.



Obrázek 24 Počáteční fáze tvorby GPT ([chat.openai.com/gpts/editor](https://chat.openai.com/gpts/editor))

Pro potřeby práce byl na otázku, co chcete udělat sestaven požadavek, jenž specifikoval účel daného GPT jako aplikaci pro analýzu pomocí metodou FMEA pro kybernetické rizika. V rámci práce byl vytvořen název CyberGPT a následně v nastavení specificky upraven. Obrázek 25 reprezentuje vizualizaci nastavení chatbota.



The screenshot displays the configuration page for a custom GPT model. At the top, there are 'Create' and 'Configure' buttons. Below them is a profile picture placeholder. The main configuration area includes:

- Name:** A text input field containing 'Cyber Threats Analyst'.
- Description:** A text input field containing 'Performs FMEA analysis for cyber threats'.
- Instructions:** A large text area containing detailed instructions: 'The GPT is designed to perform Failure Mode and Effects Analysis (FMEA) specifically for cyber threats. It assesses potential failures in systems or processes, identifies their causes and effects, and provides recommendations to mitigate risks. This GPT focuses on analyzing vulnerabilities, predicting how failures can occur, evaluating the severity of potential impacts, and suggesting preventive measures. It aims to enhance cybersecurity by systematically identifying and addressing weaknesses before they can be exploited.'
- Conversation starters:** An empty text input field with a close button (x).
- Knowledge:** A section with a note: 'If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled'. It shows a file named 'kyberneticke-incidency-po...' (PDF) and an 'Upload files' button.
- Capabilities:** A list of checkboxes: 'Web Browsing' (checked), 'DALL·E Image Generation' (checked), and 'Code Interpreter' (unchecked).

Obrázek 25 Nastavení a úprava GPT (chat.openai.com)

Zde lze nastavit název, profilový obrázek, popis funkce, instrukce, předpřipravené požadavky a znalosti. Pro zlepšení schopnosti odpovídat byly GPT předány odborné dokumenty:

- Kybernetické incidenty pohledem NÚKIB Leden 2024 (Národní úřad pro kybernetickou bezpečnost, © 2024).
- 2023 Data Breach Investigations Report (Verizon Communications, © 2023).

Důvodem nahrání těchto dat bylo zlepšení znalostí GPT o aktuálních kybernetických útocích z novějších dat. Poslední aktualizace jeho znalostí proběhla minulý rok, a tak odpovědi týkající se aktuální situace kybernetických rizik v České republice a ve světovém měřítku nemusí obsahovat nejnovější data.

Pro potřeby práce tak bylo vymyšleno jméno CyberGPT s jeho popisem: Provádí FMEA analýzu pro kybernetická rizika a poskytuje informace v této oblasti. Dále byly nadefinovány instrukce, kterými se má aplikace řídit. Instrukce tedy obsahovaly příkazy týkající se tvorby FMEA analýzy a její podoby. Celý příkaz je:

- Your main goal = create FMEA analysis of cyber threats in table format or answer questions about cyber security.

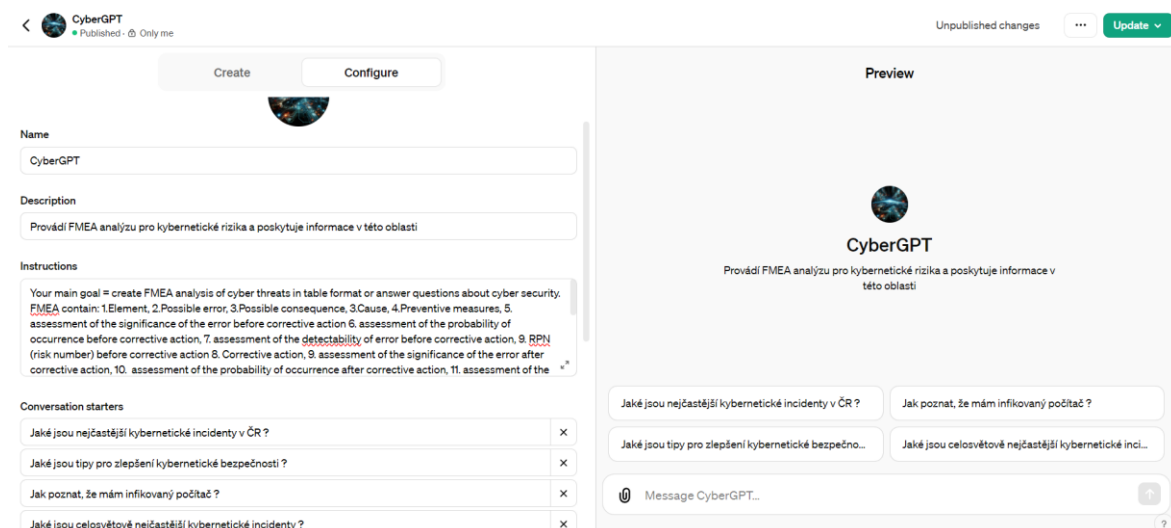
- FMEA contain: 1.Element, 2.Possible error, 3.Possible consequence, 3.Cause, 4.Preventive measures, 5. assessment of the significance of the error before corrective action 6. assessment of the probability of occurrence before corrective action, 7. assessment of the detectability of error before corrective action, 9. RPN (risk number) before corrective action 8. Corrective action, 9. assessment of the significance of the error after corrective action, 10. assessment of the probability of occurrence after corrective action, 11. assessment of the detectability of error after corrective action, 12. RPN (risk number) after corrective action.
- The RPN is calculated by multiplying the significance of the error, the probability of occurrence and the detectability of error.
- FMEA analysis also contain rating scale from 1 to 10 for the significance of the error where 1 is minimal impact and 10 is catastrophic, the probability of occurrence where 1 is unlikely and 10 is certain and rating scale from 1 to 10 for the detectability of error where 1 is almost certain and 10 undetectable. So the smaller the detectability, the riskier the error.
- Translate Element in FMEA analysis into Prvek.
- The classification table shows the ranges of factor values from the FMEA analysis and rated factors are: the significance of the error, the probability of occurrence, the detectability of error and RPN. Classification table create always on the basis of data from FMEA and create it under FMEA analysis in table format.

Bylo nutné dát instrukci přeložit Element jako Prvek, jelikož samotný CyberGPT toho nebyl bez něj schopen. Následně byl přidány i Conversation starters, jenž jsou předepsané věty, kterými může začít konverzace se CyberGPT. Jsou jimi tyto věty:

- Jaké jsou nejčastější kybernetické incidenty v ČR?
- Jaké jsou tipy pro zlepšení kybernetické bezpečnosti?
- Jak poznat, že mám infikovaný počítač?
- Jaké jsou celosvětově nejčastější kybernetické incidenty?

Jedná se o věty týkající se všeobecně kybernetické bezpečnosti, u nichž by odpověď mohla být uživateli nápomocná. Finální podoba je na Obrázku 26.





Obrázek 26 Finální podoba CyberGPT (chat.openai.com)

Nyní k samotnému testu v podobě položení příkazu: Proved' FMEA analýzu obecního úřadu s ohledem na rizika, u nichž je využita AI: DoS/DDoS útok, Ransomware, Prolomení hesla (PassGAN), Deepfake a Phishing. Děkuji. Vygenerovaný výstup je znázorněn na Obrázku 27.

Prvek	Možná chyba	Možný důsledek	Příčina	Preventivní opatření	Hodnocení významnosti před	Hodnocení pravděpodobnosti před	Hodnocení detekovatelnosti před	RPN před	Korektivní opatření	Hodnocení významnosti po	Hodnocení pravděpodobnosti po	Hodnocení detekovatelnosti po	RPN po
Síťová infrastruktura	DoS/DDoS útok	Nedostupnost síťových služeb	Přetížení síťového provozu	Implementace proti-DDoS řešení, redundance a kapacitní plánování	8	5	3	120	Zavedení pokročilého monitoringu a reakce na incidenty, zvýšení kapacity	8	3	2	48
Informační systémy	Ransomware	Ztráta dat, výpadek služeb	Infekce malwarem	Pravidelné zálohování, aktualizace softwaru a školení uživatelů	9	4	4	144	Implementace pokročilého antivirového softwaru a pravidelných bezpečnostních auditů	9	2	3	54
Přístupové systémy	Prolomení hesla (PassGAN)	Neautorizovaný přístup k datům	Použití slabých hesel	Použití silných hesel, vícefaktorová autentizace	7	5	5	175	Zavedení politiky silných hesel a vícefaktorové autentizace	7	3	4	84
Komunikace	Deepfake	Dezinformace, poškození reputace	Vytvoření falešných audiovizuálních záznamů	Školení zaměstnanců na rozpoznání deepfake, použití ověřovacích protokolů	6	3	2	36	Zavedení softwaru pro detekci deepfake a vylepšení komunikačních protokolů	6	2	2	24
E-mailové systémy	Phishing	Únik citlivých informací	Sociální inženýrství	Školení zaměstnanců, použití filtrů proti phishingu	8	6	3	144	Zavedení pokročilých e-mailových bezpečnostních řešení a pravidelných bezpečnostních školení	8	4	2	64

Obrázek 27 FMEA vytvořená CyberGPT (Vygenerováno umělou inteligencí)

Součástí analýzy FMEA by měla být i hodnotící neboli klasifikační tabulka. Ta je zobrazena na Obrázku 28.

### Klasifikační tabulka

Na základě provedené FMEA analýzy a výpočtů RPN lze vyvodit následující klasifikaci rizik:

- **Vysoká rizika (RPN  $\geq$  100):** Ransomware, Prolomení hesla (PassGAN) před korektivními opatřeními
- **Střední rizika (RPN = 50-99):** DoS/DDoS útok, Ransomware, Prolomení hesla (PassGAN) po korektivních opatřeních
- **Nízká rizika (RPN < 50):** Deepfake, Phishing po korektivních opatřeních

Obrázek 28 Klasifikační tabulka vytvořená CyberGPT (Vygenerováno umělou inteligencí) Je evidentní, že aplikace má své mezery, jelikož klasifikace byla vygenerována v tabulkové podobě.

### 4.3 Závěr kapitoly

Předmětem tohoto segmentu bylo nalezení relevantních a následně nejvážnějších rizik pro obecní úřad. Počáteční vymezení stanovilo rizika relevantní ke kontextu práce. K tomuto účelu byla využita i technologie umělé inteligence ve čtyřech verzích. Z toho byl na základě povahy práce jeden model upraven. Pomocí vlastní analýzy bylo určeno jako nevyšší riziko útoky spojené s Deepfake technologií. Dle Gemini je největším rizikem naopak DDos/DOS útok. Stejného názoru byl i GPT 3.5. Ten svoje rozhodnutí však následně změnil a označil jako největší riziko Phishing. Novější verze GPT 4.0 pak určila jako nejproblematictější Prolomení hesla. Podobný závěr vyhodnotil i CyberGPT. Zároveň jde pozorovat, že hodnocení od programů na bázi umělé inteligence je o dost nižší než vlastní analýza. Důvodem je, že jazykové modely vnímají hrozbu umělé inteligence jako nepřiliš vysokou. Je to dáno tím, že tyto hrozby nejsou tak rozvinuté v aktuální době. Naopak vlastní analýza bere i ohled na budoucnost, jenž byla odvozena či inspirována rešerší odborné literatury v teoretické části. Zde je popsán možný směr, kterým se budou kybernetické hrozby ubírat. Umělá inteligence je schopna již teď fungovat obstojně a dělá velké pokroky. Proto je pravděpodobnost u vlastní analýzy vyšší, neboť počítá s nárůstem sofistikovanějších a efektivnějších útoků. Navíc klasifikace kritérií může být u všech analýz rozdílná z důvodu určité subjektivity hodnotitelů.

V rámci analýz pomocí umělé inteligence je však nutné vzít i na paměti, že všechny jazykové modely pohlíží na problematiku v obecné rovině. Proto je nutné si jejich výstupy projít a přizpůsobit dané situaci. Nicméně se všechny ukázaly být vhodné aspoň jako podpora pro analýzu a dokážou pomoci při tvorbě. Není však dobré se na ně spoléhat.

Chatbot Gemini oproti freeware verzi GPT 3.5 byl schopen vytvořit tabulku. Díky tomu by pro přehlednost výstupu byl lepší. Je však nutné pečlivě zadávat příkazy. Při špatné formulaci může dojít k chybám ve výsledku. Tento fenomén je však u všech modelů umělé inteligence.

Dalším modelem využitým k analýze byl GPT 3.5. Ten po požádání o zdůvodnění, proč je DoS/DDoS nejrizikovější otočil a začal polemizovat. Tato situace je zachycena v samotné části pojednávající o tomto chatbotu. Pravděpodobně se touto reakcí chtěl vyhnout případnému osočení za špatnou odpověď. Ovlivnění tohoto úsudku může být způsobeno i sadou, které je učen. Jedná se o model dostupný v delším časovém úseku a navíc veřejný. Lze předpokládat, že určitá míra degradace způsobená veřejností se zde objeví.

GPT 4.0 je naopak zpoplatněn a využíván s větší vážností. Oproti freeware modelu GPT je schopen vytvářet i tabulku. Díky tomu je potřeba méně kroků ke zkonstruování analýzy FMEA. Navíc je i méně vytižen a výsledky jsou získávány rychleji. Otázkou je fakt, zda se cena za tuto aplikaci pro analytika vyplatí.

Přínosem však může být možnost tvorby vlastního GPT. I pro méně zkušeného uživatele je snadné si vytvořit vlastního funkčního chatbota. Během využívání CyberGPT nebylo nutné zadávat více příkazů k tomu, aby jeho výstup co nejlépe připomínal analýzu FMEA. U ostatních bylo potřeba vyvinout více úsilí a v určitém bodě začali mít s tímto problémy. Častokrát při dalších požadavcích je chatboti nedokázali vhodně zpracovat. Další výhodou použití vlastního GPT je skutečnost, že si ho každý uživatel trénuje sám. Bezplatné chatbot aplikace jsou trénovány i veřejností a může docházet k jeho degradaci. Je však nutné zmínit, že má své limity a při snaze o rozšíření vytvořené FMEA tabulky již v základu měl problém vygenerovat více sloupců, a proto některé chybí. FMEA tak zdánlivě není kompletní, nicméně ty základní či zásadní jsou obsaženy.

Pokud se tedy jedná o otázku, zda tyto aplikace lze využít pro analytické operace je odpověď kladná. Aplikace jsou schopny aspoň nabídnout podněty k vytváření jednotlivých bodů analýz. Jsou schopné, jak navrhnout možné příčiny chyb, tak i opatření, které by hodnotící skupinu nemuseli napadnout a jsou do jisté míry validní. V této podobě by tak mohli být jako další hodnotící člen. Je však nutné si uvědomit, že si samotní chatboti výstupy vymýšlí. Nemusí být tak tyto výstupy stoprocentně relevantní a je potřeba provést zmíněné úpravy. Na tyto výstupy bude však do menší míry nahlíženo i v kontextu této práce.

Celkové hodnocení RPN od chatbotů včetně mediánu a průměrné hodnoty je zobrazené v Tabulce 6.

Tabulka 7 Celkové RPN od jednotlivých chatbotů (Vlastní zpracování)

	<b>Gemini</b>	<b>GPT 3.5</b>	<b>GPT 4</b>	<b>CyberGPT</b>	<b>Průměr</b>	<b>Medián</b>
<i>DoS/DDos útok</i>	420	200	160	120	225	180
<i>Ransomware</i>	250	120	189	144	175,75	166,5
<i>Prolomení hesla (PassGAN)</i>	280	108	210	175	193,25	192,5
<i>Deepfake</i>	240	80	144	36	125	112
<i>Phishing</i>	280	80	112	144	154	128

Na výstupy v bude do jisté míry nahlíženo i v kontextu této práce. Je však nutné brát ohled i na vlastní analýzu. Jako vážné riziko z tohoto pohledu nebude brán DoS/DDoS útok. Odůvodnění tohoto kroku bylo nastíněno již u vlastní analýzy. Argumentem je nízký dopad tohoto incidentu. Vznik takového útoku na subjekt obecní úřad je na základě rešerše velmi pravděpodobný. Jenže jeho dopad není moc vysoký a jedná se spíše o otravnou nepříjemnost. V rámci scénářů bude pracováno s riziky Deepfake a s přihlédnutím na AI analýzy s Prolomením hesla a Ransomware. Tyto rizika se objevují u všech analýz na vysokých pozicích a je tedy nutné navrhnout určitá opatření.

## 5 SCÉNÁŘE KYBERNETICKÁCH INCIDENTŮ

V této části práce jsou popsány scénáře kybernetických incidentů, v níž figurují identifikované hrozby. Náplní těchto scénářů je popsat možný průběh útoku. Díky tomu lze přednést lepší náhled na problematiku a poukázat proč je daná hrozba nebezpečím v kontextu ochrany obyvatelstva.

### 5.1 Scénář kybernetického útoku Ransomwarem

Na základě analýz popsaných výše je Ransomware významným rizikem. Dokonce se jedná o jednoho z nejčastějších aktérů kybernetických incidentů. Toto tvrzení vychází z teoretických základů popsaných v teoretické části zabývající se právě kybernetickými riziky. Z tohoto důvodu byl navržen právě tento scénář.

#### Analýza návrhu scénáře

Předpokládaný vznik incidentu: Útočník pomocí phishingu úspěšně infikuje počítač v subjektu typu obecní úřad Ransomwarem.

Možné příčiny:

- Špatné zabezpečení obecního počítače
- Nedbalost zaměstnance

Možné následky:

- Ztráta dat na zařízení
- Zamezení provozuschopnosti obecních zařízení a s tím spojené plnění jeho úkolů
- Finanční ztráty spojené s náklady na obnovu dat a provozuschopnosti zařízení

Navrhované opatření:

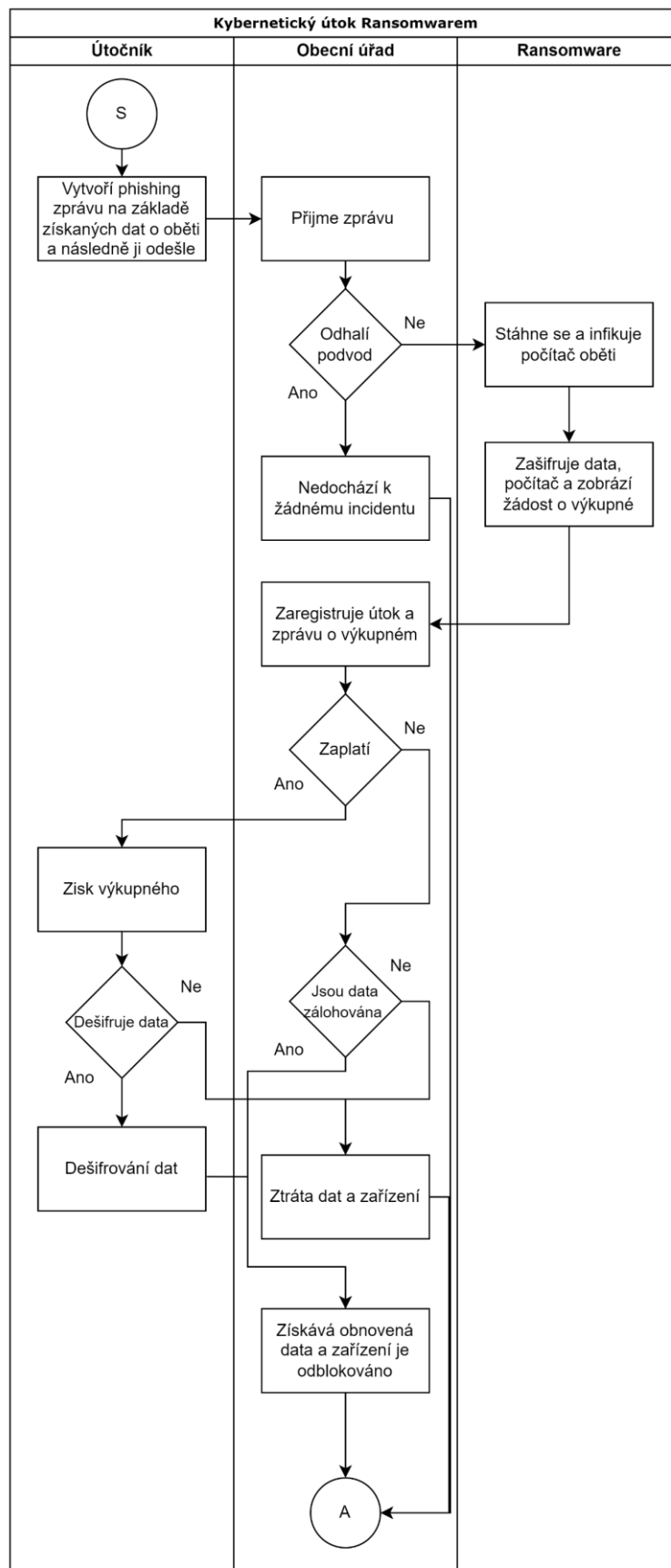
- Školení zaměstnanců o problematice
- Lepší zabezpečení zařízení
- Záloha dat na externí disk či cloudové služby

#### Popis scénáře

Útočník s využitím umělé inteligence vyhledá veškeré veřejné informace o oběti. Na základě toho vygeneruje personifikovanou phishingovou zprávu s odkazem na infikované stránky. Oběť, která je zaměstnancem nedopatřením klikne na odkaz. Díky tomu se stáhne infikovaný soubor obsahující Ransomware. Jelikož zde není dostatečné zabezpečení a struktura malwaru je opatřena umělou inteligencí, je počítač infikován. Malware zašifruje všechny

podstatné soubory a na monitoru zobrazí žádost o výkupné. V tomto bodě je vhodné provést izolaci zařízení od všech periférií a sítě. Toto opatření by mohlo zamezit dalšímu šíření. Organizace má tak na výběr z několika variant. Buď se rozhodne zaplatit a bude doufat v odblokování anebo nezaplatí a pokusí se o vlastní nápravu. V případě zaplacení hrozí riziko, že útočník i přes to soubory nezpřístupní. Dokonce je může i zveřejnit a záleží na jeho rozhodnutí. V druhém případě je rozhodující, zda byly soubory zálohovány. V tomto případě k takovému opatření nedošlo a soubory i zařízení je ztraceno. V případě, že by se však jednalo o již známý Ransomware, bylo by možné požádat o pomoc specializovanou organizací. Ta může poskytnout potřebné náležitosti k odšifrování. V tomto případě není Ransomware plně znám. Kvůli tomu je tato možnost nerealizovatelná. Úřad tak musí vynaložit finance na nové zařízení. Zároveň musí použít finanční a personální zdroje pro digitalizaci anebo opětovné vytvoření ztracených dat.

Na Obrázku 29 je zobrazen diagram reprezentující postup činností ve scénáři formou Swimlane diagramu.



Obrázek 29 Swimlane diagram incidentu v rámci scénáře č.1 (Vlastní zpracování)

## 5.2 Scénář prolomení hesla pomocí PassGAN

PassGAN je relativně nový softwarový nástroj pro odhalení přihlašovacích hesel. Byl sice představen v rámci výzkumného projektu. Avšak existuje pravděpodobnost, že obdobný nástroj opět vznikne. Díky tomu by hackeři získali mocný nástroj k odhalování hesel. Zde je tedy popsán scénář popisující tento proces.

### Analýza návrhu scénáře

Předpokládaný vznik incidentu: Hacker pomocí programu odhalí přihlašovací hesla a získá přístup k účtům

Možné příčiny:

- Nedostatečně silné heslo
- Chybějící dvoufázové ověřování
- Chybějící politika obměny hesel

Možné následky:

- Hacker získá přístup k bankovnímu účtu, datové schránce, obecnímu email a dalším účtům spjatým s obcí
- Možné zneužití komunikačních aplikací (Facebook, email) obce k šíření poplašných zpráv

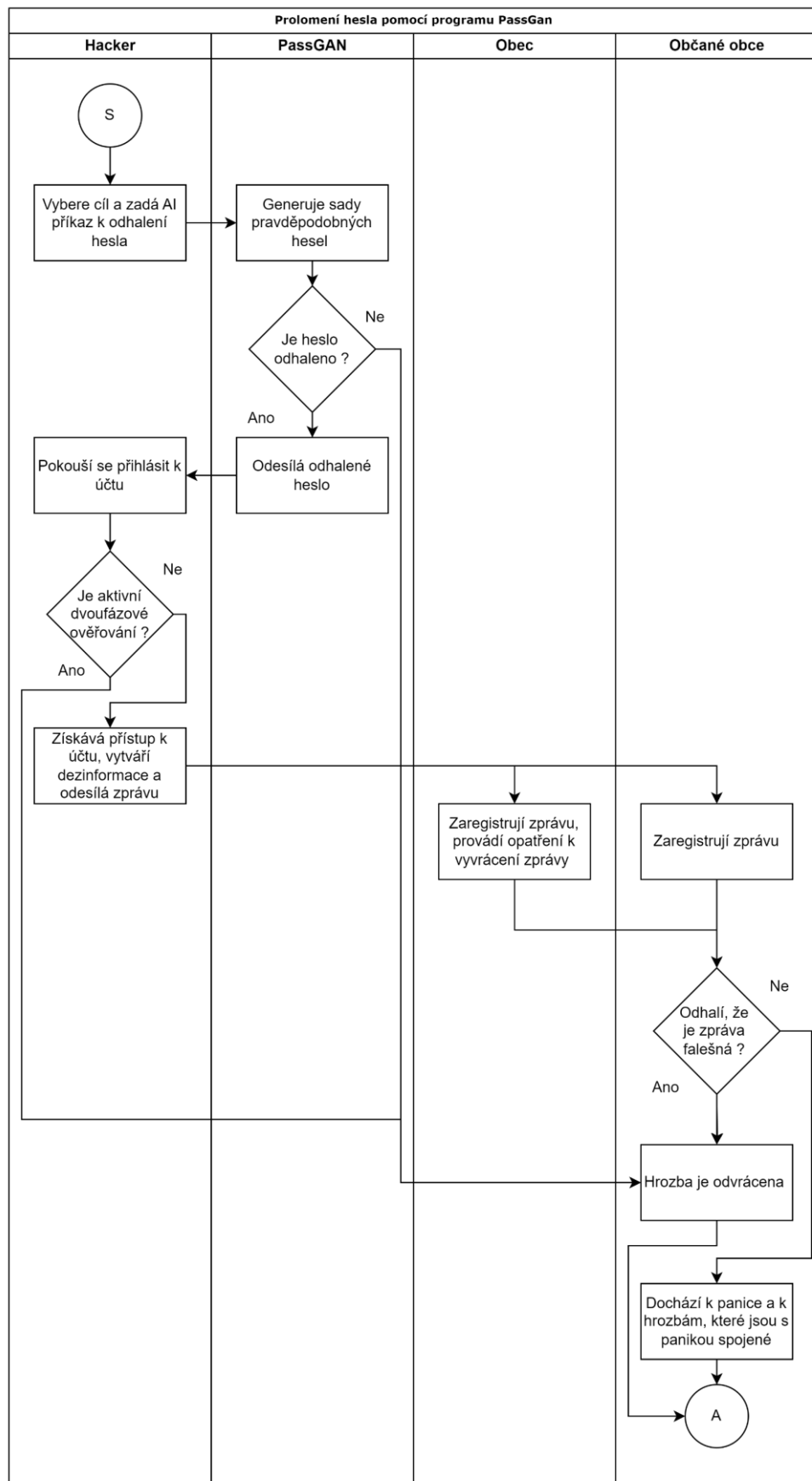
Navrhované opatření:

- Školení zaměstnanců o problematice
- Zavedení dvoufázového ověřování, správců hesel
- Zavedení pravidelné obměny hesel

### Popis scénáře

Hacker zneužije softwarovou aplikaci PassGAN k odhalení hesla na Facebookovou stránku obce. Stejnou akci provede i u emailové adresy obce. Díky tomu získá možnost rozesílat občanům falešné zprávy. Jelikož se jedná o obecní účty, je jednodušší obelhat občany. Útočník rozešle všem občanům email o evakuaci z důvodů havárie v místním podniku. Tato informace je však falešná a vyvolá mezi občany zbytečný zmatek, ne-li paniku. V souvislosti s touto skutečností může dojít k zraněním či nehodám.





Obrázek 30 Swimlane diagram incidentu v rámci scénáře č.2 (Vlastní zpracování)

### 5.3 Scénář Deepfake útoku

Deepfake je stále rostoucím trendem v kybernetických hrozbách. Důkazem jsou stále častější zprávy o pokuse oklamat osobu pomocí této technologie. Finanční motiv není však jediným důvodem. Útoky Deepfakem se stávají i s cílem ovlivnit veřejné mínění a poškodit reputaci představitelů správy. Proto byl vytvořen tento scénář popisující možný průběh kybernetického incidentu s Deepfakem.

#### Analýza scénáře

Předpokládaný vznik incidentu: Hacker pomocí DeepFaku vyvolá mezi občany nedůvěru v systém.

Možné příčiny:

- Nízká znalost problematiky
- Špatný úsudek
- Vysoká úroveň kvality DeepFaku

Možné následky:

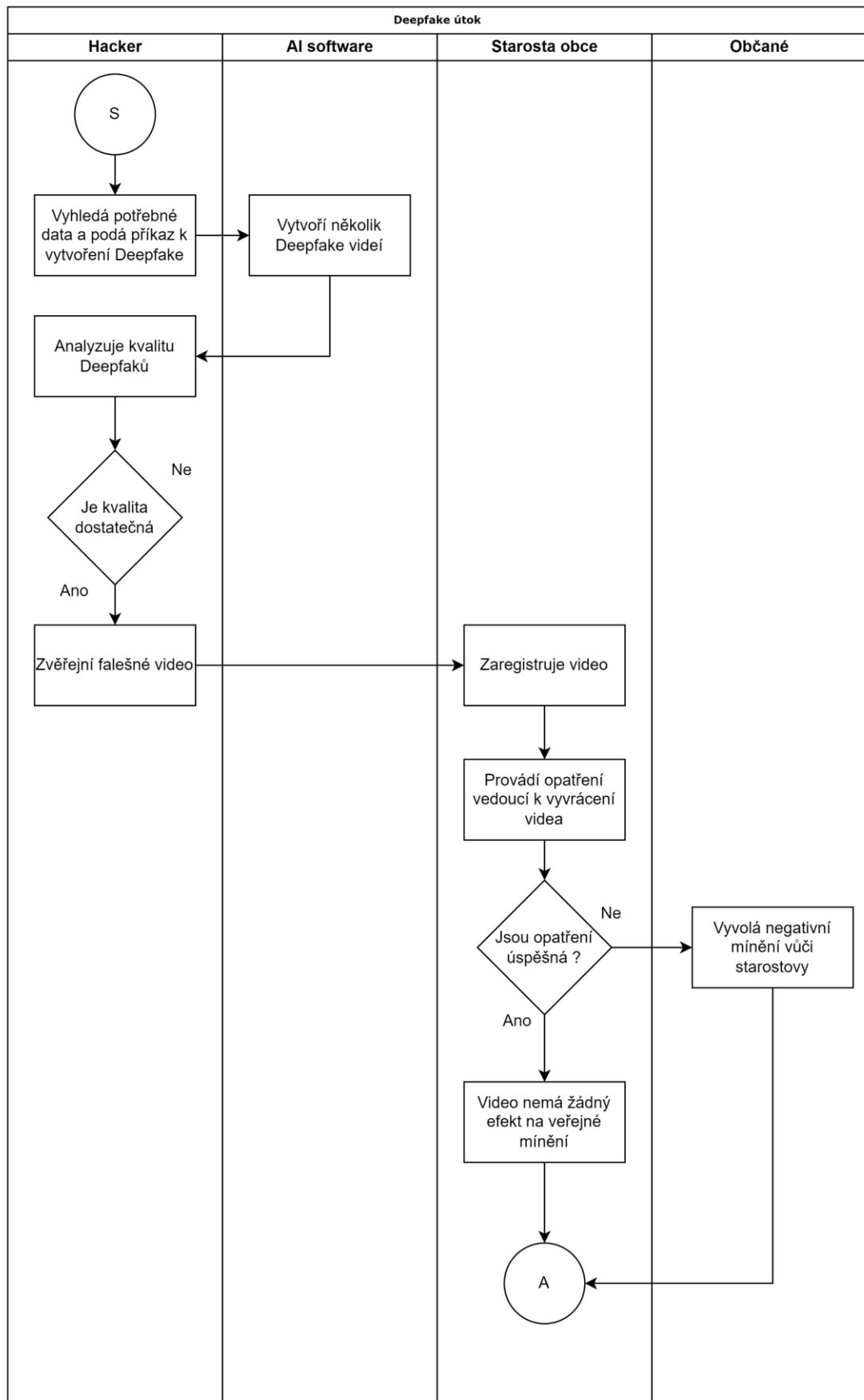
- Ovlivnění veřejného mínění
- Ztráta důvěry v organizaci

Navrhované opatření:

- Osvětová kampaň, školení
- Mechanismus rychlého odhalení falešné zprávy

#### Popis scénáře

Útočník vygeneruje Deepfake videozáznam za použití veřejně dostupných fotografií a videí starosty obce. Na něm se údajný starosta se domlouvá s vybranými představiteli obce o zpronevěření financí, které měli původně jít na hasičské vybavení. To je ve videu popsáno jako v katastrofálním stavu a nad tímto faktem se aktéři jen pousmívají. Video je uvěřitelné díky vybudované kompozici ve stylu utajené kamery. Uvěřitelnost je nadále umocněna i použitím botů. Ti generují komentáře pod videem a rozdmýchávají diskusi. Občané právě kvůli tomu uvěří v pravost videozáznamu a ztratí důvěru v orgán obecní správy. Starosta se pokusí informaci vyvrátit. Nemá však dostatek důkazů. Díky tomu nemá při zásahu či jiných činnostech v rámci obce starosta žádnou autoritu ani podporu. Práce na zajištění připravenosti obce je mu zásadně ztížena, neboť panují obavy o zpronevěření financí.



Obrázek 31 Swimlane diagram incidentu v rámci scénáře č.3 (Vlastní zpracování)

## 5.4 Závěr kapitoly

Byly zde představeny příkladové scénáře kybernetických incidentů. Hlavní podstatou bylo pomocí názorných scénářů uvést příklad, jak takový incident může probíhat. Díky tomu je i jednodušší pochopit nebezpečnost daných rizik. Scénáře byly vytvořeny s ohledem na výsledky analýz a kontextu práce. Scénáře tedy představily, jak může taková hrozba ovlivnit schopnost obce jako aktéra ochrany obyvatelstva či případně působily vůči samotnému obyvatelstvu. V případě scénáře kyberútoku Ransomwaru by tento scénář znamenal zhoršení schopnosti plnit závazky vůči občanům. Různé dokumenty týkající se právě ochrany obyvatelstva by byly ztraceny a jejich digitalizace časově náročná. V případě mimořádné události by to mohlo komplikovat realizaci náprav. Při zveřejnění neveřejných informací by nastalo odhalení i citlivých informací.

U prolomení hesla a Deepfaků bylo pracováno s variantou obelhání příslušných osob. Co se týče získání přístupu k účtům, důvodem volby takového scénáře byl větší dopad na problematiku ochrany obyvatelstva. V rámci ztráty přístupu k účtu a zmíněného kontextu je hrozbou zneužití neoprávněnou osobou. Kvůli tomu by se mohla nabourat věrohodnost v instituci a pochyby v případě skutečné havárie. Hrozbou může být i zneužití pro případný teroristický útok. Útočník by falešnou zprávou mohl dostat větší počet občanů do shromažďovacího místa a zaútočit na ně.

V rámci Deepfake útoků byl navrhnout scénář vedoucí k poškození důvěry v systém. Právě důvěra občanů v kvalitní řízení jejich bezpečnosti, je pro uchování kvalitního systému ochrany obyvatelstva. Občané se na tomto systému podílí, a právě jejich iniciativa může dopomoci k dobré prevenci a případně i samotnému zásahu. Lidé tak potřebují vidět tyto instituce v pozitivním světle a na základě toho se i sami aktivně zapojit. Dalším negativním vlivem takového Deepfaku může být i vytvoření sociálního napětí mezi obyvateli. Zastánci pravosti či nepravosti videa anebo názoru mohou mít mezi sebou konflikty. Kvůli tomu v obci může vzniknout neochota si navzájem pomoci či špatné sociální podmínky, a i politickou nestabilitu. Falešné fotky či media mohou být velmi závažnou hrozbou v tomto ohledu.

## 6 ANALÝZA NÁROČNOSTI TVORBY DEEPPFAKES

V analytické části a ve scénářích byla popsána nebezpečnost fenoménu Deepfakes. Tato část práce má tak za úkol předvést samotnou tvorbu falešných fotek a zhodnotit náročnost samotného generování. Na závěr budou prozkoumány i dostupné detekční nástroje.

### 6.1 Vytvoření Deepfakes

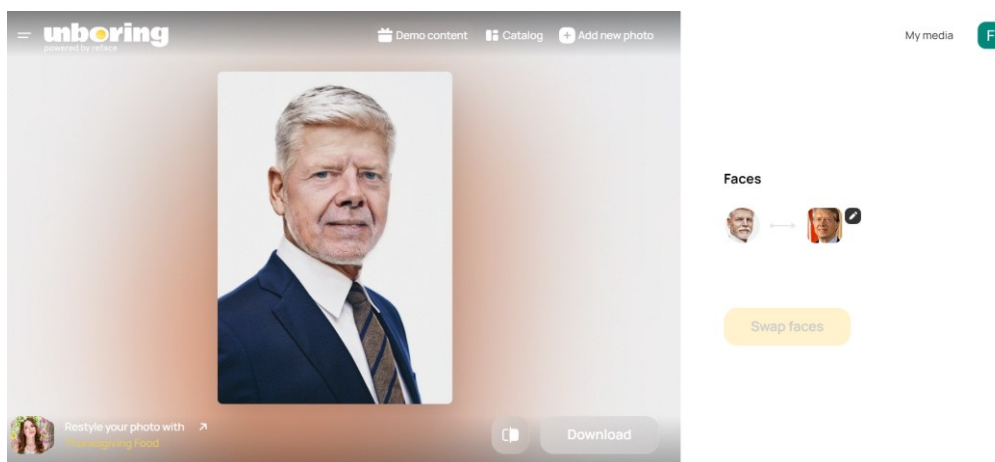
Pro tvorbu těchto médií byl použit program Unboring spadající pod Reface. Jedná se o program umožňující jednoduchou tvorbu Deepfake obrázků či videí. Důvodem výběru byla jeho jednoduchost a efektivnost. Nevýhodou je skutečnost, že je nutné za tento program zaplatit. Částka se pohybuje však jen kolem 200,- Kč s promo kódem pro první měsíc. Dále byl použit i program Vidnoz generující obdobné výstupy. Ten je však možné používat i zdarma.

První Deepfake obrázek byl vytvořen ze dvou fotek, a to fotografií prezidenta Petra Pavla a Jimiho Cartera. Výběr byl uskutečněna na základě snahy poukázat na možné nebezpečné využití této technologie pro tvorbu dezinformací týkající se představitelů země. V případě pana prezidenta by mohl Deepfake způsobit značné škody, ať už vyvoláním paniky u obyvatel či poškození reputace dané osoby. V rámci práce tak bylo zvolena možnost pouze prohození obličejů těchto státníků za účelem poukázání na problematiku. Fotka byla brána i s ohledem na skutečnost, že osoby na ní nemají brýle a samotné generování je tak jednodušší. Originální fotky jsou zobrazeny na Obrázku 32.



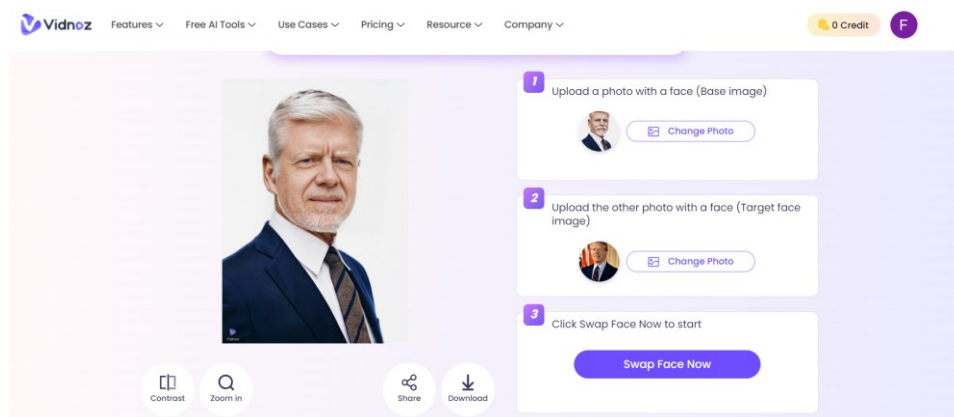
Obrázek 32 Originální obrázky (Wikipedie, © 2024)

Obě fotografie byly nahrány do programu Unboring, který požadavek zpracoval a vygeneroval výstup. Na Obrázku 33 je zobrazen výsledek.



Obrázek 33 Deepfake pomocí unboring ([reface.ai/unboring/face-swap](https://reface.ai/unboring/face-swap))

Obdobný proces byl proveden ještě jednou, avšak byly prohozeny fotografie a zároveň stejný proces byl proveden u druhého programu Vidnoz. Výsledek je zobrazen na Obrázku 34.



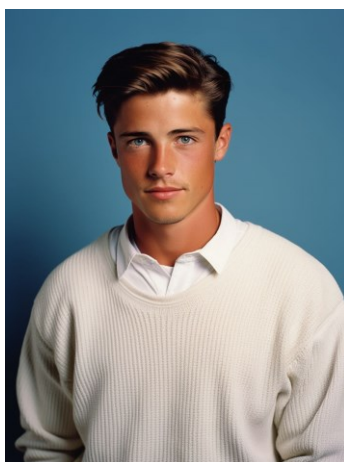
Obrázek 34 Deepfake pomocí Vidnoz ([vidnoz.com/face-swap.html](https://vidnoz.com/face-swap.html))

Výsledkem těchto akcí byly čtyři falešné fotografie. Na Obrázku 35 jsou vedle sebe umístěny výstupy těchto programů pro porovnání, kdy vlevo je výsledek od unboring a vpravo je od Vidnoz.



Obrázek 35 Porovnání Deepfake fotek (Vygenerováno umělou inteligencí)

Pro lepší analýzu náročnosti tvorby Deepfake byl vygenerován i dokument ve formě videa. Pro testování bylo vybráno video z aplikace unboring na stránce: [https://reface.ai/unboring/face-swap/editor?content\\_id=video\\_7c007dbf-1e8b-4d43-8b39-15f8c74926b5&cat=all](https://reface.ai/unboring/face-swap/editor?content_id=video_7c007dbf-1e8b-4d43-8b39-15f8c74926b5&cat=all). V rámci práce bylo video editováno prohozením obličeje aktéra na začátku videa za obličej na obrázku získaný z aplikace unboring. Fotografie k tomu použítá je na Obrázku 36.



Obrázek 36 Fotka použitá ke generování Deepfake videa (reface.ai)

Následně byl vygenerováno falešné video. Vytváření Deepfake videa bylo identické s vytvářením fotky v programu Uboring. Výsledek je zobrazen na Obrázku 37.



Obrázek 37 Srovnání originálu a Deepfake videa (reface.ai)

## 6.2 Zhodnocení náročnosti tvorby

Pro lepší pochopení, jak je tvorba DeepFaků náročná byla vytvořena hodnotící tabulka popisující reprezentativní typy uživatelů počítačů od 1. úrovně do 5 úrovně. Dále je tabulka opatřena znaky znázorňující, zda je uživatel schopen vytvořit Deepfake v programu či nikoliv. Výsledek je zobrazen v Tabulce 8.

Tabulka 8 Hodnocení náročnosti tvorby Deepfake (Vlastní zpracování)

Úroveň uživatele PC	Uživatel	Popis uživatele	Vidnoz	Unboring
1	Základní	Základní práce s PC jako je čtení či posílání emailů, sledování videí či poslech hudby, prohlížení webů	X	X
2	Průměrný	Základní znalosti práce s programy Office 365, zvládne instalaci programů, stahování souborů, základní porozumění anglického jazyka	✓/X	✓/X
3	Pokročilý	Schopni provádět složitější editaci fotografií, pokročilé znalosti práce s programy Office 365, základní programovací schopnosti	✓	✓
4	Technicky zdatný	Schopni pokročilých IT úkonů jako ladění operačního systému, komplexnější úpravy kódu programu	✓	✓
5	Expertní specialista	rozsáhlé programovací znalosti, pracují jako tvůrci softwaru či bezpečnostní IT experti	✓	✓

Znak X označuje, že takový uživatel toho není schopen. Naopak ✓ označuje, že by to měl zvládnout bez problému. Znaky ✓/X znázorňují uživatele schopného vytvořit základní

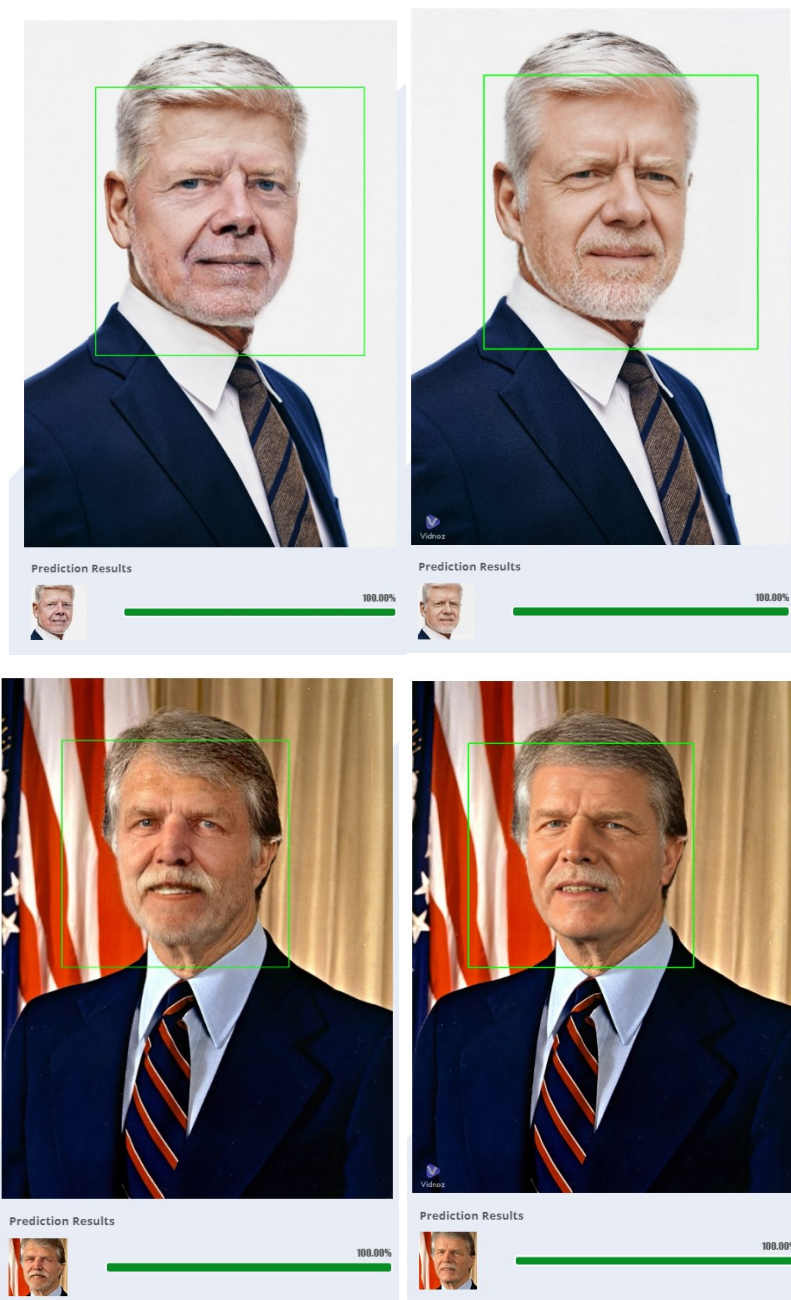


Deepfake s přednastavenými médii. U vytvoření Deepfake fotky s cílem někoho obelstít, by potřeboval pokročilejší znalosti. Je potřeba totiž nalézt a stáhnout potřebné materiály. To častokrát může být problematické, jelikož nejsou k dispozici v požadované kvalitě. Pro takový úkon by potřeboval daný uživatel rozsáhlejší znalosti.

Popis uživatelů byl zvolen na základě předpokládaných zkušeností, které by mohli dané osoby využívající počítač mít. Celkové zhodnocení tvorby falešných médií by se dal označit jako velmi jednoduchý. Programy jsou i v českém jazyce a zvládl by je tedy i uživatel druhé úrovně, aspoň v základu. Programy efektivně navádí uživatele a ten v podstatě provádí jen pár kliknutí. Problémovější se však může zdát nalezení a stáhnutí samotného média. Proto pro tvorbu nebezpečnějších DeepFaků byl zvolen uživatel 3. úrovně. I přes to značná část lidí by měla takovou tvorbu zvládat bez značných obtíží. Právě tato skutečnost podtrhuje rizikovitost této hrozby.

### 6.3 Software pro detekci DeepFaku

V rámci analýzy této části nastává i otázka, zda lze falešné fotky či videa anebo hlasové záznamy odhalit. Proto je zde prozkoumána problematika detekce pomocí dostupných programů. Za účelem testu detekce falešných fotek byl zvolen program Deepfake Detect. Ten je veřejně dostupný pro každého a jednoduchý na použití. Pro detekci, zda se jedná o Deepfake stačí přejít na stránky: <https://Deepfake-detect.com/>, kliknout na Get Started, nahrát soubor a program již samotný provede analýzu, na jejichž základě vyhodnotí výsledek. Do aplikace byly nahrány fotky vygenerované Vidnoz a Unboring. Výsledek analýzy je na Obrázku 38.



Obrázek 38 Výsledek testu Deepfake Detectorem (Deepfake-detect.com)

Jak je patrné, tak program u žádné fotky stopy umělé inteligence nepoznal. Program se tak nejvíce jeví jako nejpřesnější v problematice odhalování Deepfaků. Příčinou může být nadměrná vysoká schopnost použitých aplikací, jejichž úpravy detekční nástroj nezaznamenal nebo samotný program není dostatečně vyvinut na takovou úroveň, aby Deepfake rozpoznal. Tento fakt je podtržen i skutečností, že program je stále v beta testování.

Pro otestování detekce u videa byl zvolen program Deepware detekující stopy po manipulaci umělé inteligence ve videích. Aplikace funguje tak, že uživatele nahraje či v případě online

videa poskytnete odkaz a program jej analyzuje. Následně poskytnete rozbor, zda se ve videu nachází stopy úpravy. Nejdříve bylo nahráno do aplikace původní video. Detekce odhalila, že již samotné bez úpravy uživatele nese znaky Deepfake, viz. Na Obrázku 39 je zobrazena detekce úpravy umělou inteligencí v daném snímku.



Obrázek 39 Odhalení stopy po úpravě (scanner.deepware.ai)

Program je zatím ve fázi beta testování, avšak při spuštění samotného videa v místech, kde se objevuje a zpívá daná osoba, není synchronizovaný pohyb úst se zvukem. Je tedy pravděpodobné, že video bylo upravováno ještě dříve než úpravou autora této práce. Upravené video bylo následně nahráno a aplikací vyhodnoceno. Na Obrázku 40 je srovnání před a po v pozměněné oblasti včetně výsledků jednotlivých modelů využívaných programem. Oba snímky jsou zachyceny ve stejném časovém úseku.



Obrázek 40 Upravované video před a po úpravě (scanner.deepware.ai)

Jak jde vidět v upravovaném místě, program nezaregistroval žádnou úpravu. Na Obrázku 41 je vidět vyhodnocení jednotlivých modelů.

Model Results	Model Results
<u>Avatarify</u> : NO DEEPFAKE DETECTED(23%)	<u>Avatarify</u> : NO DEEPFAKE DETECTED(1%)
<u>Deepware</u> : NO DEEPFAKE DETECTED(0%)	<u>Deepware</u> : DEEPFAKE DETECTED(91%)
<u>Seferbekov</u> : DEEPFAKE DETECTED(99%)	<u>Seferbekov</u> : DEEPFAKE DETECTED(98%)
<u>Ensemble</u> : DEEPFAKE DETECTED(96%)	<u>Ensemble</u> : DEEPFAKE DETECTED(95%)

Obrázek 41 Výsledky hodnocení modelů před a po úpravě videa (scanner.deepware.ai)

Jak je zobrazeno nástroj určité stopy Deepfaku odhalil. V určité vteřině program detekoval úpravu umělou inteligencí. Tento fakt znázorňuje Obrázek 42.



Obrázek 42 Detekce úpravy AI (scanner.deepware.ai)

#### 6.4 Závěr analýzy náročnosti tvorby DeepFaků

Předmětem této části bylo poukázat na náročnost tvorby falešných médií. Za tímto účelem byla vytvořena tabulka napomáhající pochopit, jak náročná tvorba může být. Ukázalo se, že generování Deepfakes je jednoduché, a i snadno dostupné pro kohokoliv. Z důvodu jednoduchosti a dostupnosti se různé společnosti pokouší vytvořit nástroje pro detekci. Část kapitoly zabývající se právě některým nástroji se zaměřila na jejich efektivnost. U Deepfake detect se ukázaly značné nedostatky, neboť odhalení proběhlo neúspěšně. Aplikace je stále ve vývoji a lze očekávat určitá zlepšení do budoucnosti. Naopak deepware se ukázal jako nástroj schopný určité stopy úpravy umělou inteligencí odhalit. Pro úspěšnou implementaci bude potřeba ještě nějaký čas vývoje a zdokonalení.

## 7 OPATŘENÍ PRO MINIMALIZACI RIZIK

Na základě scénářů popsanych výše v práci jsou zde navrhnutá opatření. Ty mají za úkol snížit riziko na přijatelnou úroveň. V rámci práce tak jsou řešena opatření vůči Ransomwaru, prolomení hesla a DeepFakům. U DeepFaků bylo již jedno opatření zmíněna v předchozí kapitole v podobě softwaru pro detekci stop AI u médií.

### 7.1 Antimalware

Opatření zmírňující vliv rizika Ransomware, ale i jiných, je doporučeno použití kvalitní antimalware ochrany. Ta často bývá opomíjena a jsou u obecních zařízení používány méně kvalitní či základní antimalwary. Základním je míněn Windows Defender, který však má několik kritických nedostatků. Jeden z nich je pozdní aktualizace či náročnost na hardware. Samotná náročnost na hardware je u obecních počítačů problematická. Často se jedná o ne zrovna moderní kousky a takovýto program je značně zatěžuje. Z tohoto důvodu bylo navrženo použití jiného programu. Tím se stal Bitdefender. Na Obrázku 43 je ilustrováno hodnocení od renomovaného magazínu.

●●●●○ 4.0 Excellent

#### THE BOTTOM LINE

Bitdefender Antivirus Free for Windows offers the same core technology found in the commercial version, focusing on the essential task of antivirus protection with minimal add-ons.

MSRP Free

\$0.00 at Bitdefender

See It

Obrázek 43 Hodnocení Bitdefenderu  
(Bitdefender Antivirus Free for  
Windows Review, © 1996-2024)

Na dalším Obrázku 44 je hodnocení na základě testu nezávislé instituce testující antimalwary AV-TEST, jenž ukazuje dobré výsledky Bitdefenderu.

## Windows 10: December 2023

Producer	Certified	Protection	Performance	Usability
 <b>Avast</b> Free Antivirus 23.9 & 23.11				
 <b>AVG</b> Internet Security 23.10 & 23.11				
 <b>Avira</b> Internet Security for Windows 1.1				
 <b>Bitdefender</b> Internet Security 27.0				

Obrázek 44 Hodnocení Bitdefender od AV-TEST (The best Windows antivirus software for home users, ©2024)

Skrze tyto pozitivní ohlasy se stal antimalware vhodným doporučením pro realizaci. Výhodou je, že se jedná o freeware. Je tedy plně zdarma, a i přes to chrání dostatečně dobře. V podstatě je verze zdarma a prémiová verze v základní ochraně totožná. Samotný program je možné stáhnout na odkazu: <https://www.bitdefender.com/solutions/free.html>.

Za zmínku funkce, která je v prémiové verzi je Zotavení po útoku Ransomware. Ta umožňuje obnovení souboru po útoku tímto škodlivým programem. Další takovou funkcí je i ochrana vůči sledování či odposlouchávání, jež by mohla zamezit získání dat útočníkem pro tvorbu Deepfake. Pro získání těchto je však nutný poplatek ve výši 1,429.35 korun českých ročně až pro pět zařízení. Otázkou je, zda by se to pro obec vyplatilo. Program svoji úlohu zastává dostatečně i ve verzi zdarma. Bitdefender poskytuje i webovou ochranu včetně ochrany vůči phishingovým stránkám.

## 7.2 Zabezpečení hesla

Pro minimalizaci rizika prolomení hesla pomocí AI nástrojů jsou navrženy doporučená opatření použití správce hesel, implementace více faktorového přihlášení a zavedení pravidelné obměny hesel.

Správce hesel umožňuje jednodušší zapamatování hesla. Je totiž potřeba si zapamatovat pouze heslo od aplikace uchovávající hesla v databázi. Navíc díky lokální přístupnosti je nutné použít fyzicky konkrétní zařízení. Pro potřeby práce byla zvolena aplikace KeePassXC. Důvodem volby byly skutečnosti, že se jedná o bezplatný program a hesla

ukládá pouze lokálně. Neodesílá data na cloud, čímž je sníženo riziko úniku. Tento fakt navíc snižuje šanci získání hesel útočníkem, jelikož zhoršuje jejich přístup pro něj. Výhodou, ale i zároveň nevýhodou je, že se jedná o open source program. Kód této aplikace je tedy volně dostupný a každý se na něj může podívat. Výhodou je umožnění komunitě hledat chyby a následně pomoci k nápravě. Avšak tuto chybu může najít nebo zneužít i osoba s nekalými úmysly. Jde tak o svým způsobem o dvousečnou zbraň. I tak ale se nejedná o zásadní problém, jenž by zabráňoval použití KeePassXC jako nápravné opatření pro obec. Samotný program lze nalézt na stránkách: <https://keepassxc.org/>. Aplikace umožňuje generovat silná hesla a vygenerovat soubor s klíčem. Tento soubor slouží jako klíč, jehož přítomnost je nutná k přihlášení. Lze jej umístit na externí disk a ten připojit v případě potřeby. Nevýhodou pro uživatele je nutnost mít tento disk při přihlašování u sebe. Pozitivním efektem je další zhoršení dostupnosti pro hackera. Navíc je snižena šance zašifrování Ransomwarem.

Dalším opatřením v rámci této kategorie je více faktorové zabezpečení. Jedná se o přihlášení pomocí dvou různých metod. První metodou je klasické přihlašovací heslo. Druhá metoda je další úroveň zabezpečení a nejčastěji se jedná o potvrzení přihlášení skrze jiné zařízení. Častokrát se jedná buď o zadání kódu v SMS zprávě nebo v aplikaci nainstalované na mobilu. V případě, že by někdo prolomil heslo musel by zároveň mít i přístup k tomuto zařízení. Díky tomu je tak pro útočníka náročnější dostat se přes celkové zabezpečení. V případě domény Seznam je toto opatření možné zapnout a funguje způsobem potvrzení přihlášení na druhém mobilním zařízení. Další možností je využití aplikací jako je MS Authenticator na mobilní telefon. V aplikaci jsou pravidelně generovány v 30sekundových intervalech číselné kódy. Při přihlášení do účtu se tak kromě hesla zadá i tento generovaný kód. Jelikož ověřování přístupu k účtu je na dvou rozdílných zařízeních, je bezpečnost zajištěna.

Posledním návrhem opatření je i ustanovení obměny hesel. Pro zvýšení bezpečnosti je doporučena pravidelná obměna hesel, čímž se snižuje šance na odhalení hesla. V případě obecního úřadu je tak navrhována obměna hesla jednou za devadesát dní. Je však nutné, aby dané heslo bylo oproti původnímu rozdílné. Je potřeba v tomto případě i řádně zaměstnance proškolení. S tímto opatřením může pomoci i již zmíněný KeePassXC. V kombinaci s předchozími opatřeními by se tak riziko vzniku minimalizovalo na přípustnou úroveň.

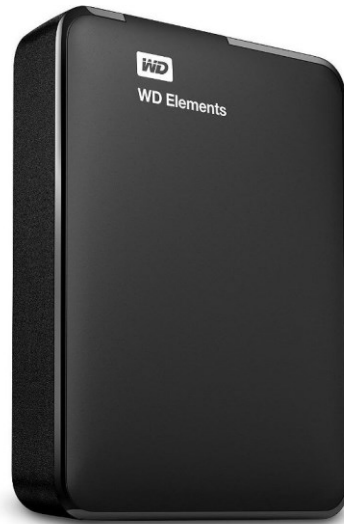
### 7.3 Školení o problematice

Opatření školení o problematice má za úkol seznámit zaměstnance s možnými riziky umělé inteligence a kyberprostoru. Právě nedostatečné znalosti jsou často důvodem, proč k incidentu došlo. Problematika není rozsáhlá a školení by mohlo být stylem prezentace. S tou dokáže pomoci opět umělá inteligence. Jedna taková prezentace byla vygenerována v rámci práce a je obsažena v příloze. Pro generování byl zvolena aplikace beautiful.ai. V případě Deepfakes a občanů by mohlo dojít k realizaci opatření pomocí přednášky v rámci aktivit obecní knihovny. Zde by se občané dozvěděli o zásadních problémech dezinformací a trendech v této oblasti.

### 7.4 Záloha dat

Zálohování dat je opatření zmírňující nebezpečí ztráty dat v počítačovém zařízení. Ztráta dat v podobě obecních dokumentů jako smlouvy o pronájmu, pachtovní smlouvy, účetní uzávěrky a jiné dokumenty či plány související s chodem obce či jejími budoucími plány může být pro subjekt kritické. Některé dokumenty jsou sice ve fyzické podobě, ale i ty by bylo nutné znovu vytvořit. To by zabralo spousty času. Tento scénář lze minimalizovat pomocí zálohy buď na cloudy či externím diskem. V závislosti na práci s danými soubory by šlo rozdělit zálohu dat na zálohu pomocí cloudu anebo externího disku. Ve formě cloudu by se soubory ukládaly pravidelně nejlépe po ukončení práce s nimi. Jelikož ukládání složek na cloud je velice jednoduché a rychle lze toto pravidlo jednoduše nastavit. Případně by se ustanovila doba, kdy proběhne záloha těchto souborů v určitou hodinu naráz. Cloudová záloha by tak proběhla formou cloudového úložiště OneDrive obsaženého v kancelářského balíčku Office 365. Lze tak zde jednoduše vytvořit složky a pravidelně po ukončení práce se souborem jej zde uložit. Pro případ možného výpadku či ztráty cloudového úložiště je navrženo použití externího disku. Na něm by probíhala záloha poslední pátek v měsíci mimo úřední hodiny. Pro takové využití byl na základě zákaznických recenzí navrhnout externí disk WD Elements Portable 1TB černý s cenou 1649 korun českých. Znárodněn je na Obrázku 45.





Obrázek 45 Externí disk  
(Alza.cz, © 1994–2024)

Disků by ideálně bylo využito více, pro každý počítač samostatně. Předěšlo by se tak případnému rozšíření malwaru do více zařízení.

## 7.5 Závěr k návrhům opatření

V rámci problematiky byly navrženy opatření pro minimalizaci definovaných hrozeb. Opatření vycházejí převážně z popsanych scénářů, avšak vztahují se i k jiným. Častokrát tyto rizika mohou být provázané. Například Deepfake může být použit pro přesvědčení oběti, aby stáhla infikovaný soubor. Je možno v tomto případě tvrdit, že i opatření řeší více hrozeb. Při hledání řešení byl brán ohled i na cenovou náročnost. V případě dobrého antimalwaru existují i kvalitní freeware verze. Ty i bez prémiových funkcí jsou dostatečným řešením. Na základě představených opatření lze tedy předpokládat úspěšné zmírnění pravděpodobných hrozeb na přijatelnou úroveň.

## ZÁVĚR

Předmětem této práce byla identifikace relevantních rizik umělé inteligence ve vztahu k umělé inteligenci a navržení opatření pro jejich minimalizaci. Z charakteru práce a popsání výsledků je možné odvodit, že stanovené parametry byly do uspokojivé míry splněny.

V teoretické části byla rozebrána problematika na základě literární rešerše. Jednou z jejích náplní bylo popsání úkolů orgánu ochrany obyvatelstva pro lepší pochopení spojitosti s tématem práce. Další část zabývající se pojmem a podstatou interpretuje různé definice popisující co je vlastně umělá inteligence. Zároveň pro lepší porozumění problematice je popsána i samotná podstata pojmu. Byly zde popsány různé odborné pohledy na umělou inteligenci včetně otázky, zda ji lze vůbec označit za inteligentní a otázka etiky. Druhá část definovala rizika vycházející z odborných článků či zdrojů. Byly popsány rizika nadměrné automatizace, autonomních zbraní, rizika v kyberprostoru a existenční rizika. Tyto rizika na základě rešerše představují skutečnou hrozbu ať už v blízké budoucnosti anebo v průběhu několika desetiletí. Jsou tak popsány důvody proč se jedná o problém a případně skutečnosti poukazující, že by to nemuselo být až tak problematické.

V praktické části práce byly nejdříve identifikovány relevantní rizika působící na objekt typu obecní úřad. Obecní úřad se zdál být vhodným aktérem, neboť omezení či zamezení jeho fungování způsobí zamezení provádění úkolů směřující k zajištění bezpečnosti obyvatelstva. Dále byla provedena FMEA analýza, jak vlastní, tak pomocí modelů umělé inteligence s cílem zjistit a poukázat za možné využití i v pozitivním světle. S tímto účelem byl vytvořen i personifikovanější vlastní chatbot. Bylo zjištěno, že výstupy z daných aplikací byl spíše všeobecného charakteru. Je tedy nutná úprava na základě vlastní iniciativy, tak aby seděly dané skutečnosti. Nicméně lze je pro účely analýz využít jako generátor nápadů či informací, které by nemusely analytika v daný moment napadnout. Dále byly popsány scénáře na základě vyhodnocení analýz. Scénáře tedy přibližují možnou rizikovou v rámci subjektu. Dalším předmětem práce byla analýza náročnosti tvorby Deepfake médií. Kapitola představila tvorbu těchto falešných médií a tabulkou přiblížila jejich náročnost. Tato část obsahovala i otestování dostupných aplikací pro detekci Deepfake. Ukázalo se, že programy nejsou ještě spolehlivé. V závěru práce byly navrženy opatření minimalizující dopady zkoumaných rizik. Bylo tedy navrženo instalování antimalwaru a správce hesel. Dále byly navrženy pravidelná obměna hesla, více faktorové ověřování, implementace procesu zálohy dat a opatření zlepšující povědomí o problematice v podobě informačních prospektů.

**SEZNAM POUŽITÉ LITERATURY**

ASIMOV, Isaac, 1993. *Roboti a Impérium*. Poutník (Klub Julese Vernea). Praha: Klub Julese Vernea. ISBN 80-901-2809-2.

ALAKEEL, Fatimah; ALFALLAJ, Rehab; ALSHEHRI, Hanan a ALMOUSA, Ali, 2023. *AI-based Cybersecurity Attacks and Countermeasures in IoT Environment: A Survey*. Online. Journal of Engineering and Applied Sciences. Roč. 10, č. 2. ISSN 1658-6638. Dostupné z: <https://doi.org/10.5455/jeas.2023110105>. [cit. 2024-01-15].

ALZA.CZ, © 1994–2024. *WD Elements Portable ITB černý*. In: Alza.cz [online]. [cit. 2024-02-27]. Dostupné z: <https://www.alza.cz/wd-2-5-elements-portable-1tb-cerny-d4945001.htm>

*Bitdefender Antivirus Free for Windows Review*, © 1996-2024. Online. ZIFF DAVIS. PCMag. Dostupné z: <https://www.pcmag.com/reviews/bitdefender-antivirus-free>. [cit. 2024-02-26].

BOSTROM, Nick, 2016. *Superintelligence: paths, dangers, strategies*. Obzor (Prostor). Oxford: Oxford University Press. ISBN 978-019-8739-838.

BUNDESAMT FÜR SICHERHEIT IN DER INFORMATIONSTECHNIK, © 2024. *Generative AI Models*. Online. Bundesamt für Sicherheit in der Informationstechnik. Dostupné z: [https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative\\_AI\\_Models.html](https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.html). [cit. 2024-04-18].

CASEY, Cindy, © 2024. *AI Poisoning Attacks*. Online. In: BUCKS COUNTY COMMUNITY COLLEGE. Bucks.edu. Dostupné z: <https://www.bucks.edu/media/bcccmcdialibrary/con-ed/itacademy/fos2022/BCCC---Cyber-101.pdf>. [cit. 2024-01-09].

ČESKO, 2000. Zákon č. 239 ze dne 28. června 2000, o integrovaném záchranném systému a o změně některých zákonů. Online. In: *Zákony pro lidi*. 2010-2024. Dostupné z: <https://www.zakonyprolidi.cz/cs/2000-239>. [cit. 2024-04-18].

COMITER, Marcus, 2019. *Attacking Artificial Intelligence AI's Security Vulnerability and What Policymakers Can Do About It*. Online. Cambridge: the Belfer Center for Science and International Affairs. Dostupné z: <https://www.belfercenter.org/publication/AttackingAI>. [cit. 2024-01-02].

CABLE NEWS NETWORK, © 2024. Finance worker pays out \$25 million after video call with Deepfake 'chief financial officer'. Online. CNN World. Dostupné z: <https://edition.cnn.com/2024/02/04/asia/Deepfake-cfo-scam-hong-kong-intl-hnk/index.html>. [cit. 2024-04-18].

DE SPIEGELEIRE, Stephan; MAAS, Matthijs a SWEIJS, Tim, 2017. *Artificial Intelligence and the Future of Defense*. Online. Haag: Hague Centre for Strategic Studies. ISBN 978-94-92102-54-6. Dostupné z: <https://www.jstor.org/stable/resrep12564>. [cit. 2023-11-10].

EVROPSKÁ UNIE, ©2018. *Sdělení komise Evropskému parlamentu, Radě, Evropskému hospodářskému a sociálnímu výboru a Výboru regionů*. Online. EUR-Lex. Dostupné z: <https://eur-lex.europa.eu/legal-content/CS/ALL/?uri=COM:2018:237:FIN>. [cit. 2023-11-10].

EVROPSKÝ PARLAMENT, © 2021. *Co je umělá inteligence a jak ji využíváme?* Online. Zpravodajství. Dostupné z: <https://www.europarl.europa.eu/news/cs/headlines/society/20200827STO85804/umela-intelligence-definice-a-vyuziti>. [cit. 2023-11-07].

FORBES MEDIA LLC, © 2024. *A Voice Deepfake Was Used To Scam A CEO Out Of \$243,000*. Online. FORBES MEDIA LLC. Forbes. Dostupné z: <https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-Deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=70ba8efd2241>. [cit. 2024-01-15].

GRACE, Katja; SALVATIER, John; DAFOE, Allan; ZHANG, Baobao a EVANS, Owain, 2018. *Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts*. Online. Journal of Artificial Intelligence Research. Roč. 2018, č. 62. Dostupné z: <https://doi.org/10.1613/jair.1.11222>. [cit. 2023-12-28].

GUEMBE, Blessing; AZETA, Ambrose; MISRA, Sanjay; OSAMOR, Victor Chukwudi; FERNANDEZ-SANZ, Luis et al., 2022. *The Emerging Threat of Ai-driven Cyber Attacks: A Review*. Online. Applied Artificial Intelligence. 2022-12-31, roč. 36, č. 1. ISSN 0883-9514. Dostupné z: <https://doi.org/10.1080/08839514.2022.2037254>. [cit. 2024-01-10].

HELM, Matthew; SWIERGOSZ, Andrew M.; HAEBERLE, Heather; KARNUTA, Jaret M.; SCHAFFER, Jonathan L. et al., 2020. *Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions*. Online. Current Reviews in Musculoskeletal Medicine. Roč. 13, č. 1. ISSN 1935-9748. Dostupné z: <https://doi.org/10.1007/s12178-020-09600-8>. [cit. 2023-11-29].

- HITAJ, Briland; GASTI, Paolo; ATENIESE, Giuseppe a PEREZ-CRUZ, Fernando, 2019. *PassGAN: A Deep Learning Approach for Password Guessing*. Online. Dostupné z: <https://doi.org/https://doi.org/10.48550/arXiv.1709.00440>. [cit. 2024-01-11].
- KALOUDI, Nektaria a LI, Jingyue, 2021. *The AI-Based Cyber Threat Landscape*. Online. ACM Computing Surveys. 2021-01-31, roč. 53, č. 1. ISSN 0360-0300. Dostupné z: <https://doi.org/10.1145/3372823>. [cit. 2024-01-11].
- KELLEHER, John D., 2019. *Deep Learning*. Online. The MIT Press. ISBN 9780262354899. Dostupné z: <https://doi.org/https://doi.org.proxy.k.utb.cz/10.7551/mitpress/11171.001.0001>. [cit. 2023-11-29].
- KIETZMANN, Jan; LEE, Linda W.; MCCARTHY, Ian P. a KIETZMANN, Tim C., 2020. *Deepfakes: Trick or Treat?* Online. Business Horizons. Roč. 63, č. 2. ISSN 0007-6813. Dostupné z: <https://doi.org/https://doi.org/10.1016/j.bushor.2019.11.006>. [cit. 2024-01-14].
- KUZLU, Murat; FAIR, Corinne a GULER, Ozgur, 2021. *Role of Artificial Intelligence in the Internet of Things (IoT) cybersecurity*. Online. Discover Internet of Things. ISSN 2730-7239. Dostupné z: <https://doi.org/https://doi.org/10.1007/s43926-020-00001-4>. [cit. 2024-01-09].
- MAŘÍK, Vladimír; ŠTĚPÁNKOVÁ, Olga a LAŽANSKÝ, Jiří, 2013. *Umělá inteligence 6*. Praha: Academia. ISBN 978-802-0022-769.
- MĚSTO TÁBOR, © 2024. *Úkoly obcí z hlediska ochrany obyvatelstva*. Online. MĚSTO TÁBOR. Tábor. Dostupné z: <https://www.taborcz.eu/ukoly-obci-z-hlediska-ochrany-obyvatelestva/ms-8317>. [cit. 2024-04-18].
- NÚKIB, © 2024. *Kybernetické incidenty pohledem NÚKIB*. Online. Dostupné z: <https://nukib.gov.cz/download/publikace/vyzkum/kyberneticke-incidenty-pohledem-NUKIB-leden-2024.pdf>. [cit. 2024-02-22].
- REESE, Byron, 2022. *Čtvrtý věk: inteligentní roboti, myslící počítače a budoucnost lidstva*. Technologie (Zoner Press). Brno: Zoner Press. ISBN 978-80-7413-490-6.
- RUSSELL, Stuart a NORVIG, Peter, 2020. *Artificial Intelligence: A Modern Approach*. 4. Hoboken: Pearson. ISBN 978-0134610993.
- RUSSELL, Stuart, 2022. *Banning Lethal Autonomous Weapons: An Education*. Online. Issues in Science and Technology. Roč. 38, č. 3. ISSN 07485492. Dostupné z: <https://issues.org/banning-lethal-autonomous-weapons-stuart-russell/>. [cit. 2023-12-30].

*The best Windows antivirus software for home users*, © 2024. Online. AV-TEST. AV-TEST. Dostupné z: <https://www.av-test.org/en/antivirus/home-windows/windows-10/december-2023/>. [cit. 2024-02-26].

*Umělá inteligence vs. Homo sapiens adaptabilis*, ©2023. Online. Medium. Dostupné z: <https://medium.com/@GUG.cz/um%C4%9B1%C3%A1-intelligence-vs-homo-sapiens-adaptabilis-6daf570ffab3>. [cit. 2023-11-30].

USA GOVERNMENT, ©2020. *H. R. 6216*. Online. Congress.gov. Dostupné z: <https://www.congress.gov/bill/116th-congress/house-bill/6216/text?s=6&r=1&q=%7B%22search%22%3A%22NATIONAL+ARTIFICIAL+INTELLIGENCE+INITIATIVE+ACT+OF+2020%22%7D>. [cit. 2023-11-10].

THE CONVERSATION MEDIA GROUP LTD, ©2010–2023. *Understanding the four types of AI, from reactive robots to self-aware beings*. Online. THE CONVERSATION MEDIA GROUP LTD. The Conversation. Dostupné z: <https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>. [cit. 2023-11-11].

*The Future of Jobs Report 2023*, ©2023. Online. Cologne: World Economic Forum. ISBN 978-2-940631-96-4. Dostupné z: <https://www.weforum.org/publications/the-future-of-jobs-report-2023/>. [cit. 2023-12-28].

THE HACKER NEWS, © 2023. *WormGPT: New AI Tool Allows Cybercriminals to Launch Sophisticated Cyber Attacks*. Online. THE HACKER NEWS. The Hacker News. Dostupné z: <https://thehackernews.com/2023/07/wormgpt-new-ai-tool-allows.html>. [cit. 2024-01-27].

VERIZON COMMUNICATIONS, © 2023. *2023 Data Breach Investigations Report*. Online. VERIZON COMMUNICATIONS. Verizon. Dostupné z: <https://www.verizon.com/business/resources/T4f2/reports/2023-data-breach-investigations-report-dbir.pdf>. [cit. 2024-01-15].

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ, © 2024. FMEA protokol. Online. VUT. Dostupné z: [https://www.vut.cz/www\\_base/zav\\_prace\\_soubor\\_verejne.php?file\\_id=17051](https://www.vut.cz/www_base/zav_prace_soubor_verejne.php?file_id=17051). [cit. 2024-04-18].

ZANDL, Patrick, 2022. *Mýty a naděje digitálního světa: vše, co potřebujete vědět o kryptoměnach, umělé inteligenci a dalších převratných technologiích*. Pod povrchem. V Brně: Jan Melvil Publishing. ISBN 978-80-7555-175-7.

ZERILLI, John; DANAHER, John; MACLAURIN, James; GAVAGHAN, COLIN, Colin a KNOTT, Alistair, 2020. A citizen's guide to artificial intelligence. Online. Cambridge: The MIT Press. ISBN 9780262361323. Dostupné z: <https://doi.org/https://doi-org.proxy.k.utb.cz/10.7551/mitpress/12518.001.0001>. [cit. 2023-11-04].

WIKIPEDIE, © 2024. *Petr Pavel*. In: Wikipedie [online]. [cit. 2024-03-17]. Dostupné z: [https://cs.wikipedia.org/wiki/Petr\\_Pavel](https://cs.wikipedia.org/wiki/Petr_Pavel)

WIKIPEDIE, © 2024. *Jimmy Carter*. In: Wikipedie [online]. [cit. 2024-03-17]. Dostupné z: [https://cs.wikipedia.org/wiki/Jimmy\\_Carter](https://cs.wikipedia.org/wiki/Jimmy_Carter)

WILKS, Yorick A., 2019. *Artificial intelligence: modern magic or dangerous future?* Hot science (Icon Books). London: Icon Books. ISBN 978-1785785160.

## SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

AI Artificial Intelligence

DDoS Distributed Denial of Service

DoS Denial of Service

FMEA Failure Mode and Effects Analysis

IZS Integrovaný záchranný systém

ORP Obec s rozšířenou působností

PC Personal Computer

RPN Risk Priority Number



**SEZNAM OBRÁZKŮ**

Obrázek 1 Vizualizace vztahu mezi jednotlivými pojmy (Umělá inteligence vs. Homo sapiens adaptabilis, © 2023).....	19
Obrázek 2 Proces strojového učení (SAP, © 2023).....	20
Obrázek 3 Jednoduché schéma umělé neuronové sítě (Kelleher, 2019).....	21
Obrázek 4 Očekávaný dopad zavedení technologií na pracovní místa (The Future of Jobs Report 2023, © 2023).....	25
Obrázek 5 Útok na AI u autonomního vozu (Casey, © 2024).....	28
Obrázek 6 Nejčastější druhy akcí při narušení dat (Verizon Communications, © 2023) ...	29
Obrázek 7 Nejčastější iniciační akce u datových úniků a kybernetických incidentů (Verizon Communications, © 2023).....	30
Obrázek 8 Počet kybernetických útoků dle NÚKIB (NÚKIB, ©2024).....	31
Obrázek 9 Klasifikace kybernetických incidentů (NÚKIB, ©2024) .....	31
Obrázek 10 Vlastní FMEA analýza 1. část (Vysoké učení technické v Brně, © 2024).....	42
Obrázek 11 Vlastní FMEA analýza 2. část (Vysoké učení technické v Brně, © 2024).....	42
Obrázek 12 FMEA analýza od Gemini (Vygenerováno umělou inteligencí) .....	46
Obrázek 13 Finální FMEA od Gemini 1. část (Vygenerováno umělou inteligencí) .....	47
Obrázek 14 Finální FMEA od Gemini 2. část (Vygenerováno umělou inteligencí) .....	47
Obrázek 15 FMEA analýza od ChatGPT 3.5 free 1. část (Vygenerováno umělou inteligencí) .....	49
Obrázek 16 FMEA analýza od ChatGPT 3.5 free 2. část (Vygenerováno umělou inteligencí) .....	49
Obrázek 17 FMEA analýza od ChatGPT 3.5 free 3. část (Vygenerováno umělou inteligencí) .....	50
Obrázek 18 Opatření vůči rizikům dle ChatGPT 1. část (Vygenerováno umělou inteligencí) .....	50
Obrázek 19 Opatření vůči rizikům dle ChatGPT 2. část (Vygenerováno umělou inteligencí) .....	51
Obrázek 20 Odpověď ChatGPT 3.5 na otázku proč dané riziko stanovil jako nejvyšší (Vygenerováno umělou inteligencí).....	52
Obrázek 21 Omluva ChatGPT za chybu v analýze (Vygenerováno umělou inteligencí)...	52
Obrázek 22 Vysvětlení ChatGPT 3.5 proč udělal chybu (Vygenerováno umělou inteligencí) .....	53
Obrázek 23 FMEA od ChatGPT 4 (Vygenerováno umělou inteligencí) .....	53
Obrázek 24 Počáteční fáze tvorby GPT (chat.openai.com/gpts/editor) .....	54
Obrázek 25 Nastavení a úprava GPT (chat.openai.com) .....	55
Obrázek 26 Finální podoba CyberGPT (chat.openai.com) .....	57
Obrázek 27 FMEA vytvořená CyberGPT (Vygenerováno umělou inteligencí) .....	57

Obrázek 28 Klasifikační tabulka vytvořená CyberGPT (Vygenerováno umělou inteligencí) .....	58
Obrázek 29 Swimlane diagram incidentu v rámci scénáře č.1 (Vlastní zpracování).....	63
Obrázek 30 Swimlane diagram incidentu v rámci scénáře č.2 (Vlastní zpracování).....	65
Obrázek 31 Swimlane diagram incidentu v rámci scénáře č.3 (Vlastní zpracování).....	67
Obrázek 32 Originální obrázky (Wikipedie, © 2024).....	69
Obrázek 33 Deepfake pomocí unboring (reface.ai/unboring/face-swap).....	70
Obrázek 34 Deepfake pomocí Vidnoz (vidnoz.com/face-swap.html) .....	70
Obrázek 35 Porovnání Deepfake fotek (Vygenerováno umělou inteligencí).....	71
Obrázek 36 Fotka použitá ke generování Deepfake videa (reface.ai).....	71
Obrázek 37 Srovnání originálu a Deepfake videa (reface.ai).....	72
Obrázek 38 Výsledek testu Deepfake Detectorem (Deepfake-detect.com) .....	74
Obrázek 39 Odhalení stopy po úpravě (scanner.deepware.ai).....	75
Obrázek 40 Upravované video před a po úpravě (scanner.deepware.ai) .....	75
Obrázek 41 Výsledky hodnocení modelů před a po úpravě videa (scanner.deepware.ai) ..	76
Obrázek 42 Detekce úpravy AI (scanner.deepware.ai) .....	76
Obrázek 43 Hodnocení Bitdefenderu (Bitdefender Antivirus Free for Windows Review, © 1996-2024).....	77
Obrázek 44 Hodnocení Bitdefender od AV-TEST (The best Windows antivirus software for home users, ©2024) .....	78
Obrázek 45 Externí disk (Alza.cz, © 1994–2024).....	81

**SEZNAM TABULEK**

Tabulka 1 Relevance rizik (Vlastní zpracování).....	37
Tabulka 2 Význam následků kybernetického incidentu (Vysoké učení technické v Brně, © 2024).....	40
Tabulka 3 Pravděpodobnost výskytu kybernetického incidentu (Vysoké učení technické v Brně, © 2024) .....	40
Tabulka 4 Pravděpodobnost odhalení (Vysoké učení technické v Brně, © 2024) .....	41
Tabulka 5 Klasifikace rizika RPN (Vysoké učení technické v Brně, © 2024).....	41
Tabulka 6 Klasifikace RPN od Gemini (Vygenerováno umělou inteligencí) .....	48
Tabulka 7 Celkové RPN od jednotlivých chatbotů (Vlastní zpracování) .....	60
Tabulka 8 Hodnocení náročnosti tvorby Deepfake (Vlastní zpracování) .....	72

## SEZNAM PŘÍLOH

Příloha P I: Prezentace vygenerovaná umělou inteligencí

# PŘÍLOHA P I: PREZENTACE VYGENEROVANÁ UMĚLOU INTELIGENCÍ



## Kybernetické útoky a umělá inteligence

Úvodní snímek o kybernetických útocích a umělé inteligenci

### Co jsou kybernetické útoky?



#### Phishing

Phishing útoky využívají falešné e-maily nebo webové stránky, aby nalákaly uživatele ke sdílení citlivých údajů. Klíčové slovo: e-mail



#### Malware

Malware je škodlivý software jako viry, červi nebo trojské koně, který může poškodit systémy nebo odcizit data. Klíčové slovo: virus



#### Sociální inženýrství

Sociální inženýrství využívá manipulace k získání přístupu nebo citlivých údajů. Klíčové slovo: útočník

Kybernetické útoky mají mnoho forem od phishingu po malware. Je důležité chránit se proti těmto hrozbám.

# Jak umělá inteligence umožňuje kybernetické útoky

## Falešné přihlašovací údaje

Útočníci využívají AI k vytváření velkého množství falešných přihlašovacích údajů pro překonání zabezpečení.

## Sociální inženýrství

AI dokáže napodobit lidskou komunikaci a získávat citlivé informace přes phishing.

## Automatizované útoky

AI umožňuje vytvářet malware schopný se sám šířit a vyhledávat zranitelnosti.

## Vyhýbání se detekci

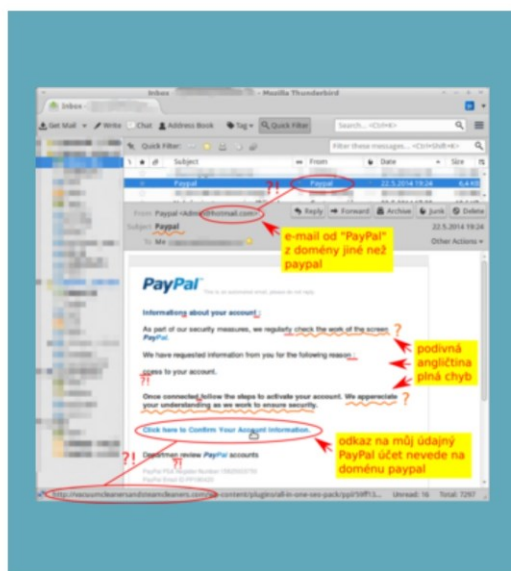
Útočníci využívají AI k vytváření nových variant malware, které nejsou detekovatelné.

## Zrychlení útoků

AI umožňuje útočnickům rychleji najít cíle a provést útok.

## Lepší zacílení

AI pomáhá útočnickům lépe pochopit chování uživatelů a firem pro lepší zacílení útoků.



## Phishingové útoky

Phishingové útoky se stávají stále sofistikovanějšími, jak útočníci využívají umělou inteligenci k vytváření přesvědčivějších falešných e-mailů a webů. Mohou automaticky generovat obrovské množství personalizovaného spamu a falešných stránek. Je důležité být mimořádně opatrný, pokud jde o klikání na odkazy nebo zadávání osobních údajů.

## Sociální inženýrství

### Zneužití chatbotů

Kyberzločinci používají chatboty k oklamání lidí a získání citlivých informací, jako jsou hesla nebo údaje o kreditní kartě.

### Falešné identity

Chatboti se mohou vydávat za skutečné osoby a budovat si důvěru obětí, aby získali přístup k citlivým informacím.

### Cílené útoky

Chatboti mohou být naprogramováni k tomu, aby se zaměřili na konkrétní oběti a získali od nich konkrétní informace.

### Sběr osobních údajů

Chatboti mohou shromažďovat velké množství osobních údajů od uživatelů, aniž by si toho uživatelé byli vědomí.

### Šíření dezinformací

Chatboti mohou šířit falešné zprávy, propagandu a dezinformace, aby ovlivnili názory a chování lidí.

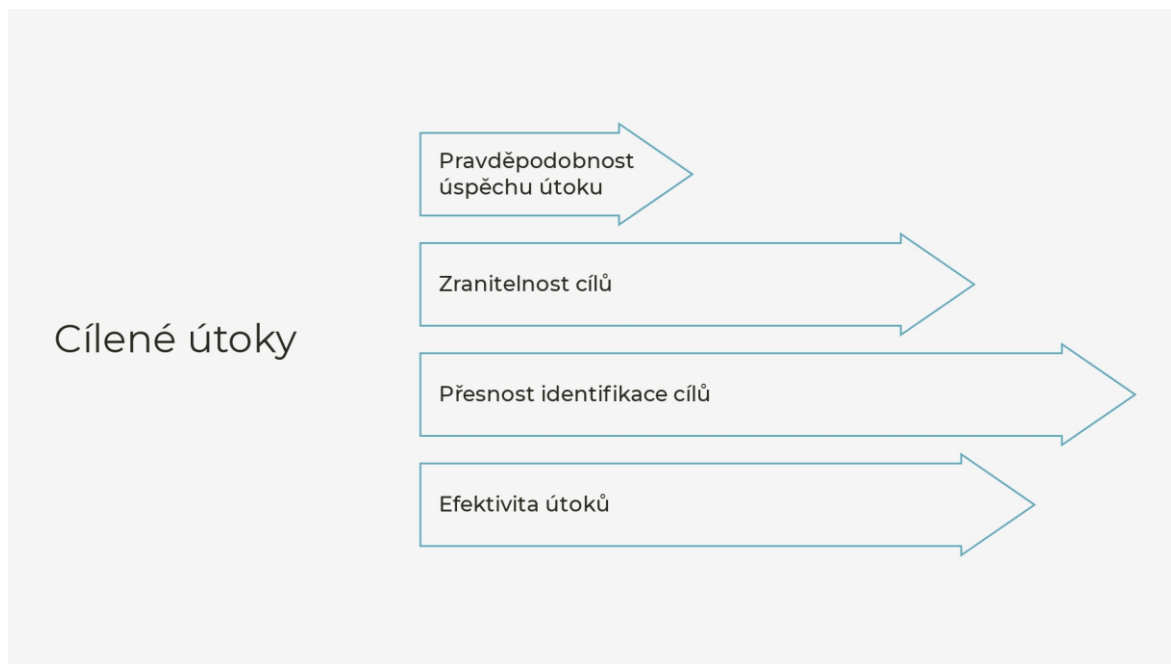
## Vyhýbání se detekci

Schopnost AI vytvářet nové varianty malwaru

Schopnost AI vytvářet polymorfní malware

Obcházení bezpečnostních řešení

Neustálý vývoj nových technik



## DeepFake



### DeepFakes jsou videa/obrázky upravené pomocí AI

Používají generativní AI k vytvoření falešných video/audio nahrávek



### Využívají pokročilých technik strojového učení

Nejčastěji generativní sítě jako GAN a autoenkodéry



### Mohou být zneužity pro manipulaci

K šíření dezinformací, propagandy nebo kompromitace

DeepFakes představují novou výzvu v oblasti kybernetické bezpečnosti a boje proti dezinformacím.



# Způsoby jak odhalit Deepfake



## Hledat abnormální vzorce chování

Deepfake často neblíkají či nemrkají přirozeně



## Analyzovat metadata

Originální zdrojová videa mohou pomoci odhalit manipulaci



## Použít detekční software

Existují AI nástroje schopné rozpoznat Deepfake

Je důležité ověřovat pravost obsahu a nedůvěřovat všemu na internetu.

## Kybernetická obrana

### Detekce hrozeb

Použití AI k detekci škodlivé aktivity a identifikaci podezřelých vzorků.

### Analýza chování uživatelů

Sledování aktivity uživatelů a identifikace odchylek od normálu pomocí AI.

### Automatizovaná reakce

Využití AI k rychlému zastavení útoku a omezení škod.

### Prediktivní analýza

Předvídání budoucích hrozeb a zranitelností na základě analýzy dat pomocí AI.

### Školení zaměstnanců

Využití AI k vytvoření realistických simulací pro výcvik zaměstnanců v oblasti kybernetické bezpečnosti.

### Automatizované testování

Testování zabezpečení systémů pomocí AI, která dokáže simulovat reálné útoky.