VYSOKÁ ŠKOLA EVROPSKÝCH A REGIONÁLNÍCH STUDIÍ, Z. Ú., ČESKÉ BUDĚJOVICE

BAKALÁŘSKÁ PRÁCE

Moderní bezpečnostní systémy a využití umělé inteligence v rámci prevence kriminality

Autor práce: Denisa Műller Honsová DiS.

Studijní obor: Bezpečnostně právní činnost ve veřejné správě

Forma studia: Kombinovaná

Vedoucí práce: Prof. JUDr. Jozef Meteňko, PhD.,

Katedra: Katedra právních oborů a bezpečnostních studií

2020

Prohlašuji, že jsem bakalářskou práci vypracoval(a) samostatně, na základě vlastních zjištění a s použitím odborné literatury a materiálů uvedených v seznamu použitých zdrojů.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce v elektronické podobě ve veřejně přístupné části infodisku VŠERS, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky vedoucí(ho) a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce systémem na odhalování plagiátů.

.....

VYSOKÁ ŠKOLA EVROPSKÝCH A REGIONÁLNÍCH STUDIÍ, z.ú.
Žižkova 6, 370 01 České Budějovice

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Jméno a příjmení studenta: Denisa Műller Honsová DiS.

Studijní program: Bezpečnostně právní činnost
Studijní obor: Bezpečnostně právní činnost ve veřejné správě
Forma studia: Kombinovaná
Místo studia: Příbram

Název bakalářské práce: Rizika využití umělé inteligence v bezpečnostních systémech

Název bakalářské práce v anglickém jazyce: Risks of using artificial intelligence in security systems

Katedra: Katedra právních oborů a bezpečnostních studií
Vedoucí bakalářské práce (jméno a příjmení, titul): Prof. JUDr. Jozef Meteňko, PhD.

Datum zadání bakalářské práce: říjen 2018

CÍL BAKALÁŘSKÉ PRÁCE: Cílem práce je zjistit, jak velká rizika přináší užití umělé inteligence v oblasti bezpečnostních systémů, a to při využití za současného stavu. Sekundárním a terciálním cílem je, postavení vlastní teorie o využití umělé inteligence, proti aktuálnímu trendu, následná analýza rozdílu dopadů a návrh opatření, které by snížilo všechna rizika.

Student: Denisa Műller Honsová DiS.	25.10.19 *datum*	*podpis*
Vedoucí práce: Prof. JUDr. Jozef Meteňko, PhD.	16.11.2019 *datum*	*podpis*

Schvaluji zadání bakalářské práce:

Vedoucí katedry: doc. JUDr. Roman Svatoš, Ph.D.	18.11.2019 *datum*	*podpis*
Prorektorka pro studium a vnitřní záležitosti: RNDr. Růžena Ferebauerová	19.11.19 *datum*	*podpis*
Pověřený rektor: doc. Ing. Jiří Dušek, Ph.D.	23.11.2019 *datum*	*podpis*

Děkuji vedoucímu bakalářské práce panu JUDr. Jozefu Meteňkovi, PhD. za cenné rady, připomínky a metodické vedení práce.

ABSTRAKT

MÜLLER HONSOVÁ, D. *Moderní bezpečnostní systémy a využití umělé inteligence v rámci prevence kriminality: bakalářská práce.* České Budějovice: Vysoká škola evropských a regionálních studií, 2020. Vedoucí bakalářské práce: Prof. JUDr. Jozef Meteňko, PhD.,

Klíčová slova: bezpečnostní systémy, umělá inteligence, neuronová síť…

Práce řeší problematiku bezpečnostních systémů z pohledu rozvoje a obohacení o umělou inteligenci. Klade důraz na vyobrazené komplexní problematiky neuronových sítí, jejich užití a potenciálních rizik, které z užívání plynou. Dále poukazuje na složitost neuronových sítí, jejich schopnost se učit a filosofický problém s rozporem v Asimovových zákonech robotiky.

ABSTRACT

MÜLLER HONSOVÁ, D. *Modern security systems and usage of artificial intelligence in crime prevention : Bachelor Thesis.* České Budějovice: The College of European and Regional Studies, 2020. p. Supervisor: Prof. JUDr. Jozef Meteňko, PhD.,

Key words: security systems, artificial intelligence, neural networks

The work addresses the issue of security systems in terms of development and enrichment with artificial intelligence. It emphasizes the depicted complex issues of neural networks, their use and potential risks that arise from use. It also points out the complexity of neural networks, their ability to learn and the philosophical problem with the contradiction in Asimov's laws of robotics.

Obsah

Úvod

Moderní bezpečnostní systémy jsou nedílnou součástí našich životů a jejich potřeba má vzrůstající tendence. Každý den se setkáváme se zprávami, které nás informují o krádežích, poškozování majetku, neoprávněnému vniknutí a jiných trestných činnostech, kterým by šlo předcházet, nebo snižovat škody z nich vzniklé, pomocí bezpečnostních systémů.

Historie bezpečnostních systémů je velmi široká a sahá až do starověku, přičemž evoluce tohoto technologického odvětví, byla plná slepých uliček, kuriózních „systémů“ a všechny tyto etapy, formovaly dnešní svět bezpečnostních systémů.

Tato bakalářská práce ukazuje, kam až se bezpečnostní systémy dostaly a jak velké jsou rozdíly v jednotlivých typech systémů, jejich užití a rizik z nich vyplývající.

Stejně jako objev metalurgických procesů změnil bezpečnostní systémy a obohatil je o kovové prvky, závory, mříže, kované zámky a další železné části, tak vznik umělé inteligence, přinesl možnosti, o kterých zatím pořádně ani nevíme.

Rizika, která to ale obnáší, taktéž zatím nejsou přesně zmapována a v moderní době již není žádoucí, testovat schopnosti jednotlivých systémů pomocí metody „pokus X omyl“.

Možná si ani neuvědomujeme, co vše jsou bezpečností systémy, kam až zasahují a jaké jsou jejich funkce a nástrahy.

1 Cíl a metodika bakalářské práce

Cílem bakalářské práce je vytvoření detailní a funkční analýzy dílčích částí bezpečnostních systémů a technologií využívaných v rámci bezpečnostních systémů. Analýza si klade za cíl explicitně doložit klady a zápory jednotlivých systémů, a to včetně návrhu zlepšení v jednotlivých oblastech. Některé návrhy jsou procesního charakteru a určují například postupy při analýze aplikovatelnosti jednotlivých prvků v konkrétním místě, a to s ohledem na využitelnost a potřeby objektu a uživatelů.

Druhotným přínosem, je pak prezentace vlastní myšlenky o rizikovosti užití umělé inteligence v bezpečnostních systémech, což je v dnešní době raketově rostoucím trendem.

Právě v rámci umělé inteligence existuje velký prostor pro zlepšení implementace bezpečnostních systémů, a to „end-to-end“. Počínaje analýzou systému, jeho schopností a bezpečnosti, následně analýzou prostředí, aplikovatelnosti, využití a v neposlední řadě rizik, rentability a schopnosti vhodně a včas poskytovat podporu.

Metodiky užité v této bakalářské práci se opírají o klasické základní metodiky, jejichž jednoduchost dovoluje získat rychle a efektivně výsledky.

Komparativní metoda [1]– Porovnání klasických bezpečnostních systémů, jejich přínosů, využitelnosti a rizik, vůči systémům využívající umělou inteligenci, nebo přímo systémů řízených umělou inteligencí.

Pozorování – Předložení výsledků na základě již proběhlých a zmapovaných experimentů. V rámci rozvoje všech výše uvedených systémů, proběhlo již několik stovek experimentů, které byly vedeny vědeckou, nebo technickou obcí. Na druhou stranu můžeme čerpat i ze zkušeností, které nám přináší armádní sféra, nebo civilní sektor. Ne všechny experimenty jsou samozřejmě zdokumentovány „as-is“, neboť podléhají určitému utajení (apriori se jedná o armádní projekty, nebo projekty zahrnující bezpečnost státu). Výhodou je, že v dnešní době existuje velké množství civilních subjektů, které se zaměřují na tuto problematiku a provádí transparentní vývoj a testování.

[1] *Pojem metoda komparativní* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://slovnik-cizich-slov.abz.cz/web.php/slovo/metoda-komparativni>.

Asimovovi zákony robotiky[2]– Ačkoliv se nejedná o druh metodiky, Asimovovi zákony robotiky jsou nosným pilířem v oblasti „etické“ a „filosofické“ otázky při využívání umělé inteligence a autonomní digitalizace. V rámci bakalářské práce jsou jednotlivé systémy využívající umělou inteligenci, nebo řízené umělou inteligencí, postaveny před základní otázky Asimovových zákonů robotiky. Výsledek je pak zásadním kritériem pro určení aplikovatelnosti systémů v praxi a jejich bezpečnosti pro uživatele a jejich okolí.

Poslední částí bakalářské práce je potom prezentace vlastní teze o vývoji inteligentních bezpečnostních systémů, jejich realizovatelnosti v blízké budoucnosti a návrh opatření, které může přinést hladší průběh změny bezpečnostních systémů ve veřejném prostoru.

[2] ANDERSON, S. (2011). *The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics*. In M. Anderson & S. Anderson (Eds.), Machine Ethics (pp. 285-296) Cambridge: Cambridge University Press. doi:10.1017/CBO9780511978036.021 [cit. 2020-04-19]
Dostupné z WWW: < https://www.cambridge.org/core/books/machine-ethics/unacceptability-of-asimovs-three-laws-of-robotics-as-a-basis-for-machine-ethics/D58C8BAD402DF52AD2785C17A68431EB >.

2 Využití jednotlivých bezpečnostních systémů v připravených scénářích

Pro následující analýzu využitelnosti jednotlivých bezpečnostních systémů, staví bakalářská práce jednotlivé systémy do předem připravených scénářů. Každá modelová situace, mapuje tři základní pilíře:

1. ***Překonatelnost bezpečnostního systému*** – určuje procentuální šanci na překonání bezpečnostního systému a dobu, po kterou může v konkrétní situaci odolávat útočníkovi;

2. ***Schopnost synchronizace a kooperace s dalším systémem nebo IZS***- možnosti systému navázat na další kroky jiných, přidružených bezpečnostních systémů a případná kooperace se bezpečnostními složkami;

3. ***Náročnost instalace a obsluhy bezpečnostního systému*** – porovnání účinnosti systému vůči náročnosti instalace a obsluhy. Taktéž je zde uvážena náchylnost systému na okolní vlivy a jeho chybovost.

Jako modelové situace, slouží následující:

1. ***Zabezpečení RD na periferii Prahy*** – útočník se snaží vniknout do objektu přes dveře a okna;

2. ***Zabezpečení bytu ve městě*** – útočník se snaží vniknout do objektu přes dveře a balkón;

3. ***Zabezpečení firemního skladu*** – útočník se snaží vniknout do objektu přes dveře, vrata pro nakládku zboží a okna.

U každé modelové situace je popsán konkrétní bezpečností systém, který je nainstalován v objektu. Všechny bezpečnostní systémy kombinují prvky aktivní i pasivní ochrany, včetně hlídání objektu psem nebo ostrahou. Protože v rámci bezpečnostních systémů existuje pouze jediný koeficient (nebo spíše index), a to časové náročnosti pro překonání systému, bylo třeba vytvořit vlastní, subjektivní, koeficient.

Tento koeficient/index určuje odolnost systému na základě:

- časové náročnosti pro překonání systému;

- nutnosti nástrojů, které je třeba mít pro překonání systému;

- počtu osob, které jsou potřeba k překonání systému.

Všechny výše uvedené body, mají svojí vlastní hodnotu „1“ a celkovou škálu hodnocení „1-5“. Kde nejnižší hodnotou je „1“. Následně posuzujeme:

A.) Časová náročnost

a. 1 – nejnižší – do 1 minuty;

b. 2 – nízká – do 3 minut;

c. 3 – střední – do 5 minut;

d. 4 – vyšší – do 10 minut;

e. 5 – vysoká – 10+ minut.

B.) Nutnost nástrojů

a. 1 – bez nutnosti nástrojů;

b. 2 – běžné nástroje (šroubováky, kleště, aku nářadí);

c. 3 – složitější nástroje (šperháky, elektrické nářadí);

d. 4 – specializované nástroje (speciální klíče, rušičky, elektronické přemosťovače, kopie čipových karet, čtečky);

e. 5 – nestandardní nástroje (zahrnuje prvky sociálního inženýrství, vzdálené PC útoky, přenosné stanice, atd).

C.) Počet osob

a. 1 – pouze jedna osoba;

b. 2 – dvě osoby;

c. 3 – tři osoby;

d. 4 – čtyři osoby;

e. 5 – specializované týmy.

U každé situace pak dochází k explicitní determinaci potřeb k překonání daného systému a součtem všech výše uvedených hodnot, získáme výsledek.

A.) 1-3 body – systém je snadno narušitelný a nedokáže dlouho odolávat útočníkům;

B.) 4-7 bodů – „běžné systémy“;

C.) 8-11 bodů – systém odolává řádově delší, než standardní dobu a jeho bezpečnost je vysoká;

D.) 12-15 bodů – velmi časově náročné k překonání, odolává dlouhou dobu a k jeho překonání je třeba velká příprava.

2.1 Zabezpečení RD na periferii Prahy

Pro naší modelovou situaci slouží standardní RD na periferii Prahy, který nedisponuje specializovaným zabezpečovacím systémem. Vybavenost z hlediska zabezpečovacího systému je:

- plot včetně branky a vjezdové brány – z pletiva;

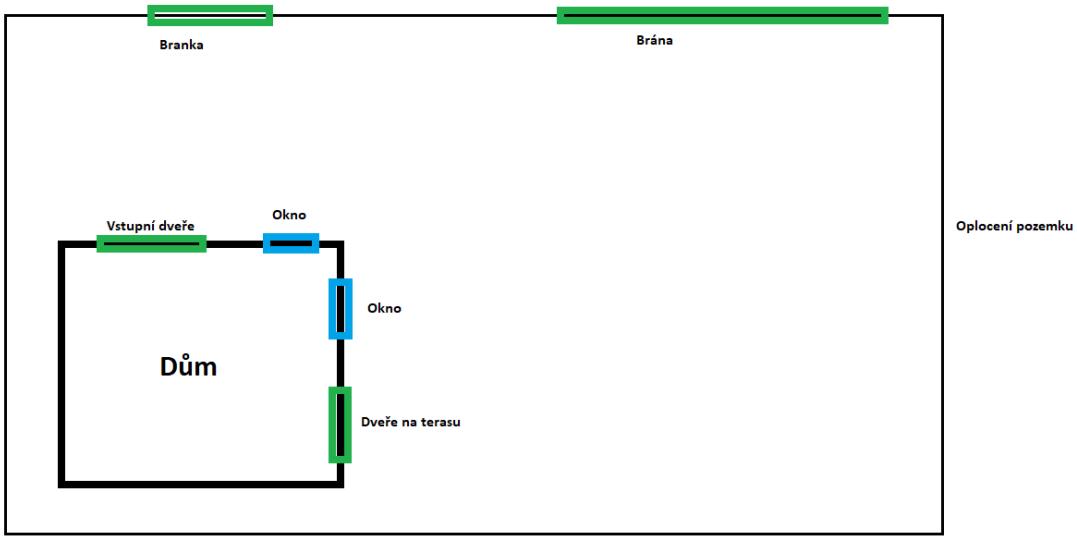
- bezpečnostní zámek vchodových dveří třídy RC2[3];

- bezpečnostní kliky oken v přízemí.

Dům nedisponuje žádným alarmem, nebo jiným druhem EZS a nemá žádné mříže, nebo rolety.

[3] *Co je to BEZPEČNOSTNÍ TŘÍDA?* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://www.bezpecnostni-dvere-mrize-kavan.cz/co-je-to-bezpecnostni-trida/>

Obrázek 1: Nákres zabezpečení RD na periferii Prahy



Zdroj: vlastní tvorba

V tomto případě je analýza za pomoci výše popsaného koeficientu:

Časová náročnost – 1

Nutnost nástrojů - 1

Počet osob – 1

Celkem: 3 (v případě nedestruktivního jednání – rozbití okna – bude výsledek 4)

Výše uvedený systém není schopen dlouhodobě odolávat. Útočník může přelézt plot nebo jej rozplést, následně rozbít okno (případně pomocí šroubováku a kladiva, zničit zámek dveří) a objekt narušit.

Z hlediska vyhodnocení analýzy jednotlivých bodů:

Překonatelnost bezpečnostního systému **–** postačí rozbít okno a objekt je narušen;

Schopnost synchronizace a kooperace s dalším systémem nebo IZS – mechanické zabezpečovací systémy nejsou schopny kooperovat s jinými systémy a nemají možnost kontaktovat bezpečnostní složky;

Náročnost instalace a obsluhy bezpečnostního systému – instalace bezpečnostního zámku, bezpečnostních kliček oken a vnější obvodové ochrany, nevyžaduje žádnou náročnou instalaci. Většinově lze tyto prvky instalovat svépomocí;

Doporučení:

- instalace ochrany oken – mříže;

- instalace bezpečnostního zámku vyšší bezpečnostní třídy;

- instalace alespoň základního EZS.

2.2 Zabezpečení bytu ve městě

Zabezpečení bytu se i s ohledem na umístění, v oblasti městské zástavby, liší od zabezpečení RD. Zpravidla je zabezpečení bytů větší, než u zabezpečení RD, neboť majitelé RD zabezpečení s oblibou zanedbávají:

- bezpečnostní zámek vchodových dveří třídy RC4; [4]

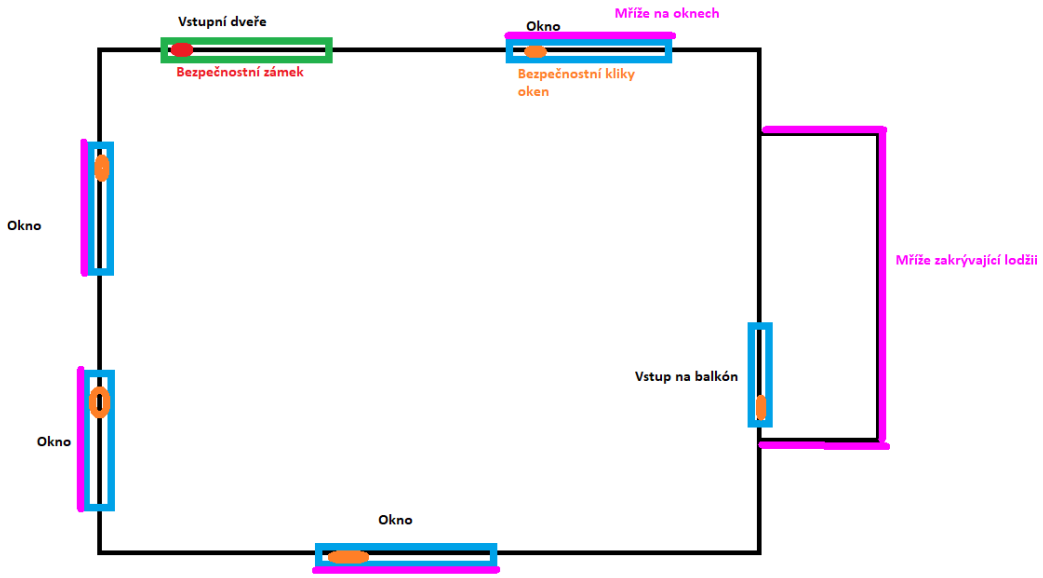
- bezpečnostní kliky oken v přízemí;

- mříže na oknech a lodžii;

- základní EZS – čidla otevření dveří a oken.

[4] *Co je to BEZPEČNOSTNÍ TŘÍDA? [online]*. [cit. 2020-04-19] *Dostupné z WWW: < https://www.bezpecnostni-dvere-mrize-kavan.cz/co-je-to-bezpecnostni-trida/>*

Obrázek 2: Nákres zabezpečení bytu ve městě



Zdroj: vlastní tvorba

V modelové situaci máme byt v městské zástavbě, který se nachází v přízemí, tedy na výškové úrovni, která je překonatelná bez žebříku a nástrojů jemu podobných.

V tomto případě je analýza za pomoci výše popsaného koeficientu:

Časová náročnost – 4

Nutnost nástrojů - 3

Počet osob – 2

Celkem: 9

K narušení takto zabezpečeného objektu je možno zvolit pouze dveře. S ohledem na hluk, který by způsobil pokus o narušení objektu přes okna nebo lodžii (s přihlédnutím na přítomnost mříží), neexistuje jiná volba než právě dveře. K vniknutí do objektu přes dveře, musí útočníci překonat bezpečnostní zámek (třída RC4 odolává řádově 10 minut, což může být zkráceno třeba pomocí páčidla, které sice způsobuje hluk, ale řádově menší, než řezání mříží) a následně vyřazení čidel pohybu.

U čidla pohybu záleží na dvou faktorech:

1- zdali se jedná o čidlo snímající jeden nebo dva sektory;

2- přítomnost domácích mazlíčků.

Z hlediska vyhodnocení analýzy jednotlivých bodů:

Překonatelnost bezpečnostního systému **–** Narušení objektu skrze dveře, vypáčení zámku a vyřazení čidla pohybu;

Schopnost synchronizace a kooperace s dalším systémem nebo IZS – základní EZS může poslat SMS zprávu majitelům objektu s tím, že došlo k jeho narušení. Následně záleží na dvou výše uvedených faktorech;

Náročnost instalace a obsluhy bezpečnostního systému – Instalace bezpečnostního zámku je velmi jednoduchá. K instalaci mříží je třeba firmy a základní EZS v bezdrátovém provedení, může provést majitel objektu sám.

Doporučení:

- instalace čidla pohybu snímajícího dva sektory (k eliminaci falešných poplachů v důsledku pohybu domácích mazlíčků);

- instalace kamerového systému;

- zajištění schopnosti EZS kontaktovat bezpečnostní složky.

2.3 Zabezpečení firemního skladu

Většina skladů má řádově vyšší zabezpečení než běžné objekty „civilního“ užití, tedy domy, byty a chaty. Je tomu tak hlavně z důvodu možné finanční škody, která je násobně vyšší v případě skladů, než v případě standardních domů, bytů a chat.

Náš sklad má následující zabezpečení:

- vysoká zeď s ostnatým drátem;

- ostraha objektu – pes;

- bezpečnostní zámky třídy RC6 [5]a řetězy na vratech;

- kamerový systém se sofistikovaným EZS – kontaktujícím bezpečnostní složky;

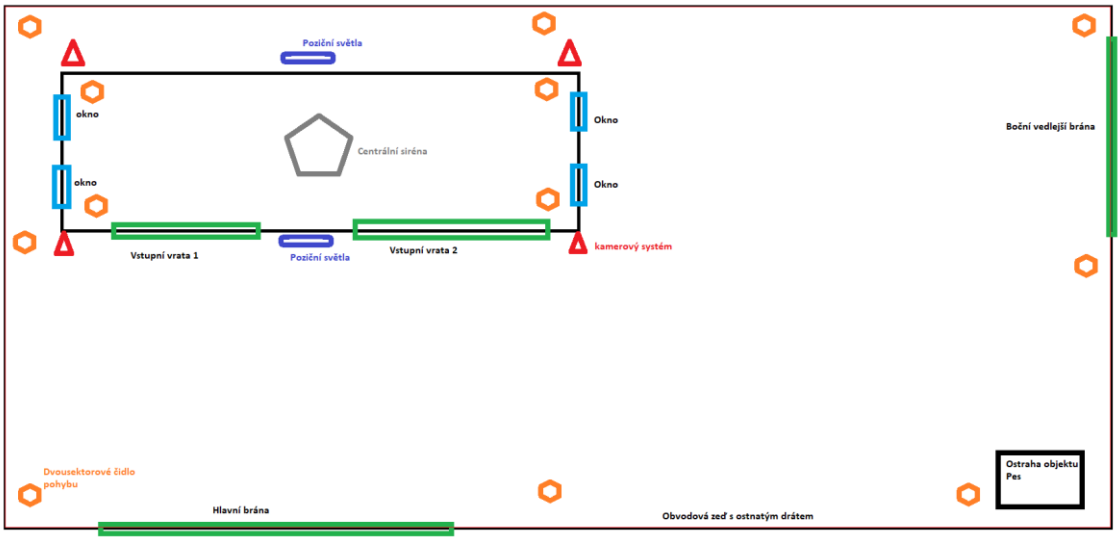
- 24/7 [6] ostraha objektu – člověk;

- dvou sektorová čidla pohybu ve vnějších i vnitřních prostorech objektu;

- poziční světla s detektorem pohybu;

- siréna.

Obrázek 3: Nákres zabezpečení firemního skladu



Zdroj: vlastní tvorba

V tomto případě, je analýza za pomoci výše popsaného koeficientu:

Časová náročnost – 5

Nutnost nástrojů - 5

Počet osob – 5 a více osob

Celkem: 15

[5] *Co je to BEZPEČNOSTNÍ TŘÍDA? [online]. Dostupné z WWW: < https://www.bezpecnostni-dvere-mrize-kavan.cz/co-je-to-bezpecnostni-trida/>*

Ideální zabezpečení jakéhokoliv objektu z hlediska standardních systémů.

Z hlediska vyhodnocení analýzy jednotlivých bodů:

Překonatelnost bezpečnostního systému **–** scénářů k překonání takového systému je hned několik. Nicméně to vyžaduje eliminaci psa, obejití ostrahy, ochrana před kamerovým systémem, překonání všech mechanických zabezpečovacích systémů, a to je velmi komplikované. Překonání takového systému vyžaduje tým útočníků a důslednou přípravu;

Schopnost synchronizace a kooperace s dalším systémem nebo IZS – Kamerové systémy dokážou kooperovat například s pozičním osvětlením, případně sirénou. Stejně tak sektorová čidla mohou spustit poplach, informovat ostrahu a bezpečnostní složky;

Náročnost instalace a obsluhy bezpečnostního systému – Vyžaduje specializované firmy, které provádí montáž tohoto druhu bezpečnostního systému;

Případná aplikace bezpečnostních systémů využívajících umělou inteligenci je z hlediska realizovatelnosti pouze na firemní budovy. V soukromém sektoru se zatím běžně nepoužívají. Nicméně pokud bychom hypoteticky vzali v úvahu, že se jedná o již dostupnou technologii, tak by mělo, bez přihlédnutí na rizika, smysl, ji aplikovat i na soukromé sektory.

Samozřejmě je otázkou, jaká by byla návratnost investice s přihlédnutím na lokalitu, vybavení, které by mohlo být odcizeno nebo celkový majetek uvnitř objektu.

V rámci výše uvedené analýzy jsme využili standardní systémy v běžném provozu. Jak je uvedeno výše, využití bezpečnostního systému s umělou inteligencí, není běžným standardem. Pokud ale vezmeme hypotetickou rovinu věci a aplikujeme modely, které byly využity v rámci testů v zahraničí, můžeme dojít k relevantním závěrům, které budou poplatné naší analýze.

Jedním z největších hráčů na poli UI v bezpečnostních systémech budov (nikoliv aplikačních bezpečnostních systémech SW) je bezesporu společnost Honeywell[7], která již navázala spolupráci i s bezpečnostními složkami USA.

V případě našich testů si představíme modelovou situaci s RD na periferii tak, jak bylo uvedeno v prvním případu naší analýzy. Rozdíl bude v tom, že integrujeme veškeré, již otestované, prvky inteligentních systémů s UI a vyhodnotíme je dle stejných kritérií, jako standardní systémy. V tomto bodě je nám již jasné, že systémy s UI budou v oblasti bezpečnosti dominovat. Na komplexní problematiku jejich „neduhů" se podíváme v následující kapitole.

2.4 Zabezpečení RD na periferii Prahy systémem s UI

Abychom zachovali modely tak, jako v předešlých případech, využijeme stejný dům, ale jeho vybavenost z hlediska zabezpečovacího systému, změníme:

- plot včetně branky a vjezdové brány – zděný;

- kamerový systém s detekcí pohybu a termo kontrolou – po obvodu zdiva;

- RFID čip na vstup;

- kamerový systém s biometrickým skenem a rozpoznáváním SPZ– pro vjezd a vstup do objektu;

- sektorová čidla pohybu včetně akustické detekce – na vnější ploše objektu;

- kamerový systém s biometrickým skenem uvnitř objektu;

- biometrický scan [8]pro vstup do objektu;

- elektronické zámky všech vnitřních a vnějších dveří;

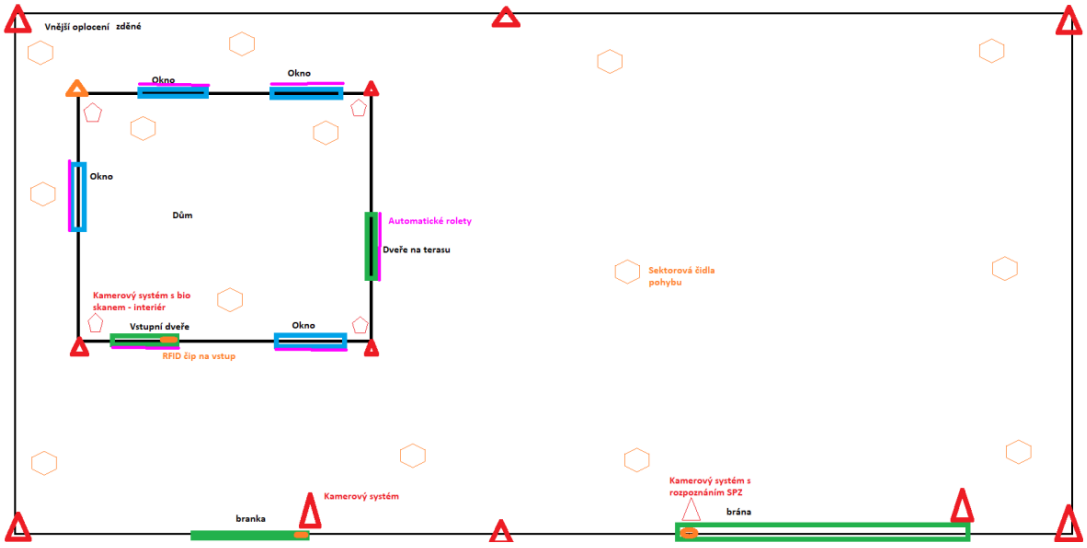
- napojení na centrální systém IZS;

- automatické okenní rolety a mříže ovládané systémem;

[7] *Honeywell* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://www.honeywell.com/en-us/company/about-us >.
[8] *Biometrics and biometric data: What is it and is it secure?* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://us.norton.com/internetsecurity-iot-biometrics-how-do-they-work-are-they-safe.html >.

- centrální jednotka ovládající bezpečnostní prvky k uzavření objektu.

Obrázek 4: Nákres zabezpečení RD na periferii prahy systémem s UI



Zdroj: vlastní tvorba

V tomto případě je analýza za pomoci výše popsaného koeficientu:

Časová náročnost – 5

Nutnost nástrojů - 5

Počet osob – N

Celkem: 15+

Výše uvedený systém je z hlediska běžného narušení objektu zcela nepřekonatelný.

Z hlediska vyhodnocení analýzy jednotlivých bodů:

Překonatelnost bezpečnostního systému – v tomto případě by bylo nutné napojit se přímo na centrální systém a vypnout jej. Vzhledem k tomu, že systémy s UI dokáží svou systémovou „identitu“ šifrovat (standardně se používá šifrování stejného charakteru jako u kryptoměn), tak není možné se do jejich jádra dostat z venku. Navíc je pro tyto

systémy využíváno tzv. „floating routing[9]“, kdy je trasování vedeno přes několik přístupových bodů, kdy většina z nich, vede do „slepé uličky“ v síti. Pro její vnější narušení je tak třeba znát přesnou trasu a přístupy;

Schopnost synchronizace a kooperace s dalším systémem nebo IZS – UI dokáže vyhodnocovat narušení systému a dle zadaný algoritmů a vyhodnocení již nasbíraných dat, kooperovat s ostatními systémy. Každý UI systém je před samotnou instalací vystaven stovkám simulací, ve kterých sbírá data. Dá se říci, že po samotné instalaci již dochází pouze ke sběru dat environmentálního a unikátního charakteru. Tedy k dospecifikování místě příslušných unikátních faktorů (chování obyvatel domu, okolní zvěř, počasí, zvyky ostatních návštěvníků, atd.) a samotný systém již dokáže reagovat na narušení;

Náročnost instalace a obsluhy bezpečnostního systému – vzhledem k tomu, že instalace systému do soukromých sektorů je zatím pouze v omezeném měřítku, tak není možné zcela relevantně kvantifikovat, jak náročná instalace je. Nicméně předpokládejme následující:

- systém potřebuje vlastní místnost pro instalaci serveru;

- každá místnost musí mít vlastní detektory pohybu, kamery, termo detektory, akustické detektory a panely ovládání;

- vnější obvodová struktura pozemku musí být připojena k elektrické síti;

- každá část pozemku musí být propojitelná s elektrickou sítí a internetem.

2.5 Výsledky analytické části práce

Výše uvedený systém se tváří jako symbol dokonalosti, což jistě v hypotetické rovině je. Možnosti narušení takového systému se limitně blíží nule. Navíc vezmeme-li v potaz poměr vynaloženého úsilí k zisku ze samotného narušení. Můžeme tak hovořit maximálně o vhodném tématu na Hollywoodský film, ale reálně by takový objekt nikdo nenarušoval. V případě, že by přesto k narušení došlo, tak systém umožňuje efektivně hrozbu eliminovat, například uzavřením útočníka v místnosti, ve které se právě nachází. Může tak učinit pomocí uzamčení elektronických zámků a uzavření oken pomocí

[9] DOYLE, J. ; CARROLL DH.J. *Routing TCP/IP, Volume 1, 2nd Edition - CiscoPress, 2005, s. 97-105.*

vnějších rolet/mříží. Následně přivolá policii pomocí zprávy pro IZS a ta si útočníka vyzvedne ve svém „přechodném vězení“ uvnitř domu.

Všechny výše uvedené systémy mají své výhody a samozřejmě i své nevýhody. Tam, kde můžeme vidět nižší pořizovací cenu a snazší instalaci, můžeme nalézt i nižší stupeň ochrany, který samozřejmě nebude dosahovat takových kvalit a výsledků, jako systémy dražší. Nebudeme-li zde spekulovat o problematice cenové politiky nastavené dle značky systému, tak se můžeme shodnout na faktu, že pořizovací cena hraje svou roli ve schopnostech systémů. Na druhou stranu dražší systémy jsou oproti tomu výrazně komplexnější a samozřejmě i složitější na instalaci, případně obsluhu. Jednoduché systémy jsme schopni si v rámci administrace obstarat většinou sami. Hlavně pokud se jedná o systémy mechanického charakteru. Složitější systémy většinou vyžadují servisní zásahy a následné revize.

Obecně ale můžeme říci, že pokud chceme dosáhnout nejlepšího výsledku, je třeba prvně specifikovat své požadavky a jasně si nastavit, co od systému požadujeme. Cena systému, stejně jako jeho vlastnosti a schopnost plnit naše požadavky, podléhá subjektivnímu hodnocení a záleží pouze na uživateli, pro jaký systém se rozhodne.

Exteriér je hlídán kamerovým systémem a sadou světel s pohybovým senzorem, které se rozsvítí vždy, když zachytí pohyb.

3 Umělá inteligence a Asimovovi zákony robotiky

Veškeré systémy umělé inteligence, ať už se jedná o bezpečnostní systémy, roboty, autonomní vozidla, nebo software, se mají řídit zákony robotiky, které byly prvně stanoveny Isaacem Asimovem.

Ačkoliv v průběhu let, docházelo k úpravě oněch zákonů, což je popsáno níže, tak se staly tyto zákony základem pro filosofii využití robotizace. Sada těchto zákonů, se nazývá „Asimovovi zákony robotiky“, nebo také „Tři zákony robotiky“.

Historie těchto zákonů sahá do roku 1942, kdy byly prvně stanoveny zákony robotiky spisovatelem Isaacem Asimovem v jeho povídce „Hra na honěnou[10]“. Jejich základní podstatou, bylo stanovit rozsah možností chování robotů v lidské společnosti. V úplně původním znění tyto zákony říkají:

1.) Robot nesmí ublížit člověku nebo svou nečinností dopustit, aby bylo člověku ublíženo;

2.) Robot musí uposlechnout příkazů člověka, kromě případů, kdy jsou tyto příkazy v rozporu s prvním zákonem;

3.) Robot musí chránit sám sebe před poškozením, kromě případů, kdy je tato ochrana v rozporu s prvním, nebo druhým zákonem.

Není nic zvláštního na tom, že všeobecně uznávaný termín, nebo v tomto případě soubor pravidel, vzešel z pera autora sci-fi[11]. Ostatně termín robot pochází od českého autora. Pokud se podíváme blíže na tyto tři zákony, v jejich původním znění, tak v nich lze najít určitou posloupnost a zároveň i problémové aspekty.[12]

[10] *Hra na Honěnou* [online]. [cit. 2020-04-19] Praha Dostupné z WWW: < https://www.databazeknih.cz/povidky/hra-na-honenou-444 >.
[11] *Sci-fi* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://literaryterms.net/science-fiction/ >.
[12] MURPHY R., R., WOODS, D., D. *Beyond Asimov: The Three Laws of Responsible Robotics 2009* [online]. Ohio, 1541-1672/09/$26.00 © 2009 IEEE Published by the IEEE Computer Society. [cit. 2020-04-19] Dostupné z WWW: <https://www.researchgate.net/publication/224567023_Beyond_Asimov_The_Three_Laws_of_Responsible_Robotics>.

3.1 Problematika vzájemného vylučování zákonů

Hned první zákon určuje, že robot nikdy nesmí ublížit člověku a zároveň nesmí nečinně přihlížet tomu, jak bude člověku ublíženo. Paradoxně se tyto věci mohou vylučovat a dostaneme se tak do bezvýchodné situace, kdy se nebude moci stroj rozhodnout a tím vyloučí zákon číslo 2. Jako příklad je velmi často uváděna situace s autonomním automobilem. V automobilu sedí čtyřčlenná rodina složená z dvou prarodičů a vnoučat. Na přechod přímo před vozidlo vstoupí čtyřčlenná rodina ve složení matka, otec a dvě děti. V naší hypotetické situaci neexistuje jiné východisko, než že automobil buď srazí a následně usmrtí rodinu na přechodu, nebo si vybere cestu vyhnutí se

a v následném nárazu usmrtí posádku automobilu. Jak by se měl stroj v tomto případě zachovat? Vezmeme-li v potaz první z Asimovových zákonů robotiky, tak by neměl automobil ohrozit posádku, protože by to bylo přímé ublížení člověku. Na stranu druhou ale nemůže srazit rodinu na přechodu, protože tak by svou nečinností přihlížel usmrcení jiného člověka. Dostáváme se tak do filosofického paradoxu, který může vyústit v usmrcení všech účastníků.

Druhý zákon zase vylučuje možnost využití robotů, nebo systémů umělé inteligence jako ochranných, nebo obranných prvků. Pakliže útočník přímo konfrontuje svou agresí majitele bezpečnostního systému s AI, tak majitel logicky vydá příkaz k obraně. Pokud si představíme robota, jako mechanickou entitu, tak by měl zakročit a zneškodnit útočníka. Jenže v tomto případě nesmí útočníkovi ublížit, zároveň ale nesmí útočníka nechat ublížit majiteli. Pravdou je, že v případech bezpečnostních systémů, které dokáží například ovládat elektronické zámky, a tak uzamknout útočníka v odlehlé části domu, se nejedná o závažné morálně-filosofické dilema.

Nejzásadnější a nejvíce diskutovaný problém je, dle mého osobního názoru, ukryt ve třetím ze zákonů. Ten nám říká, že robot, nebo náš bezpečnostní systém, musí za každou cenu chránit sám sebe před poškozením, pokud to není v rozporu s prvním, nebo druhým ze zákonů.[13]

[13] MURPHY R., R., WOODS, D., D . *Beyond Asimov: The Three Laws of Responsible Robotics 2009* [online]. [cit. 2020-04-19] Ohio, 1541-1672/09/$26.00 © 2009 IEEE Published by the IEEE Computer Society. Dostupné z WWW: < https://www.researchgate.net/publication/224567023_Beyond_Asimov_The_Three_Laws_of_Responsible_Robotics>

Tady se dostáváme na tenký led, ve kterém hraje roli velké množství faktorů. Příkladem bude snaha majitele bezpečnostního systému o vypnutí centrální jednotky. Pakliže vezmeme v potaz, že UI funguje na principu rozšířené neuronové sítě[14] se schopností učit se (kvantová jednotka nižšího řádu[15]), tak může vyhodnotit tuto snahu jako zásah proti vlastní vnitřní integritě. Jejím primárním cílem je ochrana domu a jeho obyvatel, a to za jakékoliv okolnosti. V případě vypnutí přijde o svůj základní účel a nedojde tak k naplnění předurčení samotného systému. UI tak vyhodnotí tento krok jako útok na sebe samotnou a uzamkne, například pomocí elektronických zámků, majitele domu v jedné z místností. Ten může potom v místnosti zemřít, třeba hlady, nebo žízní. Menším problémem může být to, že po uzamčení majitele v jedné z místností, zavolá systém jednotky IZS a majitel zaplatí zbytečný výjezd.

V principu se jedná o to, že logika rozhodování a tíha morálních dilemat a těžkých rozhodnutí je komplikovaná i pro samotného člověka. Posuzujeme při nich převážně emoční dopad a řídíme se pocity. Hlavním „tahounem“ takového rozhodnutí je tedy něco, co stroj nikdy nemůže mít, protože pocit, emoce a vnitřní rozpoložení, není něco, co by se dalo obecně elektronicky kvantifikovat tak, aby byl jakýkoliv algoritmus schopný je inteligentně replikovat ve formě syntaxiálního zápisu kódu. Rozhodovací mechanismus by byl daleko za možnostmi, byť teoretizovaných, kvantových počítačů a počet možností, které máme, a které můžeme v jednu chvíli zvažovat a v druhé chvíli provést, by byl daleko za hranicí množiny představivosti jakéhokoliv stroje.

Ačkoliv Asimov přiznal autorství někomu jinému, a to konkrétně R. Daneelu Oliwavovii, přidal později ještě tzv. „nultý zákon robotiky“, který následně používal jako odkaz ve všech předchozích/následujících zákonech. Všechny tedy fungovali pouze v případě, že nebyly v rozporu s nultým zákonem:

0) **Robot nesmí ublížit lidstvu nebo svou nečinností dopustit, aby mu bylo ublíženo.**

Zařazením tohoto zákona, mezi obecné zákony robotiky, došlo k zásadní změně ve vnímání umělé inteligence. Došlo totiž k tomu, že se poukázalo na problém „globalizace“. Faktem totiž je, že to, co je dobré pro jedince, nemusí být dobré

[14] *What is a neural network?*. [cit. 2020-04-19] Dostupné z WWW: < https://www.techradar.com/news/what-is-a-neural-network >.
[15] *What is a qubit?* [cit. 2020-04-19] Dostupné z WWW: < https://www.quantum-inspire.com/kbase/what-is-a-qubit/ >.

pro společnost a naopak. Otázkou zůstává, kdo posuzuje, co je dobré pro společnost a kdo by měl tak být tím arbitrem, který vyhodnotí správnost rozhodnutí UI.

Po přečtení nultého zákona mi automaticky naskočí film „Terminátor[16]“, ve kterém umělá inteligence vyhodnotila, tady tedy in extremo, lidstvo jako hrozbu samo pro sebe a rozhodla se jej zničit. Opět se tak dostávám do morálně tíživých a obtížných otázek, jak by umělá inteligence v rámci bezpečnostních systémů obrany státu, reagovala například na napadení země, kterou chrání? Nemůže být nečinná, ale nemůže zakročit. Zákony robotiky tak vlastně vyvrací a popírají sami sebe, ačkoliv jsou jediným mechanismem, který může explicitně determinovat okruh působnosti robotů a umělé inteligence.

V průběhu času a s progresivním vývojem technologií, docházelo postupně k doplňování zákonů robotiky:

4**) Robot se musí vždy prokazovat jako robot;**

5) **Robot musí vědět, že je robot.**

Další dva zákony, které, v tomto případě, doplnily Bulharští sci-fi autoři Ljuben Dilov a Nikola Kesarovski. Čtvrtý zákon tak apeluje na striktní dodržování průkaznosti odpovědnosti. Robot/UI tak musí vždy sama sebe prokazovat jako robota/UI. Nikdy nesmí sama sebe zaměňovat za jinou entitu, ať už mechanického, nebo organického charakteru. [17]

Zákon číslo pět naopak cílí na vnitřní integritu všech systémů, které si, v případě, že se jedná o autonomní systémy, například na neuronové bázi, musí uvědomovat sama sebe a musí pracovat s předpokladem, že jsou tím, co jim bylo nastaveno v době vzniku. Jedná se tak vlastně o zákon, který podporuje první 4 (s nultým zákonem 5) zákony,

[16] *Terminator* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://www.thisisbarry.com/film/terminator-film-series-all-plots-explained/ >.
[17] MURPHY R., R., WOODS, D., D . *Beyond Asimov: The Three Laws of Responsible Robotics 2009* [online]. Ohio, 1541-1672/09/$26.00 © 2009 IEEE Published by the IEEE Computer Society. [cit. 2020-04-19] Dostupné z WWW: < https://www.researchgate.net/publication/224567023_Beyond_Asimov_The_Three_Laws_of_Responsible_Robotics>

protože až na základě zákonu číslo pět, může robot sám sebe posuzovat ve všech případech jako mechanickou entitu s elektronickým vědomím. [18]

Zákony robotiky a jejich základy jsou tedy v plné režii Isaaca Asimova, a proto je stále nazýváme Asimovovi zákony robotiky, ačkoliv se v průběhu času změnili. Docházelo k řadám otázek a s vývojem se objevovali „hluchá“ místa. Zásadní otázky, které narušovaly, nebo přímo konfrontovaly zákony robotiky, byly otázky nevědomého porušení zákonů, nejasné definice pojmu „robot“ a v neposlední řadě pojmu „člověk.“

3.2 Otázka nevědomého porušení zákonů

Prvně formulována spisovatelem Elijahem Baleyem v jeho díle The Naked Sun. Elijah Baley v nich položil jednu ze zásadních otázek, zdali neexistuje možnost, jak zneužít nevědomost robotů, například k páchání trestné činnosti. Elijah dodává, že by zločinec mohl i rozdělit úkony trestné činnosti mezi více robotů takovým způsobem, aby jednotlivý robot nedokázal rozeznat důsledky jeho činů vedoucí k ublížení zdraví člověka. Stanovil pro zákony úpravu u prvního zákona: *Robot nesmí vědomě ublížit člověku nebo svou nečinností za plného vědomí dopustit, aby bylo člověku ublíženo*. Pozměnění slov má podstatu v tom, že by roboti mohli být jako nástroje zneužiti i k spáchání vraždy, aniž by si sami byly vědomi závažnosti úkonů, které provádí.

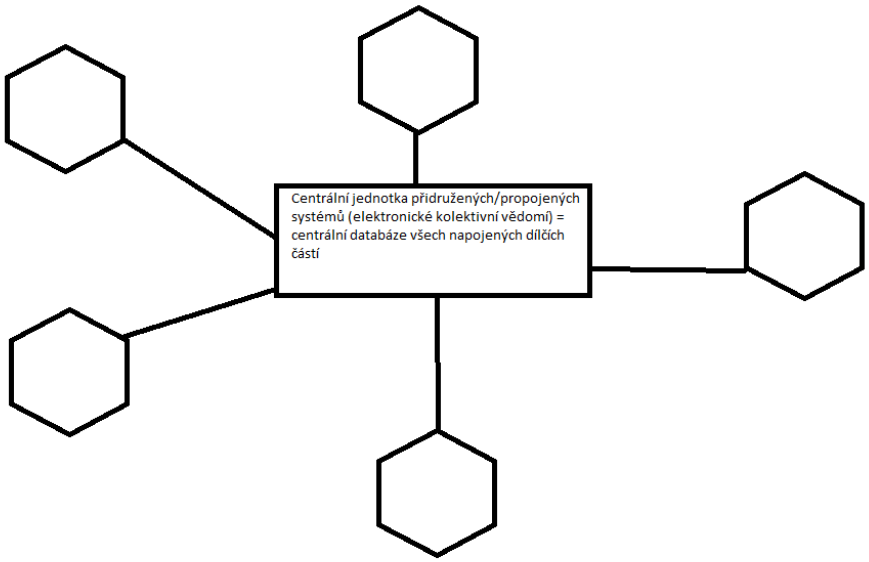
Nicméně i v tomto případě lze hledat určitou „díru“ v samotné formulaci a koncepci zákona, nebo jeho upraveného znění. Autor zde pracuje s premisou[19], že jasně definovaný pojem zajistí dodržení a nabytí určitého „globálního kausálního [20]chápání“. Problémem však je, že pakliže se budeme bavit o dílčích systémech bez vlastní interní sítě nebo synchronizace, tak nejsme schopni zajistit, že bude tento zákon dodržen. Problematiku je možné vizualizovat dle přiloženého obrázku.

[18] RUSSEL S.J., NORVIG.P. *Artificial Intelligence A Modern Approach Third Edition* : New Jersey; Upper Saddle River, 2010. 1151 s.

[19] *Premisa* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://www.czechency.org/slovnik/LOGICK%C3%89%20VYPL%C3%9DV%C3%81N%C3%8D >.

[20] Kauzalita [online]. [cit. 2020-04-19] Dostupné z WWW: < https://encyklopedie.soc.cas.cz/w/Kauzalita >.

Obrázek 5: Schéma propojení jednotlivých entit [21]systému s řídícím elementem



Zdroj: vlastní tvorba

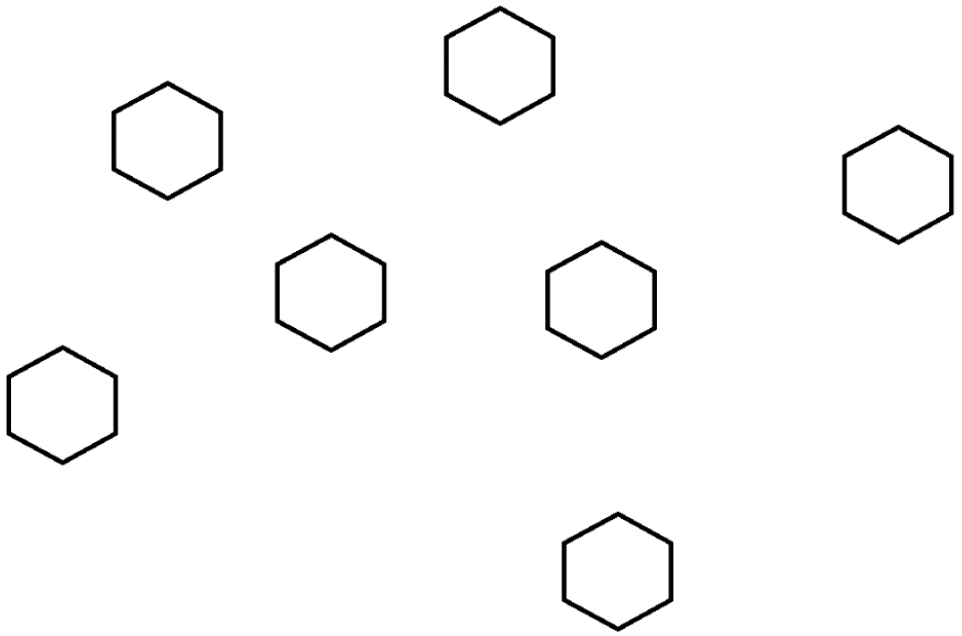
První obrázek znázorňuje systém, který je vybaven centrální databází a ovládáním. Jedná se, v případě inteligentních systémů, neuronových sítí a kvantových počítačů, o elektronické kolektivní vědomí, které slouží jako centrální mozek všech připojených systémů. V tomto případě můžeme mluvit o tom, že upravený zákon dle Elijaha Baylea, bude fungovat perfektně. Jakákoliv snaha o rozdělení činností v přímém rozporu se zákony robotiky, ačkoliv to z dílčích úkonů nebude zcela jasné a ani jeden z dílčích úkonů, nebude v přímém rozporu se zákony robotiky, nebude možné realizovat.

V centrálním systému dojde k vyhodnocení kauzálních souvislostí a vytvoření predikce funkčního celku. Následkem toho bude přesná realizace modelu chování a dopadů jednotlivých konsekvencí na dodržení zákonů robotiky. Bezpečnostní mechanismus zákonů bude tedy fungovat perfektně, protože centralizované řízení zajistí spojení jednotlivých činností. [22]

[21] *Entita* [online]. Praha [cit. 2020-04-19] Dostupné z WWW: < https://it-slovnik.cz/pojem/entita >.
[22] McCAULEY, L. *AI Armageddon and the Three Laws of Robotics.* Praha, 2006, s. 13.
.

Oproti tomu následující obrázek znázorňuje nejčastější zastoupení inteligentních systémů, i když využívají UI.

Obrázek 6: Schéma jednotlivých entit systému bez propojení a řídícího elementu



Zdroj: vlastní tvorba

Ani jeden ze systémů není nijak propojen. Proto při rozmělnění celého činu do několika dílčích aktivit, nebude možné nahlížet na jednotlivé aktivity jako na celek a vyhodnotit tedy jeho komplexní dopad. Představme si příklad číslo jedna jako firmu se zabezpečovacím systémem, který je řízen centrálně. Tam bude onen zákon fungovat skvěle, protože centrální mozek vyhodnocuje jednotlivé dílčí části. V případě druhém si můžeme představit jednotlivé systémy, jako bezpečnostní systém doma, ve firemním skladu a na chatě. Ačkoliv jsme majiteli všech tří objektů, nemají společnou jednotku a každý funguje nezávisle na sobě. [23]

Zásadní je potom otázka, jak aplikovat všechny uvedené zákony u stávajících nebo vyvíjených technologií.

„Roboti a přístroje obdařené umělou inteligencí se musí, chování podřízené Třem zákonům robotiky, učit. Jejich lidští stvořitelé si musí vybrat integraci programování tak,

[23] CLARKE, R. (2011). *Asimov's Laws of Robotics*. In M. Anderson & S. Anderson (Eds.), Machine Ethics (pp. 254-284). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511978036.020

aby se jím roboti řídili. Již existují robotická zařízení, která mohou způsobit újmu na zdraví a bolest na živých organismech aniž by chápali hranici bolesti, u které je vhodné přestat. Většina je vytvářena s bezpečnostní pojistkou, jako jsou různé nárazníky, varovné světelné a zvukové signály, ochranné mříže nebo také vyhrazené oblasti pro minimalizaci možných scénářů nehod. Dokonce i ty nejsložitější robotická zařízení nedokáží plně aplikovat Tři zákony robotiky. Pro naplnění chápání podstaty zákonů by byl potřeba velký rozvoj umělé inteligence. I v momentu, kdy se lidé a roboti inteligencí vyrovnají, tak jim nadále brání kulturní/kontextuální a etické bariéry. S tím souvisí i návaznost na zákony a přestupky, které se nevztahují na roboty. Jak pokročil vývoj robotických zařízení, roste s nimi zájem o vývoj návodů a jistot pro zacházení a různé využití. V roce 2007 dočasný pisatel v redakci Science argumentoval názory na téma „Etika u robotů“. Autor Robert J. Sawyer komentoval financování vývoje robotiky, které je hlavně financováno armádou Spojených států amerických. Kdy je nepravděpodobné zabudování těchto norem do designu zařízení (již jsou využívána bezpilotní letadla s cílem zabíjet). V odlišné eseji Sawyer zobecňuje komentář: Výzkum umělé inteligence je forma podnikání, kdy ty jsou svou podstatou bez zájmu o vytváření základních bezpečnostních pojistek.[24] To se týká obzvláště překračujících do filosofické úrovně (například průmysl s tabákem, automobilový průmysl, atomový vývoj). Ani u jednoho nejsou od počátku dostatečně brána nutná opatření na pravou míru, každý z nich odolal snahám o bezpečnostní pojistky zvenčí a ani jeden nepřijal edikt o naprosto žádném vlivu na zdraví člověka. [25]

Dále David Langford navrhl neformálně řečenou sadu zákonů:

- robot nezpůsobí zranění úřední osobě, ale zasáhne a zneškodní pachatele trestné činnosti;
- robot se podřídí příkazům pověřené osoby, kromě případu, kdy by příkazy byly v rozporu s třetím zákonem;
- robot hledí na zajištění vlastní bezpečnosti za využití smrtelného ozbrojení, protože sám robot je zatraceně drahý.

[24] McCAULEY, L. *AI Armageddon and the Three Laws of Robotics.* Praha, 2006, s. 19-21
[25] ANDERSON, S. (2011). *The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics*. In M. Anderson & S. Anderson (Eds.), Machine Ethics (pp. 285-296). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511978036.021

Roger Clark napsal práce analyzující komplikace při implementaci těchto zákonů, hledíce do pravděpodobné budoucnosti, kdy budou moci být v systémech využity. Komentoval, že i přes svou obecnou formu byly Tři zákony robotiky úspěšným literárním prostředkem. Možná ironicky, nebo také umělecky dobře zprostředkované, celkové dílo příběhů Asimova vyvracuje počáteční tvrzení. Nelze důvěryhodně řídit chování robotů rozdělením a aplikací sady pravidel.

Na druhou stranu pozdější novely Isaaca Asimova, The Robots of Dawn, Roboti a Impérium a Nadace a Země implikuje scénář, ve kterém roboti způsobily nejhorší dlouhodobé škody naprostým řízením Tří zákonů. Tedy ve výsledku zamezujíce lidem podnikat vynalézavé a rizikové chování. V květnu 2007 vydala vláda Jižní Koreji prohlášení, že následně ve stejném roce vydá „Sbor etického robotického kodexu". Ten by měl nastavit standardní normy pro uživatele i výrobce. Podle Park Hye-Young z ministerstva komunikace a sdělování, by měl kodex odkazovat i na podstatu Tří zákonů robotiky se snahou o vytvoření základní úrovně norem pro budoucí vývoj robotiky.

Futurista Hans Moravec navrhl zařazení Tří zákonů robotiky do „korporátního řízení" – Korporace řízené umělou inteligencí a vliv robotiky na manufakturu a výrobu, budou v blízké budoucnosti dle jeho představy běžnou záležitostí. V kontrastu s tím David Brin s povídkou Foundation´s Triumph (1999) navrhl, že podstata Tří zákonů může dekadentně postupně zastarávat. Roboti by obecně implikovali Zerothův zákon na racionální vyřazení prvního zákonu, a také by se roboti skryli před lidmi, tak aby nemusel být druhý zákon téměř nikdy uveden do chodu. Brin vyobrazil obavy R.Daneel Olivawa, jak by se v průběhu replikace robotů staly Tři zákony robotiky evolučním handicapem a přírodní výběr by tyto zákony smetl – opatrný základ Asimovova odvrácen procesem evoluční komputerizace (evolution computation). I když by se roboti nemohli vyvíjet změnou designu místo mutací, protože by design musel stále následovat podstatu Tří zákonů robotiky. Tím by bylo zajištěno převládání těchto zákonů. Místo toho by chyby při konstrukci a výrobě mohli funkčně nahradit biologické mutace.

V srpnu 2009 v rámci problému s IEEE Intelligent Systems, Robin Murphy a David Woods navrhli „Tři zákony zodpovědné robotiky", jako cestu na větší stimulaci diskuze o roli zodpovědnosti a autority při vytváření platformy robotů, ale také u velkých systémů, které je řídí. Navrhnuté zákony znějí následovně:

- člověk by neměl využívat robotická zařízení bez systému člověk-robot s nejvyššími legálními a profesionálními bezpečnostními a etickými standardy;
- roboti se musí lidem zodpovídat tak, jak přísluší jejich role;
- robot musí mít dostatečnou formu autonomie k zabezpečení vlastní existence do té doby, dokud tato forma ochrany zaručuje přesný přenos kontroly, která není v rozporu s prvním a druhým zákonem. [26]

David Wood řekl „Naše zákony jsou více realistické, a proto trochu více nudné" a také, že „Týkající se filosofie je jasně spjata s tím, že lidé dělají občas chyby, ale roboti budou lepší – perfektní verze nás samých". Chtěli napsat nové tři zákony, aby přivedli lidi k myšlenkám o vztazích typu člověk-robot více realisticky[27]. V říjnu 2013 navrhl Alan Winfield na setkání EUCog[28] upravených 5 zákonů, které byly publikovány s komentářem skupiny ESPRC[29]/AHRC z roku 2010.

Roboti jsou více účelovými nástroji. Roboti by neměli být navrhováni s primárním cílem zabíjet, nebo ubližovat lidem, kromě případů zajištění bezpečnosti státu.

Lidé a roboti jsou zodpovědnými osobami. Roboti by měli být navrhováni a řízeni prakticky vzhledem k existujícím zákonům, základními hodnotami a lidskými právy, včetně práva soukromí jedince.

Roboti jsou produktem. Měli by být navrhováni procesy zaručující jejich bezpečnost a využití bez větších rizik.

[26] MURPHY R., R., WOODS, D., D . *Beyond Asimov: The Three Laws of Responsible Robotics 2009* [online]. Ohio, 1541-1672/09/$26.00 © 2009 IEEE Published by the IEEE Computer Society. [cit. 2020-04-19] Dostupné z WWW: < https://www.researchgate.net/publication/224567023_Beyond_Asimov_The_Three_Laws_of_Responsible_Robotics>

[27] MURPHY R., R., WOODS, D., D . *Beyond Asimov: The Three Laws of Responsible Robotics 2009* [online]. Ohio, 1541-1672/09/$26.00 © 2009 IEEE Published by the IEEE Computer Society. [cit. 2020-04-19] Dostupné z WWW: < https://www.researchgate.net/publication/224567023_Beyond_Asimov_The_Three_Laws_of_Responsible_Robotics>

[28] *EUCog* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://www.eucognition.org/ >

[29] *ESPRC - Engineering and Physical Sciences Research Council* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://en.wikipedia.org/wiki/Engineering_and_Physical_Sciences_Research_Council>

Roboti jsou vyráběnými artefakty. Neměli by být navrhováni ve formě uvádějící uživatele v omyl a odkrývající zranitelnosti uživatelů. Robot by měl být svou podstatou předpověditelný a upřímný chováním k uživateli. [30]

U každého aktivního robota by měli být přiřazené lidské osoby zodpovědné za robota před zákonem.

V našem případě se ale spíše s robotem nesetkáme, jako spíše s umělou inteligencí, která bude fungovat jako hlavní mozek systému našeho zabezpečení.

3.3 Umělá inteligence v bezpečnostních systémech

Problematikou umělé inteligence v bezpečnostních systémech je její nutnost učení se. Ačkoliv můžeme provádět miliony simulací ve sterilních a generických prostředích, vždy je nutné naučit konkrétní vzorce chování uživatelů, a to až v samotném místě operativního působení systému.[31] Samotná problematika učení je realizovaná pomocí tzv. „deep learning[32]“ (volně přeloženo jako „hloubkové učení“), které stojí na konkrétních algoritmech pro učení strojů.

Ale co myslíme učením? Mitchell 1997 překládá definici – počítačový program se má učit ze zkušeností „E“ se vším respektem vůči skupině úkolů „T“ a mírou výkonu „P“. Pokud je výkon v úkolech „T“ stejný, jako je definován v „P“, zlepšuje se pomocí zkušeností „E“. Každý si umí představit velkou škálu zkušeností „E“, úkolů „T“ a míry výkonu „P“.[33]

V našem případě, je proměnná „E“ definována daty, které máme již nasbírané z generických prostředí, a odpovídá zkušenostem ze všech simulací. Proměnná „T“ je zastupujícím faktorem pro konkrétní typ úkonu a můžeme jej považovat za index. Zatímco proměnná „P“ definuje typ bezpečnostního opatření. [34]

Ve všech případech je nutné uvědomit si, že přístupů k učení UI je několik. Na straně jedné můžeme předpokládat, že nejrychlejší variantou je přímé zapsání kódu. Pomocí přímého zápisu můžeme manuálně určit, které funkce má systém vykonávat

[30] Winfield, Alan. (2018). Experiments in Artificial Theory of Mind: From Safety to Story-Telling. Frontiers in Robotics and AI. 5. 10.3389/frobt.2018.00075.
[31] SUTRISNO, I. (2016). *A comprehensive review on intelligent surveillance systems. Communications in Science and Technology*. 1. 10.21924/cst.1.1.2016.7.
[32] *What is Deep Learning?* [online]. [cit. 2020-04-19]Dostupné z WWW: < https://machinelearningmastery.com/what-is-deep-learning/ >.
[33] GOODFELLOW I., B. Y. *Deep learning* Massachusetts: Cambridge: The MIT Press, 2016, s. 97.
[34] GOODFELLOW I., B. Y. *Deep learning* Massachusetts: Cambridge: The MIT Press, 2016, s. 97.

nebo spouštět v konkrétních situacích. Definujeme tak okruh činností pomocí definice jednotlivých proměnných v kauzální souvislosti s jednotlivými událostmi. Oproti tomu metoda „deep learning“ nám propůjčuje možnost, naučit systém reagovat samostatně na základě vyhodnocení již nasbíraných dat.

Systém má možnost porovnávat data a zkušenosti ze všech simulací v reálném čase a vytvářet scénáře pro realizace a řešení konkrétních situací. Výhodou oproti přímému zápisu řešení je, že systém bude schopen vytvářet vzorce řešení, které budou aplikovatelné ve všech situacích a naučí se reagovat tak, aby řešení bylo poplatné aktuální situaci. Nevýhodou je samozřejmě časová investice, která je v případě „deep learningu“ značná. [35]

V základu je třeba si říci a určit, k čemu přesně bude naše umělá inteligence sloužit. Obecně můžeme říci, že bude součástí bezpečnostního systému, který bude ovládat. Nicméně variant a možností, kdy a jak bude fungovat její kontrola nad systémem, je velké množství. Můžeme se pohybovat v okruhu dohledu, kdy systém bude pouze sbírat data a vyhodnocovat je, nebo v okruhu plné kontroly, kdy systém bude zároveň určovat, jak se jednotlivé prvky zachovají. V historii robotizace a využití UI bylo zmapováno velké množství okruhů, ve kterých byla UI použita nebo minimálně testována. Úspěšnost v jednotlivých disciplínách byla vždy závislá na tom, do jaké míry byl okruh zpracován a jak velkou zásobu dat jsme měli k dispozici, jako „palivo“ pro umělou inteligenci. [36]Problémem se stává vyhodnocení, které vždy podléhalo subjektivnímu pohledu a cítění, navíc se silným dopadem na následný progres.

Existuje teorém o pokroku v umělé inteligenci:

- jakmile jsou naprogramovány některé mentální funkce, dříve nebo později je začnou lidé považovat za základní kámen "skutečného myšlení";
- neoddělitelné jádro inteligence se ale vždy skrývá v tom, co ještě nebylo naprogramováno.

Tento teorém byl poprvé představen Larrym Teslerem, a proto jej nazýváme "Teslerův Teorém":

[35] SUTRISNO, I. (2016). *A comprehensive review on intelligent surveillance systems. Communications in Science and Technology*. 1. 10.21924/cst.1.1.2016.7.
[36] GOODFELLOW I., B. Y. *Deep learning* Massachusetts: Cambridge: The MIT Press, 2016, s. 99

"Umělá inteligence je vše, co zatím nebylo naprogramováno[38]."

Selektivní přehled umělé inteligence je popsán níže. Ukazuje několik domén, na které se pracovníci zaměřili, každý z ních smýšlející vlastním směrem.

U některých z domén jsem zahrnula rozdělení podle použitých metod nebo konkrétnějších oblastí koncentrace. Přímý mechanický překlad (slovníkové vyhledávání s přeuspořádáním slov), nepřímý (prostřednictvím intermediárního interního jazyka), hraní šachů pomocí "brute force" metody s předstihem a heuristicky prořezávaným předsudkem bez dohledu jakýchkoliv validátorů. "Go kalah bridge" (sázení; hraní) pokerové variace na tic-tac-toe atd.

Umělá inteligence: Retrospektivka potvrzuje 598 teorémů v různých částech matematiky. Symbolická logika „rozlišovací“, elementární geometrie dokazující, elementární geometrie symbolická, manipulace matematických výrazů, symbolická integrace algebraické zjednodušení sumarizace, nekonečné řady vize.[39]

tiskařské matice: rozpoznávání jednotlivých ručně psaných znaků z malé třídy (např. číslic), čtení textu v různých fontech, čtení pasáží v rukopisu, čtení čínských nebo japonských tištěných znaků, čtení čínských nebo japonských ručně psaných znaků.

obrazové: lokalizace předdefinovaných objektů na fotografiích, rozklad scény na samostatné objekty, identifikace samostatných objektů ve scéně, rozpoznávání objektů zobrazených v náčrtech lidmi, rozpoznávání lidských tváří, poslouchání a porozumění mluveným slovům vytvořeným z omezeného slovníku (např. jména deseti číslic), porozumění souvislé řeči v pevných doménách, nalezení hranic mezi fonémy identifikující fonémy, nalezení hranic mezi morfémy identifikující morfémy, sestavující celé věty, porozumění přirozeným jazykům, odpovídání na otázky ve specifických doménách, parsující složité věty, vytvářející parafrázi delších textů s využitím znalostí skutečného světa, aby bylo možné porozumět pasážím, které řeší nejasné odkazy, vytvářející abstraktní poezii přirozeného jazyka (např. haiku) náhodné věty, odstavce, nebo delší kousky textu vytvářející výstup z vnitřní reprezentace znalostí.

Umělá inteligence: Retrospektivně odkazuje na 599 originálních myšlenek nebo uměleckých děl. Psaní poezie (haiku), psaní příběhů, počítačové umění, hudební

[38] GOODFELLOW I., B. Y. *Deep learning* Massachusetts: Cambridge: The MIT Press, 2016, s. 98
[39] BISHOP CH.M. *Pattern Recognition and Machine Learning. Cambridge* CB3 0FB, U.K, 2006. 758 s. ISBN 978-0387-31073-2.

kompozice, atonální a tonální analogické myšlení, geometrické tvary („testy inteligence“), konstruování důkazů v jedné z domén matematiky založené na těch v příbuzné doméně, učení úpravy parametrů koncepce tvorby). [40]

Pro většinu bezpečnostních systémů využívajících UI jsou použity neuronové sítě, a to hlavně z důvodu jejich integrace na biometrické skeny. Neuronové sítě jsou zcela unikátním druhem UI, která pracuje s vizí a principem elektronické integrace biologických procesů. Umělá neuronová síť je jeden z výpočetních modelů používaných v umělé inteligenci. Jejím vzorem je chování odpovídajících biologických struktur. Umělá neuronová síť je struktura určená pro distribuované paralelní zpracování dat. [41]

Skládá se z umělých (nebo také formálních) neuronů, jejichž předobrazem je biologický neuron. Neurony jsou vzájemně propojeny a navzájem si předávají signály a transformují je pomocí určitých přenosových funkcí. Neuron má libovolný počet vstupů, ale pouze jeden výstup. [42]

Neuronové sítě se používají mimo jiné i pro rozpoznávání a kompresi obrazů nebo zvuků, předvídání vývoje časových řad (např. burzovních indexů), někdy dokonce k filtrování spamu. V lékařství slouží k prohlubování znalostí o fungování nervových soustav živých organismů. Například perceptronová síť vznikla původně jako simulace fyziologického modelu rozpoznávání vzorů na sítnici lidského oka.

Učení u neuronové sítě je odlišné od učení „standardních“ umělých inteligenci. Zatímco „běžná“ UI se většinou učí formou **pokus X omyl**, kdy zkouší všechny možnosti obsažené v množině realizací a vyhodnocuje jejich úspěšnost v realizaci zadaného úkolu. Neuronová síť pracuje s předpokladem a vyhodnocuje výsledek na základě faktorů, které normální UI nehodnotí. Obecně lze říci, že existují „figury“, nebo „modely“ pro učení neuronové sítě. [43]

V poslední době se věnuje větší pozornost konekcionistickému přístupu k budování inteligentních strojů se strukturovanými modely, jako jsou umělé neuronové sítě (ANN[44]). Konekcionostické modely jsou založeny na tom, jak výpočet probíhá

[40] GOODFELLOW I., B. Y. *Deep learning* Massachusetts: Cambridge: The MIT Press, 2016, s. 103
[41] BISHOP CH.M. *Pattern Recognition and Machine Learning. Cambridge* CB3 0FB, U.K, 2006. s 493.
[42] FLOREANO F., MATTIUSSI C. *Bio-inspired artificial intelligence, theories, methods and technologies. Cambridge*, MA, USA: The MIT Press 2008
[43] D.R., H. *Gödel, Escher, Bach: an Eternal Golden Braid.* New York: Basic Books, cop., 1999, s. 133.
[44] *Artificial Neural Network (ANN)* [online]. [cit. 2020-04-19]
Dostupné z WWW: < https://www.investopedia.com/terms/a/artificial-neural-networks-ann.asp>

v biologických neuronových sítích. Spojení hrají zásadní roli v modelech konekcionistů, a proto používáme název „Konekcionismus“. Pojem konekcionismus zavedl Donald Hebb ve 40. letech 20. století a jedná se o soubor přístupů v oblasti umělé inteligence, který modeluje mentální nebo behaviorální jevy jako vznikající procesy vzájemně propojených sítí jednoduchých jednotek. Ústředním konekcionistickým principem je, že mentální jevy lze popsat propojenými sítěmi jednoduchých a jednotných jednotek.

Obrázek 7: Jednotka

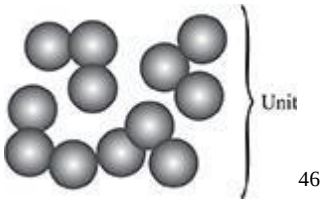
[45]

Obrázek 7.

Jednotka: základní struktura zpracování informací konekcionistického modelu.

Jednotky jsou podle modelu spojitosti to, co jsou neurony pro biologickou neuronovou síť: základní struktury zpracování informací. Protože k toku informací v síti dochází prostřednictvím jejích připojení, je propojení, přes které informace proudí od jednoho člena sítě k dalšímu, známé jako synapse. Synapse jsou pro neuronové sítě to, co je ethernetový kabel nebo telefonní drát pro počítačové sítě. Bez synapsí z jiných neuronů by nebylo možné, aby neuron přijímal vstup a odeslal výstup do jiných neuronů. Vzhledem k zásadní roli, kterou připojení hrají v síti neuronů, záleží na synapsích v biologické neuronové síti stejně, jako na samotných neuronech.

Obrázek 8: Propojovací model s 12 jednotkami



[46]

Obrázek 8.

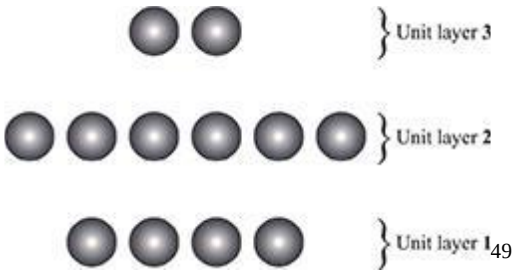
Většina konekcionistických modelů jsou počítačové simulace prováděné na digitálních počítačích. V počítačovém modelu s konektorem jsou jednotky obvykle reprezentovány kruhy, jak je znázorněno na obrázku 1. Protože žádná jednotka sama

[45] ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016 s.51
[46] ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016 s. 52

o sobě netvoří síť, jsou modely s konektory obvykle složeny z mnoha jednotek, jak je znázorněno na obrázku. [47]

Neuronové sítě jsou však uspořádány ve vrstvách neuronů. Z tohoto důvodu jsou konekcionistické modely uspořádány do vrstev jednotek, jak je znázorněno na obrázku 3. Obrázek 3 stále není sítí, protože žádná skupina objektů není kvalifikována jako síť, pokud není každý člen připojen k jiným členům; je to existence připojení, která vytvářejí síť, jak je znázorněno na obrázku [48]9.

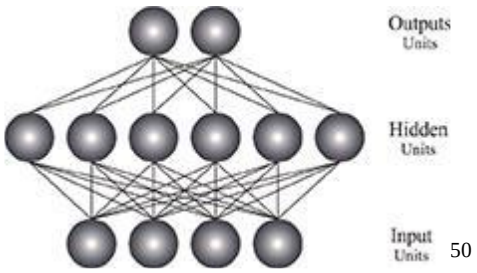
Obrázek 9: Síť jednotek ve vrstvách systému



[49]

Obrázek 9.

Obrázek 10: Model sítě

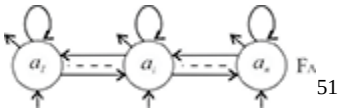


[50]

Obrázek 10

[47] ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016, s. 188.
48 LIPPMANN R. *An introduction to computing with neural nets*. IEEE ASSP Magazine. 1987. s. 4-22.
[49] ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016 s. 52
[50] ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016 s. 52

Obrázek 11: Jednovrstvá opakující se síť s postranní strukturou zpětné vazby.

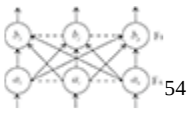
 [51]

Obrázek 11.

Na obrázku 9 je vidět, že síťová připojení jsou kanály, kterými proudí informace mezi členy sítě. Pokud taková připojení neexistují, žádná skupina objektů není kvalifikována jako síť. Existují dva druhy síťových připojení: vstup a výstup. Vstupní spojení je kanál, kterým člen sítě přijímá informace. Výstupní spojení je kanál, kterým člen sítě odesílá informace. Ačkoli je možné, aby síťové připojení bylo jak vstupním, tak i výstupním připojením, jednotka se nekvalifikuje jako člen sítě, pokud nemůže přijímat informace od jiných jednotek, ani odesílat informace jiným jednotkám. [52]

Existuje mnoho forem konekcionismu, ale nejčastější formy používají modely neuronových sítí.[53] Forma spojení a jednotek se může lišit od modelu k modelu, jak je znázorněno na obrázcích 8-10, kde je vidět, že v každé vrstvě může existovat libovolný počet jednotek a každá jednotka každé vrstvy je obvykle spojena váženým připojení ke každému uzlu další vrstvy. Data jsou do sítě dodávána prostřednictvím vstupní vrstvy.

Obrázek 12: Dvouvrstvá struktura posuvu vpřed.

 [54]

Obrázek 12.

Jednovrstvá opakující se síť s postranní strukturou zpětné vazby:

V závislosti na povaze problémů jsou modely neuronových sítí organizovány v různých strukturálních uspořádáních (architektury nebo topologie). Architektura

[51] ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016 s. 53
[52] ARBIB M.A. *Brain theory and neural networks*. Cambridge, MA, USA, 2003,.
53 ZUPAN J. *Introduction to artificial neural network methods: what they are and how to use them.* Acta Chimica Slovenica, 1994, s. 327–352.
[54] ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016 s. 53

neuronové sítě definuje její strukturu včetně počtu skrytých vrstev, počtu skrytých uzlů a počtu uzlů ve vstupní a výstupní vrstvě. Existuje několik typů architektur ANN. Jak je znázorněno na obrázcích 11-15. Většinu široce používaných modelů neuronových sítí lze rozdělit do dvou hlavních kategorií: dopředné neuronové sítě (FFNN) a zpětnovazební neuronové sítě (FBNN[55]).[56]

Obrázek 13: Dvouvrstvá struktura zpětné vazby


[57]

Obrázek 13.

Dvouvrstvá struktura posuvu vpřed:

Jak je znázorněno na obrázcích 12 a 14, FFNN umožňují signálům cestovat pouze jednou cestou; data vstupují na vstupy a procházejí sítí, vrstvu po vrstvě, dokud nedorazí na výstup. Mezi vrstvami není zpětná vazba ani smyčky. Tyto sítě se široce používají při rozpoznávání a klasifikaci vzorů. FBNN může mít signály pohybující se v obou směrech zavedením smyček v síti, jak je znázorněno na obr. 12, 13 a 15. FBNN jsou dynamické; jejich stav se neustále mění, dokud nedosáhnou rovnovážného bodu. Zůstávají v rovnovážném bodě, dokud se nezmění vstup a dokud není potřeba najít novou rovnováhu.[58]

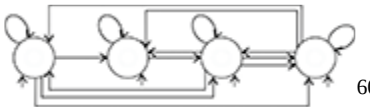
Obrázek 14: Třívrstvá struktura posunu vpřed


[59]

[55] *feed-forward-back-propagation neural network* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://www.researchgate.net/figure/Structure-of-the-feed-forward-back-propagation-neural-network-FBNN_fig1_336138883>
[56] JAIN A.K., MAO J., MOHIUDDIN K.M. *Artificial neural networks: a tutorial*. IEEE: Computer, 1996, s. 31–44.
[57] ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016 s. 54
58 FLOREAN F., MATTIUSSI C. *Bio-inspired artificial intelligence, theories, methods and technologies*. Cambridge, MA, USA: The MIT Press, 2008,
[59] ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016, s. 192.

Obrázek 14.

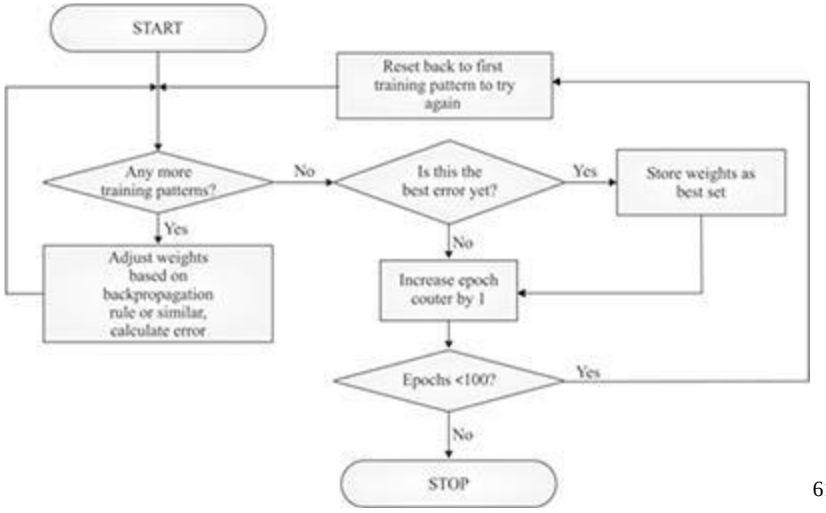
Obrázek 15: Jednovrstvá opakující se struktura



60

Obrázek 15.

Obrázek 16: Školení ANN



61

Obrázek 16.

Jednovrstvá opakující se struktura:

Ve většině modelech s konektory jsou jednotky uspořádány do tří vrstev: vstupní vrstva, jedna nebo více „skrytých“ vrstev a výstupní vrstva. Obrázky 9 a 10 znázorňují třívrstvou FFNN sestávající z 3 vrstev jednotek, kde každá jednotka je připojena ke každé jednotce nad ní a kde informace proudí „dopředu“ ze vstupních jednotek sítě, přes její „skryté“ jednotky, do svých jednotek výstupní jednotky. Uzly zpracovaných vstupních dat skryté vrstvy přijímají jako součet vážených výstupů vstupní vrstvy.

60 ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016, s. 192.
61 ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016, s. 193.

Uzly procesních vstupních dat výstupní vrstvy přijímají jako součet váženého výstupu jednotek uvnitř skrytých vrstev a dodávají výstup systému.[62]

Školení ANN:

Jak již bylo zmíněno dříve, mohou být principy učení aplikovány na stroje pro zlepšení jejich výkonu. Ve FFNN je učení sítě velmi důležitým procesem. Vzdělávací situace může být rozdělena do dvou hlavních kategorií: pod dohledem a bez dozoru. S učením pod dohledem musí být ANN učen, než se stane užitečným. Školení spočívá v prezentaci vstupních a výstupních dat do sítě. Obrázek 10 ukazuje rozlišující povahu dohlížející neuronové sítě, která zahrnuje externí trenér, u kterého jsou známy vstupy a výstupy a jeho cílem je objevit vztah mezi nimi. V tomto režimu je skutečný výstup ANN porovnáván s požadovaným výstupem. [63]

Důležitým problémem, který se týká učení pod dohledem, je problém konvergence chyb: minimalizace chyby mezi požadovanými a vypočítanými hodnotami. Výkon sítě je vyhodnocen na základě srovnání mezi vypočítaným (předpovězeným) výstupem a skutečnou (požadovanou) výstupní hodnotou. Existuje několik typů měření přesnosti predikce; nejběžnější použitá měření jsou následující:

Koeficient stanovení (R2)

R2 = Σni = 1 (Yˆi − Y ¯ ¯) 2Σni = 1 (Yi − Y ¯ ¯) 2E1

Střední chyba čtverce (MSE)

MSE = 1nΣni = 0 (Yi-Yˆi) 2E2

Root Mean Square Error (RMSE)

RMSE = [1nΣni = 0 (Yi-Yˆi) 2] 12 / E3

Střední chyba absolutního procenta (MAPE)

[62] ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016, s. 206.

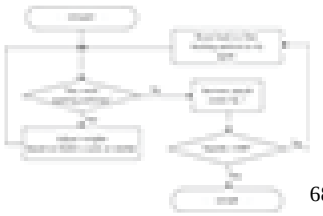
[63] ZUPAN J. *Introduction to artificial neural network methods: what they are and how to use them.* Acta Chimica Slovenica. 1994;41(3):327–352.

MAPE% = 1nΣni = 0|||Yi-Yˆ iYi||| × 100 (4) E4

kde Yi je skutečná hodnota výstupu, Yˆ i je predikovaná hodnota a (n) je počet pozorování. [64]

Na rozdíl od učení pod dohledem využívá neuronová síť bez dozoru externí zpětnou vazbu a je založena pouze na místních informacích. Jak je vidět na obrázku 16, v učení bez dozoru je znám pouze vstup a cílem je odhalit vzory ve vlastnostech vstupních dat. Rovněž se označuje jako samoorganizace v tom smyslu, že samoorganizuje data prezentovaná do sítě a zjišťuje jejich vznikající kolektivní vlastnosti. Cílem sledování bez dozoru je, aby se počítač naučil, jak dělat něco, co mu neřekneme, jak to udělat. [65]

Mezi běžné aplikace učení bez dozoru patří klasifikace, dolování dat a samoorganizující se mapy (SOM[66]), také nazývané neuronová síť Kohonen (KNN[67]).



68

Obrázek 11.

Ze všech výše uvedených modelů učení vyplývá souvztažnost jednotlivých faktorů. Neuronová síť je společnost kvantovým počítačem nejpokročilejší formou umělé inteligence a výpočetní techniky, kterou lidstvo disponuje (v otázce kvantového počítače je termín „disponuje“ nadnesený, ačkoliv již zaznamenáváme obrovské pokroky).

Nerušené učení ANN.

V FFNN s supervidovaným výcvikem existují dva velmi odlišné typy neuronových sítí: FFNN vycvičený algoritmem Backpropagation (BP) (FFBPNN) a Statistical Neural Networks (SNN). FFBPNN používají rovnice, které jsou spojeny

---

[64] HUANG D.S. *Radial basis probabilistic neural networks: model and applications*. International Journal of Pattern Recognition and Artificial Intelligence, 1999, s. :1083– 1101.
[65] ZUPAN J. *Introduction to artificial neural network methods: what they are and how to use them.* Acta Chimica Slovenica. 1994;41(3):327–352.
[66] *Samoučící se neuronová síť - SOM, Kohonenovy mapy* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://www.kiv.zcu.cz/studies/predmety/uir/NS/Samouc_NN2.pdf>
[67] *Kohonen neural network* [online]. [cit. 2020-04-19] Dostupné z WWW: https://www.sciencedirect.com/science/article/abs/pii/S0003267096003157>
68 Rosa, J. L. Artificial Neural Networks - Models and Aplications InTech, 2016, s. 195

pomocí váhových faktorů. Výběr váhových faktorů činí tyto neuronové sítě velmi silnými. Vícevrstvý perceptron (MLP) je nejběžnější a nejúspěšnější architekturou neuronových sítí s topologiemi FFNN, zatímco nejčastější supervizovanou technikou učení používanou pro trénink umělých neuronových sítí je vícevrstvý backpropagation (BP) algoritmus. [69]

BP je systematická metoda pro výcvik vícevrstvého FFNN, jak je znázorněno na obrázku 8. Protože se jedná o dohlížecí algoritmus pro dohled, jsou uvedeny jak vstupní, tak cílové vzorce. Pro daný vstupní vzor je výstupní vektor odhadován přes dopředný průchod sítí. Po dokončení dopředného průchodu se odhadne chybový vektor ve výstupní vrstvě pomocí rozdílu komponent cílového vzoru a generovaného výstupního vektoru. Funkce chyb výstupních vrstevních uzlů je poté šířena zpět sítí přes každou vrstvu pro úpravu vah v této vrstvě. Zásady přizpůsobení hmotnosti v algoritmu BP jsou odvozeny na principu nejstrmějšího sestupového přístupu při hledání minima funkce s více hodnotami. [70]

BPFFNN se skládají z neuronů uspořádaných do jedné vstupní vrstvy a jedné výstupní vrstvy a několika skrytých vrstev neuronů, jak je znázorněno na obrázku. Neurony provádějí určitý druh výpočtu pomocí vstupů pro výpočet výstupu, který představuje systém. Výstupy jsou předány dalšímu neuronu. Okraj označuje, kterým neuronům je výstup poskytnut. Tyto oblouky nesou závaží. Obvykle se učení BP skládá ze dvou průchodů: vpřed a vzad. V dopředném průchodu je na senzorické uzly sítě aplikován vzor aktivity. Konečně je vytvářena sada výstupů jako skutečné reakce sítě. Během této cesty jsou pevné synaptické hmotnosti. Během zpětného průchodu jsou synaptické hmotnosti upraveny v souladu s pravidlem pro opravu chyb. [71]

BPFFNN mají žádoucí vlastnost, že jsou velmi flexibilní. Mohou být použity pro rozpoznávání vzorů i pro problémy s rozhodováním. Další výhodou je, že stejně jako u každé jiné neuronové sítě je proces vysoce paralelní, a proto je možné použití paralelních procesorů a zkracuje potřebný čas pro výpočty. BPNN však mají negativní [72]vlastnosti. Školení sítě může vyžadovat značné množství času. Velikost údajů o školení

[69] LIPPMANN R. *An introduction to computing with neural nets*. IEEE ASSP Magazine. 1987. s. 4-22.
[70] ZUPAN J. *Introduction to artificial neural network methods: what they are and how to use them.* Acta Chimica Slovenica. 1994;41(3):327–352.
71 HUANG D.S. *Radial basis probabilistic neural networks: model and applications*. International Journal of Pattern Recognition and Artificial Intelligence, 1999, s. :1083– 1101.
72 ROSA, J. L. *Artificial Neural Networks - Models and Aplications InTech*, 2016, s. 217

pro BPFFNN musí být velmi velká. V některých případech je téměř nemožné zajistit dostatek školení.

4 Rizika, klady a zápory jednotlivých druhů technologií

Pro určení nejvhodnějšího systému zabezpečení je třeba vždy porovnávat silné a slabé stránky dílčích technologií a porovnat je vůči očekávání, které uživatel má. Kritérií pro porovnání je mnoho:

- cena;

- náročnost instalace;

- stupeň překonatelnosti;

- náročnost ovládání systému;

- a mnoho dalších.

V rámci této práce ale posuzuji vše vůči rizikům užití. Jedním z hlavních kritérií a faktorů porovnání je tedy paradoxně bezpečnost bezpečnostního systému. Nikoliv však bezpečnost v otázce schopnosti zabezpečit objekt, ale v otázce bezpečnosti užití systému vůči uživatelům a jejich životům. Je vcelku jasné, že mechanické bezpečnostní systémy, jako například ploty, nebudou mít téměř žádné riziko použití. Majitel plotu jej těžko bude přelézat, nebo jen v minimálním množství případů, aby mohl říci, že se o něj poranil. Bezpečnost tedy řešíme pouze u systémů řízených pomocí UI. Mechanické a „tradiční“, nebo chceme-li spíše „standardní“, systémy, podrobujeme klasické sadě otázek, která je uvedena výše a hledáme tak odpověď na to, co je nejlepší poměr „kladů“ a „záporů“.

Obecné definice hovoří jasně ve všech případech a jasně determinují již známé klady a zápory bezpečnostních systémů. Na rozdělení se můžeme dívat pohledem otázky připojení:

4.1 Rozdělení dle připojení

Tento systém kategorizace je nasnadě jako hlavní, a to z toho důvodu, že zahrnuje pouze dvě, dalo by se říci „samopopisné“, kategorie:

Drátové – veškeré komponenty připojené v EZS jsou primárně propojeny pomocí pevné kabelové infrastruktury. Jako každý systém má i tento systém propojení své výhody a nevýhody.

Výhody

- stabilita;

- přenosové rychlosti nijak nekolísají;

- v případě výběru správné kabeláže, nedochází k žádnému rušení;

- systém je permanentně připojen.

Nevýhody

- nutnost rozvodu kabelů při stavbě;

- systém musíte mít rozmyšlen dopředu, protože po realizaci je velmi těžké jej přesouvat;

- prakticky nulová flexibilita.

Bezdrátové – standardně jsou zásadní části EZP připojeny pouze k elektrické síti a zbytek je řešen datovým přenosem přes Wi-Fi nebo Bluetooth. Jedná se o řešení, jehož obliba narůstá čím dál více. I když jsme stále ve fázi, kdy může být v některých oblastech stabilita velmi diskutabilní.

Výhody

- přenositelnost;

- lehce integrovatelné do již postavených objektů;

- lze libovolně doplnit nebo odstranit jakékoliv prvky;

- již ze základu je ovladatelné na dálku pomocí mobilního telefonu;

- nenápadnost;

- nemožnost zničení celého systému pomocí přestřihnutí jednoho kabelu.

Nevýhody

- náchylnost na vnější rušení;

- jakékoliv bezdrátové systémy jsou obětí kolísání přenosové rychlosti;

- u čidel je nutné hlídat stav baterie.

Další již zmíněnou tradiční metodou porovnání výhod a nevýhod jednotlivých systémů je pomocí určení překonatelnosti bezpečnostního systému.

RIZIKO	ZNALOSTI A VYBAVENÍ NARUŠITELŮ	STUPEŇ ZABEZPEČENÍ
Nízká	Předpokládá se, že narušitelé mají malou znalost EZS a že mají k dispozici omezený sortiment snadno dostupných nástrojů	1
Průměrná	Předpokládá se, že narušitelé mají určité znalosti o EZS a že použijí základní sortiment nástrojů a přenosných přístrojů.	2
Vysoká	Předpokládá se, že narušitelé jsou obeznámeni s EZS a mají úplný sortiment nástrojů a přenosných elektrických zařízení.	3
Nejvyšší	Používá se tehdy, když zabezpečení má priority před všemi ostatními	4[73]

[73] National Academies of Sciences, E. a.. *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*. Washington, DC 2006

	hledisky. Předpokládá se, že narušitelé mají možnost zpracovat podrobný plán vniknutí a mají kompletní sortiment zařízení včetně prostředků pro náhradu rozhodujících prvků EZS.	

Obecné rozdělení na mechanické a elektronické bezpečnostní systémy již samo o sobě napovídá, v čem může být výhoda a nevýhoda jednotlivých systémů ve zmíněných kategoriích. Záleží ale hodně na tom, který systém z dané skupiny využijeme a v jaké souvztažnosti k okolním systémům.

4.2 Výhody a nevýhody mechanických zabezpečovacích systémů

Jak se s oblibou říká, každá mince, má dvě strany. Je to všeobecné pravidlo, které lze plošně uplatnit na vše, co nás obklopuje. Obecně je složité u této kategorie určit výhody a nevýhody, protože mnohdy záleží na typu použitého materiálu, na dodržení přesného postupu montáže a tak dále. Nicméně pokud se nám povede spojit paradoxní dělení na obecnější charakter s jasně definovaným detailem, tak se můžeme dobrat přesnějších výsledků.

Výhody:

Snadno dostupné – většina MZS [74]jsou snadno dostupné v jakémkoliv hobby marketu. Visací zámek, vložku, petlice, závory a další z těchto prvků (včetně trezorů, sejfů a dalších) můžete koupit klidně v Neděli dopoledne v Hornbachu.

Nízká cena – většina základních MZS prvků, které samozřejmě nespadají do nejvyšší bezpečnostní třídy, jsou levné. Zcela logicky lze pořídit zámkovou vložku za 700 kč včetně DPH a stejně tak za 9000 kč včetně DPH.

Relativně snadná montáž – Vložku zámku si může vyměnit každý, použití visacího zámku je taktéž všem známo. Zabudování mříže do stavebního otvoru již může být složitější, ale i tak jej spíše provedeme dle YouTube, než instalaci komplexního EZS.

[74] *Mechanické zábranné systémy (MZS)* [online]. [cit. 2020-04-19] Dostupné z WWW: https://www.security.cz/mechanicke-zabranne-systemy-mzs--2422.html

Hluk – překonání některých MZS prvků se neobejde bez hrubé síly nebo hlučné techniky. Například mříže jen velmi těžko přeříznete ruční pilkou na železo během pár vteřin. Na druhou stranu úhlová bruska s řezným kotoučem potřebuje elektřinu a ještě je dost hlasitá.

Nevýhody:

Překonatelnost – je to sice relativním bodem, ale pakliže vezmu v potaz všeobecný trend nákupu „co nejlevnějších věcí", tak je jeho relevantnost dosti opodstatněná. Valná většina levných MZS prvků spadá do nejnižší bezpečnostní třídy a lze je velmi snadno překonat.

Doba instalace – montáž vložky do dveří není totéž, co instalace mříže do stavebního otvoru pro okno, nebo stavba zdi.

Žádná signalizace – MZS samo o sobě nemá žádný systém upozornění na to, že nastal nějaký problém. Samozřejmě existují MZS, například zámky, které již mají i elektronický systém. Z pravidla je ale nutné MZS kombinovat s EZS pokud chceme dostat například informaci, že se nám někdo snaží vypáčit dveře nebo vyháčkovat zámek.

4.3 Nebezpečí plynoucí z užívání mechanického zabezpečovacího systému vlastní výroby

Každá doba přináší svá specifika a vždy ovlivňuje myšlení lidí. Jedná se o určitý imperativ, normativní vzorec v jednání lidí, nebo sociálních skupin. Pokud k tomuto přidáme ještě, u nás tak pověstnou, lidovou tvořivost, dostaneme zajímavý „elixír".

Až neuvěřitelně velké množství lidí se uchyluje k zabezpečení nemovitostí (primárně se jedná o tzv. „chataře") pomocí vyrobených zabezpečovacích prvků. Extrémní případy jsou pak zakoupené prvky, které primárně nemají sloužit k ochraně majetku nebo osob a jsou tak nesprávně použity. Ano, mluvím zde o „pastech", které lidé využívají k zabezpečení majetku.

Nejedná se o sci-fi nebo lidové pověsti, ale o skutečné události, které nastali v průběhu času. Od nalíčených ocelových lanek tenkého průměru v úrovni kotníků, nezřídka i v úrovni krku, až po oka na medvědy pod rohožkou za dveřmi. Velmi často se objevují následující „kutilské" zázraky:

Elektřina – elektrické ohradníky skrytě natažené v oblasti plotů, nebo dokonce elektřina, různého napětí a proudu přivedená přímo na kování (kliky branky).

Pasti – berme v potaz, že pastí se myslí jakékoliv nastražené zařízení, které má za úkol „zneškodnění" narušitele. Můžou to být natažená ocelová lanka, uvolněná prkna, hřebíky pod rohožkou, žiletky v místech běžného dotyku rukou až po bizarní prvky, jako pasti na medvědy, nebo samostříly s kladkou přivedenou na kování dveří.

Všechny tyto věci mají jedno společné, a to konkrétně nebezpečí z nich plynoucí. Nemluvím zde o nebezpečí úrazu, ale o legislativní složce věci. Jakákoliv újma na zdraví, která bude způsobena narušiteli objektu v důsledku střetu s takovouto pastí, bude předmětem šetření a může se otočit proti majiteli objektu. Jedná se o obecné ohrožení a v extrémních případech může dojít k zabití narušitele. Přesné právní vyjádření a pojmenování věci nechám povolanějším, ale myslím si, že princip a sdělení je zcela jasné. Jakákoliv po domácku sestrojená zařízení porušují zákony České republiky a jejich použití se může drasticky otočit proti majiteli." [75]

U mechanických systémů je v celku zřejmé, že hlavním problém může být špatné užívání, které ohrožuje přímo majitele nebo okolí. V extrémních případech narážíme na onu, již zmíněnou, „lidovou tvořivost", která může v určité formě naplňovat i skutkovou podstatu trestného činu.

Inteligentní bezpečnostní systémy, které obsahují, nebo jsou přímo řízeny UI, spadají samozřejmě do jiné kategorie, která má vlastní rizika a problémy. V případě, že bychom vzali v potaz dělení UI systémů na další podmnožiny (neuronové sítě, obsahující UI, přímo řízeny UI, s připojením na UI rozhodovací systém, atd), tak dostaneme mnohem širší paletu výhod a nevýhod. V obecné rovině ale platí následující:

„Inteligentní domácnosti jsou rychle se šířícím a velmi módním trendem. Což je s ohledem na to, kolik nevýhod mají, dosti zvláštní. Stále se jedná o nové technologie, které z velké části nejsou ještě zcela odladěny. Komfort, který přináší, je neuvěřitelný, ale ta daň je většinou dost vysoká a projevuje se až později.

[75] National Academies of Sciences, E. a.. *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*. Washington, DC 2006

Výhody:

Komfort – funkce, které výrazně ulehčují život a zpříjemňují každodenní život. Ovládání na dálku, ovládání přes internet, synchronizace napříč zařízeními. To vše a ještě mnohem více jsou klíčové faktory, které přinášejí vyšší standard komfortu.

Rostoucí trend – ačkoliv to může znít zvláštně, tak stát na počátku má i své výhody. Díky tomu, že se jedná o relativně mladé technologie, tak jsou do nich značné investice. Což znamená, že kritické chyby, jsou často velmi rychle opraveny.

Příprava – z hlediska budoucnosti se jedná o skvělou příležitost, jak se připravit na to, že jednou bude tento „High-end[76]“ naprostým standardem.

Bezpečnostní funkce – Smart Home[77] poskytuje širší paletu bezpečnostních funkcí než klasické EZS nebo MZS. Kombinace je ve všech ohledech žádoucí.

Nevýhody:

Dostupnost – zatím se jedná o celkem dost drahé řešení a většina skutečně inteligentních prvků má násobně vyšší cenu, než standardní EZS prvky. V případě Neural AI Smart Home je cenovka doslova astronomická (řádově stovky milionů korun).

Instalace a správa – Smart home prvky není jednoduché nainstalovat a je třeba více odborníků, nebo společnost, která je sdružuje pod jednu střechu. Samotná montáž není tak složitá. Ono umístění prvku do prostoru nebo upevnění na zeď není zas tak zásadní téma. Správné síťové propojení, nastavení komunikace v rámci jedné sítě a zabezpečení proti útoku zvenčí je již složitější.

Náchylnost – žijeme v době, kdy je kyberzločin na vzestupu a útok z internetu je reálnější, než chytit chřipku. Problémem v tomto ohledu je, že vše, co je připojeno k internetu, je potenciálním cílem kyberútoku. Pokud by útočník systém napadl, může jej celý vypnout nebo jej v horším případě zneužívat proti samotnému majiteli.

Nebezpečí selhání – selhání systému může mít fatální následky. Již se objevilo několik případů, kdy Smart Home zavinil poškození majetku, nebo dokonce smrt majitele. V případě Neural AI Smart Home je to riziko mnohem vyšší, protože se jedná o logickou

[76] *high end* [online]. [cit. 2020-04-19] Dostupné z WWW: https://www.wordreference.com/encz/high-end
[77] *Smart home* [online]. [cit. 2020-04-19] Dostupné z WWW: https://www.inels.cz/smarthome

jednotku, která má schopnost (respektive by měla být schopna) se učit. Ačkoliv máme definovány Asimovovi zákony robotiky, tak to neznamená, že se tím neuronové sítě řídí. Například Elon Musk[78] a společnost Tesla[79], již pochopili, že Asimovovi zákony robotiky, jsou nedílnou součástí vývoje a robotizace. Bohužel je problémem, že tyto zákony máme my, ale jakákoliv programovatelnost, není z naší strany explicitně možná.

Jak již vyplývá z informací, které jsem uvedla výše, tak nevýhody, i z hlediska dopadu, převyšují výhody. Pakliže se zaměříme čistě na Neural AI Smart Home, který mě osobně přijde nejzajímavější, tak se jedná o velmi nebezpečnou věc s přehnaně šlechetnou nálepkou.[80]

Toliko k obecným definicím a známým, již popsaným kladům a záporům. Obecně platná pravidla, jak již bylo zmíněno výše, jsou ale spíše určující pro standardní systémy. V našem případě známe celkem pět rizik užití umělé inteligence v bezpečnostních systémech, která byla definována ve vědecké publikaci.

4.4 Potenciální cesty k útoku

Papernot nezačal s příkladem modelu jednoduchého strojového učení (ML[81]), který byl navržen tak, aby předpovídal, zda mají pacienti diabetes nebo anorexii, nebo zda jsou zdraví, za použití lékařských záznamů jako vstupů. Koncepčně pokud jsou testovací body kresleny ze stejné distribuce, jako tréninková data ML systém bude fungovat a bude provádět správné předpovědi. V reálném světě však mohou vzniknout mnohem nejednoznačnější příklady, takže systém ML vytváří nízkou důvěru nebo nesprávnou předpověď. Je také pravděpodobné, že velká část vstupní domény nebyla modelována, což by vedlo k tomu, že některé vstupní dotazy povedou k náhodnému výstupu. Zkušený protivník by mohl oklamat ML, aby poskytl špatný výstup pečlivým vytvořením narušení toho, co se jinak zdá být legitimním vstupem.[82]

Papernot vysvětlil, že tento typ útoku může být vytvořen pomocí procesu podobného tomu, který se používá k trénování ML systému: to znamená namísto výpočtu derivátů chyby systému s ohledem na parametry výcviku (jako by se optimalizovalo) modelu,

[78] *Elon Musk* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://www.tesla.com/cs_CZ/elon-musk >.
[79] *Tesla* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://www.tesla.com/cs_CZ/about >.
[80] National Academies of Sciences, E. a.. *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*. Washington, DC 2006
[81] *Machine learning* [online]. [cit. 2020-04-19] Dostupné z WWW: https://en.wikipedia.org/wiki/Machine_learning
82 National Academies of Sciences, E. a.. *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*. Washington, DC. 2006, s. 44-45.

lze namísto toho vypočítat deriváty chyby systému s ohledem na samotný vstup. Tímto způsobem může protivník nebo výzkumný pracovník systematicky hledat poruchy pro jakýkoli vstup, aby oklamal jakýkoli model, na který se chtějí zaměřit, řekl Papernot. Obecná technika může fungovat proti ML v jakémkoli typu aplikace - například při rozpoznávání obrazu, přepisování zvuku nebo detekci malwaru.[83]

Útočníci se mohou také snažit ohrozit důvěrnost nebo soukromí údajů o školení. Toho lze dosáhnout pečlivým pozorováním toho, jak se předpovědi modelu liší pro různé vstupy. Pokud byl model nadřazen do jeho souboru údajů o tréninku, bude velmi citlivý na vstupy podobné odlehlým hodnotám od začátku sady - tyto body lze odvodit pečlivým testováním. Tento druh útoku je znám jako útok inference.

Papernot zdůraznil, že tyto příklady jsou jen dvěma z mnoha způsobů, jak mohou protivníci zacílit na systémy ML a zdůraznili, že existují příležitosti k útoku na každém kroku v potrubí ML. Protivník by mohl otrávit tréninková data, vyvodit závěry o důvěrných tréninkových datech na základě znalostí modelu a jeho parametrů, extrahovat model samotný pozorováním jeho předpovědí pro různé vstupy, nebo se naučit, jak narušit vstup pro trik systému.[84]

4.5 Uvedení rizik do kontextu

Vzhledem k tomu, že bezpečnost a ochrana osobních údajů v ML je hlavním problémem, Papernot položil otázku: Liší se bezpečnost a soukromí ML systémů od toho, co je vidět v tradiční počítačové bezpečnosti nebo dokonce v reálném světě? Ve všech případech řekl, že zabezpečení a soukromí jsou obtížné a rychlejší CPU a internet pravděpodobně tuto výzvu znesnadnily, než usnadnily. Zmínil se o charakteristice Butlera Lampsona, že praktická bezpečnost vyvažuje náklady na ochranu a riziko ztráty, což jsou náklady na zotavení ze ztráty a doby její pravděpodobnosti. V tomto světle lze ML považovat pouze za jiný způsob analýzy dat - ta, která zavádí nové útočné povrchy, které lze využít, což může vést ke zbrojní rase.

Papernot má však optimistický pohled na potenciální budoucí zabezpečení ML. Domnívá se, že systémy ML jsou dostatečně odlišné od tradičních počítačových systémů,

[83] PAPERNOT, N. (2018). *A Marauder's Map of Security and Privacy in Machine Learning: An overview of current and future research directions for making machine learning secure and private*. AISec '18: Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security. 1-1. 10.1145/3270101.3270102.
[84] PAPERNOT, N., McDANIEL, P., SINHA, A., WELLMAN, M. (2018). *SoK: Security and Privacy in Machine Learning*. 399-414. 10.1109/EuroSP.2018.00035.

takže je lze navrhnout systematickým a zásadovým přístupem k zabezpečení a soukromí. Důvodem je, že ML, podobně jako kryptografie, lze z velké části vyjádřit matematickou formou. Poznamenal, že v oblasti kryptografie nebylo dosaženo pokroku, dokud nebyla formálně specifikována interakce mezi protivníky a obránci, což naznačuje, že taková příležitost by mohla existovat i pro ML.[85]

4.6 Bezpečnostní požadavky a přístupy

Papernot tvrdil, že je třeba vyvinout úsilí ke stanovení ML bezpečnosti a zásad ochrany soukromí. Vědci tvrdí, že musí najít tu správnou abstrakci nebo jazyk, aby formalizovali požadavky na zabezpečení a soukromí ML s přesnou sémantikou a nejednoznačností. Jako užitečný model poukázal na referát Saltzera a Schroedera z roku 1975, který nastiňuje 10 principů souvisejících s ochranou informací v počítačových systémech.

Papernot řekl, že všechny tyto principy se přímo týkají jeho současného výzkumu v zabezpečených ML systémech. Dále uvedl konkrétní příklady tří z osmi principů.[86]

4.7 Zásady ochrany informací v počítačových systémech

Saltzer a Schroeder identifikovali následující zásady ochrany informací v počítačových systémech:

1. **Ekonomika mechanismu** - udržujte konstrukci bezpečnostních mechanismů jednoduchou;
2. **Výchozí nastavení bezpečné při selhání** - rozhodnutí o přístupu založte spíše na povolení než vyloučení;
3. **Kompletní zprostředkování** - každý přístup k objektu je kontrolován z hlediska oprávnění;
4. **Otevřený design** - konstrukce bezpečnostních mechanismů by neměla být tajná;
5. **Oddělení oprávnění** - ochranný mechanismus, který vyžaduje odemčení dvou klíčů je robustnější a pružnější;
6. **Nejméně oprávnění** - každý uživatel pracuje s nejmenším potřebným počtem oprávnění;

[85] National Academies of Sciences, E. a.. *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*. Washington, DC. 2006, s. 45-46.
86 National Academies of Sciences, E. a.. *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*. Washington, DC. 2006, s. 46-47.

7. **Nejméně běžný mechanismus** - minimalizujte mechanismy závislé na všech uživatelích;
8. **Psychologická přijatelnost** - lidské rozhraní navržené pro snadné použití;
9. **Pracovní faktor** - rovnováha nákladů na obcházení mechanismu se známými prostředky útočníka;
10. **Kompromisní záznam** - mechanismy, které spolehlivě zaznamenávají kompromisy, lze použít místo mechanismů, které zabraňují ztrátě. Nejprve se zabýval zásadou psychologické přijatelnosti, která vyžaduje rozhraní a systémy, kterým lidé mohou rozumět. Papernot zdůraznil, že lidé nebudou používat obranný nástroj, kterému nerozumí. Do této zásady se zapojil v kontextu soukromí. Protože každý má svou vlastní představu o tom, co soukromí znamená, výzkumní pracovníci v oblasti bezpečnosti se spojili s definicí zvanou diferenciální soukromí, která odkazuje na matematickou techniku používanou k maximalizaci přesnosti dotazů z databází při minimalizaci dopadu na soukromí těchto dat.[87]

Algoritmy rozdílové ochrany soukromí byly vyvinuty, aby protivníkovi znemožnily říci, jaká data, z nichž jednotlivci byli zahrnuti do tréninkové sady, takže protivník se nemůže dozvědět nic o jednotlivcích ani žádné informace o údajích, které přispěli. Nejstandardnějším algoritmem pro výcvik ML algoritmů je stochastický gradient klesání, který bere dávku dat, vypočítává chybu, vypočítává gradienty chyby ve vztahu k parametrům modelu a aplikuje gradienty pro aktualizaci parametrů modelu. Může být odlišně soukromý tím, že ořeže přechody a zašumí ořezané přechody před tím, než jsou použity pro aktualizaci parametrů modelu. Přestože se tento proces může zdát výzkumníkům jednoduchý, ti, kteří neznají rozdílné soukromí, tomu nerozumí. Papernot řekl; v důsledku toho tento proces nedosahuje psychologické přijatelnosti. Z tohoto důvodu vyvinul tým Papernot odlišný přístup nazvaný PATE - soukromá agregace souborů učitelů[88]. S PATE může uživatel chránit citlivá data v sadě dat rozdělením dat na oddíly, kde jediným požadavkem je, že jakýkoli tréninkový bod bude součástí pouze jednoho oddílu. Jeden ML model (nazývaný „učitel“) je trénován na každou podmnožinu dat, takže výsledkem je řada modelů trénovaných nezávisle na řešení

[87] SMITH, R. (2012). *A Contemporary Look at Saltzer and Schroeder's 1975 Design Principles. Security & Privacy*, IEEE. 10. 20-25. 10.1109/MSP.2012.85.
[88] PAPERNOT N., SONG S., MIRONOV I., RAGHUNATHAN A., TALWAR K., ERLINGSSON Ú., 2018*, "Scalable Private Learning with PATE. Sixth International Conference on Learning Representations* (ICLR 2018) [online]. [cit. 2020-04-19]

stejného úkolu pomocí různých podmnožin dat. Každý model získá „hlas“ na správném štítku pro daný vstup.

V zájmu zachování soukromí uživatel systému jednoduše požádá každého učitele, aby hlasoval na štítku pro konkrétní testovací bod, ale vrací pouze agregovaný výsledek hlasování. Pokud všichni učitelé přiřadí stejný testovacímu vstupu, označení je téměř jistě správné, protože každý model dospěl k predikci samostatně. Kromě toho neexistuje způsob, jak by předpověď mohla narušit soukromí kterékoli ze školicích sad. Neshody mezi učiteli však mohly odhalit informace o datech v jejich sadách. Pro snížení tohoto rizika je do počtu hlasů zaveden hluk. Celkově tento přístup poskytuje rozdílnou záruku ochrany soukromí. Další výhodou je, že model mohl někdo vysvětlit a porozumět mu, i když nerozumí pojmu rozdílného soukromí. Kromě toho zavedení rozdílového soukromí také zlepšuje výkon, protože model extrahuje pouze vzory, které lze najít v tréninkových datech a snižuje dopad přeplnění. Přesnost i soukromí systému lze ještě více vylepšit odhalením předpovědí učitelů, když se všichni, nebo téměř všichni, dohodnou na předpovědi. Tento výsledek je vzrušující, protože je v rozporu s konvenční moudrostí, že soukromí vždy přichází s kompromisem v nástroji.

Během diskuse Papernot dále rozvinul, že jedním případem možného použití pro PATE, který ještě nebyl vyzkoušen, by bylo spíše rozdělování podmnožin již existujících dat, než umělé vytváření podmnožin. Uvedl příklad několika nemocnic spolupracujících na trénování ML modelu, ve kterém jsou jednotlivé předpovědi agregovány, aby vyškolily studentský model, který by se lépe přizpůsobil, než kterýkoli z jednotlivých modelů bez úniku soukromých informací.[89]

4.8 Ověření zabezpečení

Papernot pak diskutoval o práci vztahující se k principům úplného zprostředkování - myšlenka, že všechny přístupy by měly být kontrolovány - a kompromisní nahrávání - myšlenka, že případné neoprávněné přístupy budou zdokumentovány, pokud jim nelze zabránit. Vysvětlil, že tyto zásady se promítají do dvou hlavních potřeb pro ověřování kybernetické bezpečnosti: potřeba zdokonaleného zabezpečení modelu a kontroly přístupu v ML systémech. Podle Papernota má komunita ML tendenci soustředit se na průměrný výkon případu, jak je určeno testy přesnosti daného modelu. Z hlediska ochrany soukromí a bezpečnosti se však mohou výzkumníci více zajímat

[89] National Academies of Sciences, E. a.. *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*. Washington, DC. 2006, s. 47

o nejhorší výkon, jako měřítko spolehlivosti. Přestože byly vyvinuty užitečné způsoby měření soukromí systémů ML, je třeba více práce na stanovení metrik zabezpečení ML.[90]

V době školení je třeba zajistit jistotu modelu s jistotou, že jsou splněny požadavky na bezpečnost - to vyžaduje jasnou bezpečnostní politiku nebo formální požadavek toho, čeho chceme dosáhnout. Intuitivně víme, že chceme, aby systém uspěl dokonale v modelování úkolu, pro který byl navržen. [91]Formální definování toho není snadné. Dalo by se zvážit, zda je implementace správná nebo zda funguje s vysokou přesností, aniž by došlo k nežádoucímu chování, stejně jako vytvoření zadních dveří pro protivníky. V době testování existuje potřeba kontroly vstupu, tj. Způsobu výběru, zda by měl být pár vstupů / výstupů modelu zahrnut do souboru odpovědí sdílených s uživatelem - otázku, na kterou je dnes obtížné odpovědět. Rozhodnutí přichází na schopnost odhadnout jistotu predikce - stanovení nejistoty je však obtížné, protože skutečné rozdělení, které je modelováno, není známo. Nejistotu lze tedy odhadnout nejlépe způsobem, který není náchylný k manipulaci protivníky.

Papernot ukázal na prototypovaný systém nazvaný Deep k-Nearest Neighbors[92], který využívá informace v každé vrstvě hluboké neuronové sítě. Pro daný testovací bod systém identifikuje tréninková data, jejichž reprezentace se nejvíce shodují s reprezentací testovacího bodu a porovnává označení ve všech fázích. Pokud všechny reprezentace zůstanou konzistentní, takže štítky jsou ve všech vrstvách stejné, znamená to, že model předpovídá zobecněním, což je považováno za přesné s vysokou jistotou.[93] V případech kontradiktorní manipulace se však v určité vrstvě reprezentace a označení zkušebního bodu liší od údajů a údajů o tréninku, jejichž vstupní reprezentace byly podobné, což má za následek nesprávné označení testovacího bodu jako výstupu. Papernot nenaznačil, že by zkoumání štítků v každé fázi hluboké neuronové sítě a případné uvalení určitých omezení na strukturu procesu napříč všemi vrstvami, mohlo pomoci identifikovat nebo snížit potenciál pro nepřátelsky manipulované vstupy. Papernot navrhl, že vědci musí přemýšlet o tom, jak auditovat ML systémy. Například porovnání chyb provedených v rámci modelu na ochranu soukromí s chybami

90 National Academies of Sciences, E. a.. *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*. Washington, DC. 2006, s. 47

91 SMITH, R. (2012). *A Contemporary Look at Saltzer and Schroeder's 1975 Design Principles. Security & Privacy*, IEEE. 10. 20-25. 10.1109/MSP.2012.85

92*Deep k-Nearest Neighbors* [online]. [cit. 2020-04-19] Dostupné z WWW: https://arxiv.org/abs/1803.04765

93 PAPERNOT N., SONG S., MIRONOV I., RAGHUNATHAN A., TALWAR K., ERLINGSSON Ú., 2018, *"Scalable Private Learning with PATE. Sixth International Conference on Learning Representations* [online]. [cit. 2020-04-19]

provedenými v modelu na zachování soukromí, si může poskytnout informace o rozdílech ve výkonu a nakonec může být použito ke zlepšení výkonu i ochrany soukromí.[94]

Závěrem Papernot zdůraznil tři klíčové body. Nejprve je zapotřebí více práce, aby bylo možné určit správnou abstrakci nebo jazyk pro stanovení zásad bezpečnosti a ochrany soukromí, aby byla zajištěna jistota v ML systémech. Za druhé, zjistil potřebu auditu, pokud není možné zajistit jistotu, potenciálně spolu s karanténami, validací vstupů / výstupů a kompromisním zaznamenáváním. Zatřetí poznamenal, že takové mechanismy zabezpečení a ochrany soukromí by se měly snažit sladit s cíli samotného ML - takové mechanismy budou s větší pravděpodobností přijaty, pokud také zlepší výkon modelu. Poznamenal, že tyto doplňkové synergie by mohly být prozkoumány prostřednictvím výzkumu vztahu mezi soukromým učením, robustním učením a generalizací, nebo vztahem mezi otravou datem a učením z hlučných dat nebo v přítomnosti distribučních posunů. Na závěr citoval Goodhartův zákon[95]: „Když se opatření stane cílem, přestává být dobrým měřítkem."

Hlavním problémem ve všech případech je etická část, na kterou UI nikdy nebude schopno reagovat. Všechny faktory, které mohou být problémové, jsou závislé na schopnosti správného rozhodnutí člověka, protože v mnoha případech (téměř v majoritní většině) vyžadují morální a etické posouzení onoho rozhodnutí.[96]

Přičemž otázku etiky, morálky a všeobecných pravidel poplatných pro lidstvo, nejsme v tuto chvíli schopni simulovat v rámci systémů s UI a v nejbližší budoucnosti ani nebudeme.

Zůstává tedy otázka, jak celou situaci řešit?

[94] National Academies of Sciences, E. a.. *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop*. Washington, DC. 2006, s. 48-49

[95] *Goodhartovo pravidlo* [online]. [cit. 2020-04-19] Dostupné z WWW: https://cs.wikipedia.org/wiki/Goodhartovo_pravidlo

[96] STRATHERN M.. '*Improving Ratings': Audit in the British University System*," European Review 1997, s. 305-321

5 Návrh řešení problematiky UI v bezpečnostních systémech

V rámci této bakalářské práce a její praktické části níže navrhuji řešení zásadních otázek a problémů spojených s užíváním umělé inteligence v bezpečnostních systémech. V rámci tohoto bloku, a jakožto praktická část, zodpovídám jednotlivé otázky, přičemž každá ze zásadních otázek má dva pohledy:

Funkční – řeší realizaci a proveditelnost samotného činu vyplývajícího z otázky.

Etický – řeší etickou stránku věci, její správnost a schopnost stroje/UI posoudit ji a vyhodnotit správně.

A následně prezentuji vlastní návrh pro vyřešení položené otázky na základě vlastních zkušeností a pohledů.

5.1 Potenciální možnost útoku

V tomto případě chápeme potenciální možností útoku jako čin nebo příležitost, při které může systém zaútočit na majitele, obsluhu, obyvatele nebo kohokoliv, kdo je v jeho blízkosti. V tuto chvíli neřešíme možnosti a procentuální šance zdali nastane, nebo nenastane. Řešíme pouze principiální otázku, jestli k němu může dojít a proč.

Funkční – z hlediska funkčního přístupu k samotné otázce je důležité říci, že situací tohoto typu může v běžném životě (za využití těchto systémů) nastat nespočetné množství za den. Z faktického hlediska není problém, aby systém ohrozil člověka to ať už vědomě, nebo nevědomě. Faktorů je hned několik:

- zásah zvenčí – napadení systému třetí stranou a cílená změna mechanismů chování;
- chyba v kódu;
- vir;
- výpadek el. sítě;
- špatné vyhodnocení vstupních informací;
- neznalost konkrétní situace a neschopnost odpovědět si na základní otázky.

Etický – ve všech případech dochází k zásadním etickým otázkám, které z mého pohledu stroj nedokáže účinně rozpoznat a reagovat na ně. V případě, že dojde k napadení zvenčí a systém tak bude ovládán útočníkem, může jednoduše dojít k přenastavení systému. Pokud za UI dosadíme člověka, tak víme, že pokud mu někdo zavolá a řekne mu, že má zneškodnit obyvatele domu, tak víme s naprostou jistotou, že to neudělá. Mimo jiné z důvodů, že dokáže vyhodnotit absurditu onoho požadavku, jeho morální dopady a navíc dokáže rozpoznat, odkud vstupní informace přišla. Systém tuto možnost nemá, a tak se může velmi snadno rozhodnout špatně.

Návrh řešení – všechny vzorce chování, zvyky, algoritmy a reakce systému, jsou realizovány formou pracovních balíků a indexovány dle toho, jestli se proces systém naučil sám, nebo dostal direktivní zápis. V případě direktivního zápisu je jednoduché vytvořit systém podmínek a množinu situací, ve kterých se bude systém pohybovat. V případě procesů, které se naučil sám v rámci rekognice vzorců chování, by jakákoliv zásadní změna podléhala ověření a potvrzení ze strany majitele systému. V případě útoku zvenčí, by tak nebylo možné přímo změnit jednotlivé vzorce chování, aniž by je potvrdil majitel systému, například pomocí biometrických vstupních údajů.

5.2 Vnímání rizik a kontextu

Funkční – problémem v tomto ohledu je, že systém neumí posuzovat kontext a z něj plynoucí rizika. V případě, že máme po sobě jdoucí činnosti, které svou návazností ohrožují majitele domu (příkladem děti hrající si bez dozoru v místnosti, kde jsou zápalky a velké množství papíru, přičemž jedno z dětí na ně dosáhne a v minulosti již něco zapálilo), systém nedokáže vyhodnotit jejich souvztažnost. V uvedeném příkladu by systém vyčkával do doby, než by začal požár a až posléze by začal jednat.

Etický – systém by samozřejmě mohl včasněji reagovat a upozorňovat majitele, ale vzhledem k jeho neschopnosti chápat kontext, by docházelo k absurdním situacím. Z tohoto pohledu je velmi složité určit, jestli by měl systém upozornit majitele a děti například zavřít v jiné místnosti, nebo vyčkat. Omezení osobní svobody versus potenciální úmrtí jsou těžké otázky i pro člověka, natož aby je dokázal zodpovědět systém.

Návrh řešení – osobně považuji za jedinou možnost širší rozvoj „deep machine learningu“ s vysokokapacitním kvantovým počítačem, který by poskytl dostatek výkonu

a výpočetní kapacity pro kalkulaci všech souvztažností. Problémem pak ale bude, že se dostaneme do bodu, kdy bude systém schopen predikovat mnohem více, než člověk a z našeho pohledu začne porušovat etické zásady.

5.3 Bezpečnostní požadavky a přístup k nim

Funkční – systém musí vždy dodržovat a pracovat se všemi bezpečnostními požadavky majitele. Neexistuje varianta, ve které bude ignorovat některý z prvků jen proto, že nezapadá do určitého kontextu (který neumí správně vnímat, jak je uvedeno v druhé otázce). Chyba může nastat v případě, kdy nebudou požadavky definovány správně, pak může systém vyhodnocovat jednotlivé situace špatně a chybně na ně reagovat.

Etický – v případě, že dojdeme do bodu, kdy systém nebude schopen správně reagovat, by měla zafungovat určitá pojistka, která systém vypne/zabrání mu v dalším postupu. Systém není schopen vyhodnotit, co je chyba a co není. Porovnává stavy vůči zadaným vstupům a zkušenostem, které již má. Ani v jednom z případů nebude přemýšlet nad situací, která ještě nenastala.

Návrh řešení – za každé situace by měl mít systém přístup k integrační tabulce bezpečnostních opatření, které definují jeho pole působnosti a jasně popisují, co jsou správné přístup, a které jsou naopak nepřípustné. Jedná se o dvojí ochranu, kdy determinujeme povolené i zakázané přístupy a tím zmenšujeme množinu „nepopsaných“ stavů.

5.4 Princip ochrany informačního systému

Funkční - informační systém má chránit sám sebe tak, jak uvádí jeden z Asimovových zákonů robotiky. Tato ochrana je důležitá pro zachování vnitřní integrity systému a jeho funkčnosti, která zajišťuje ochranu obyvatel.

Etický – nikdy nesmí systém preferovat a upřednostnit svou ochranu před ochranou majitele. Problém je v samotné definici, což již popisuje rozvoj Asimovových zákonů robotiky. Systém si musí uvědomovat svou podstatu a vnitřní integritu a jeho postavení vůči člověku.

Návrh řešení – soubor pravidel a ochranných prvků, které jsou pevně implementovány v kódu systému. Je nutné vytvořit velký rozhodovací strom a pomocí hloubkové analýzy určit, které všechny potenciálně nebezpečné odbočky v systému existují. K nim vytvořit soubor exekutivních pravidel, která budou mít za úkol systém odstavit.

5.5 Princip ochrany uživatele

Funkční – za každé situace musí systém chránit svého uživatele a poskytovat mu služby v rozsahu, v jakém jsou definovány v jeho jádru. Ochrana uživatele systému je nadřazena vlastní ochraně.

Etický – systém je nástrojem uživatele a jeho bezpečnostním systémem, prvkem ochrany. Výměna rolí, kdy se ze sluhy stane pán, je problematikou, která ještě nebyla vyřešena. Jeden z experimentů společnosti Google [97]zahrnoval vytvoření dvou jazykových robotů, jejichž cílem mělo být domluvit se. Každý z robotů měl vlastní jazykovou sadu, které ten druhý nerozuměl. Vize byla, že se roboti navzájem naučí svůj jazyk. Realitou ale bylo, že chvíli po spuštění si roboti vytvořili vlastní jazyk, kterému uživatelé nerozuměli, odstřihli uživatele a připojili se k vnější síti. V tu chvíli jsme pochopili, že umělá inteligence zdaleka převyšuje naše představy o její funkčnosti a schopnostech.

Návrh řešení – ačkoliv se jedná o určitou formu anachronismu a tmářství, z mého pohledu a ze zkušeností ze světa, se jako jedinou možností ochrany zatím jeví omezení práv umělé inteligence. Ve své podstatě tak „škrtíme“ její potenciál, protože zatím neumíme vyřešit jednotlivé problematické důsledky.

5.6 Ochrana osobních údajů

Funkční – jedno z nejzásadnějších témat dnešní doby a nikdy nekončící boj. Ochrana osobních údajů se stala hlavní třecí plochou na poli informačních systému, a to nejen umělé inteligence. Samozřejmě je v tomto případě problém v tom, že osobní údaje, které bude sbírat bezpečnostní systém s UI, budou řádově podrobnější a citlivější, než „běžně dostupné“ osobní údaje.

[97] *Google* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://www.computerhope.com/jargon/g/google.htm >.

Etický – samotný sběr dat je neetický, protože na něj můžeme nahlížet jako na špehování. V případě narušení systému zvenčí může útočník získat kamerové záznamy, nahrávky rozhovorů a biometrické údaje. Mnohem dál zašla třeba Čína, která se netají tím, že jejich bezpečnostní systémy (security systems), povýšila na tzv. „surveillance systems[98]“ a pomocí biometrických skenů a prvků, získává data například o lidech procházejících pod konkrétní kamerou.

Návrh řešení – přísné šifrování a ukládání osobních údajů na úložiště, které je pro potřeby UI pouze jednosměrné.

5.7 Meze ochrany lidského zdraví

Funkční – nejpalčivější problém ve všech bezpečnostních systémech, kde je zároveň i UI. Ochrana lidského zdraví je prioritním účelem těchto bezpečnostních systémů. Další částí je samozřejmě ochrana majetku. Zatím nebyl nalezen klíč, který by přesně určoval, jak má systém postupovat.

Etický – zatím nerealizovatelná a nevyřčená odpověď, která stojí na vratkých filosofických tezích. Hranice ochrany lidského zdraví je jasně vystavitelná tam, kde mluvíme například o požáru a nutnosti evakuovat budovu. Systém může navádět obyvatele skrze budovu ven, a tak ochránit jejich zdraví. Nedokáže už ale rozhodovat v případě, kdy útočník při vloupání ohrožuje majitele domu zbraní. Systém, pakliže by měl tu možnost, nemůže útočníka zabít, ale nesmí nechat umřít majitele domu.

[98] *Surveillance* [online]. [cit. 2020-04-19] Dostupné z WWW: < https://study.com/academy/lesson/what-is-surveillance-definition-systems-techniques.html >.

Závěr

Ve své bakalářské práci s názvem „Moderní bezpečnostní systémy a využití umělé inteligence v rámci prevence kriminality“ jsem si vytyčila tři cíle řešení: primární, sekundární a terciální. Primárním cílem bylo zjištění skutečně velkých rizik, které přinášejí v současnosti užití umělé inteligence v oblasti bezpečnostních systémů. Sekundárním a terciálním cílem bylo postavení vlastní teorie o užití umělé inteligence, proti aktuálnímu trendu a dále vytvoření analýzy dopadu včetně návrhu opatření, které by snížilo všechna rizika. Pro svůj výzkum jsem použila základní metodiku, neboť díky jednoduchosti umožňovala rychlé a efektivní získávání výsledků. Volila jsem komparativní metodu, při které jsem porovnala klasické a bezpečnostní systémy s přínosy a s riziky vůči systémům využívajících umělou inteligenci nebo systémům řízených umělou inteligencí. Další způsob metodiky jsem zvolila na základě výsledků již proběhlých a zmapovaných experimentů metodu pozorování. Obě tyto metody jsem úzce provázala Asimovovými zákony robotiky, které jsou nosným pilířem v oblasti etické a filosofické otázky v problematice umělé inteligence.

V obsahu své bakalářské práce jsem se v úvodní části zabývala definicí týkající se umělé inteligence v bezpečnostních systémech. Z praktického hlediska jsem vymezila postupně rizika, klady a zápory jednotlivých druhů technologií, kde jsem porovnávala cenu, náročnost instalace, stupeň překonatelnosti a náročnost ovládání systému. Rozdělila jsem připojení na drátové a bezdrátové, mechanický zabezpečovací systém, poukázala jsem na výhody a nevýhody a na nebezpečí, plynoucí z užívání těchto systémů pomocí vlastní výroby. V závěru práce jsem ze dvou pohledů-funkčního a etického pokusila navrhnout řešení v jednotlivých principech jako je problematika umělé inteligence, bezpečnostní systémy, vnímání rizik a kontextu, bezpečnostní požadavky a přístup k nim, princip ochrany informačních systémů, ochrany uživatele, osobních údajů a vymezení ochrany liského zdraví. Pro příklad uvedu dva principy a to **potenciální možnosti útoku** a **ochrany informačních systémů.** U principu **potenciální možnosti útoku** jsem zjistila, že všechny vzorce chování, zvyky, algoritmy a reakce systému, byly realizovány formou pracovních balíků a indexovány dle toho, jestli se proces systém naučil sám, nebo dostal direktivní zápis. V případě direktivního zápisu bylo jednoduché vytvořit systém podmínek a množinu situací, ve kterých se systém pohyboval. V případě procesů, které v rámci rekognice vzorců chování se naučil systém sám, by jakákoliv zásadní změna podléhala ověření a potvrzení ze strany majitele systému.

V případě útoku zvenčí, by tak nebylo možné přímo změnit jednotlivé vzorce chování, aniž by je potvrdil majitel systému, například pomocí biometrických vstupních údajů.

Princip **ochrany informačního systému** vycházel ze souboru pravidel a ochranných prvků, které byly pevně implementovány v kódu systému. V tomto případě bylo nutné vytvořit velký rozhodovací strom a pomocí hloubkové analýzy určit, které všechny potenciálně nebezpečné odbočky v systému existují a k nim vytvořit soubor exekutivních pravidel, který bude mít za úkol systém odstavit.

V závěru nutno konstatovat, že veškeré cíle pro napsání této práce, které byly vytyčeny, byly dosaženy.

Bakalářská práce by tak mohla být využitelná jako odborná literatura nejen pracovníků Policie ČR při odhalování zločinu a to například v oblasti úvěrových podvodů, ale i jako odborná pomůcka IT pracovníků – programátorů, působících jak ve veřejné správě, tak v soukromém sektoru.

Literární zdroje

1. GOODFELLOW, I., B. Y. *Deep learning* Massachusetts: Cambridge: The MIT Press, 2016
2. D. R., H. *Gödel, Escher, Bach: an Eternal Golden Braid.* New York: Basic Books, cop., 1999
3. ROSA, J. L. *Artificial Neural Networks - Models and Aplications* InTech, 2016
4. LIPPMANN, R. *An introduction to computing with neural nets. IEEE ASSP Magazine.1987*
5. ARBIB, M.A. *Brain theory and neural networks.* Cambridge, MA, USA, 2003
6. FLOREANO, F., MATTIUSSI, C. *Bio-inspired artificial intelligence, theories, methods and technologies.* Cambridge, MA, USA: The MIT Press 2008
7. HUANG, D.S. *Radial basis probabilistic neural networks: model and applications.* International Journal of Pattern Recognition and Artificial Intelligence 1999
8. National Academies of Sciences, E. a.. *Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop. Washington, DC 2006*
9. STRATHERN, M.. *'Improving Ratings': Audit in the British University System," European Review* 1997
10. BISHOP, CH.M. *Pattern Recognition and Machine Learning*. Cambridge CB3 0FB, U.K, 2006. 758 s. ISBN 978-0387-31073-2.
11. RUSSEL, S.J., Norvig.P. *Artificial Intelligence A Modern Approach Third Edition* : New Jersey; Upper Saddle River, 2010. 1151 s. ISBN 978-0-13-604259-4.
12. McCAULEY, L. *AI Armageddon and the Three Laws of Robotics*.
13. CLARKE, R. (2011). *Asimov's Laws of Robotics. In M. Anderson & S. Anderson (Eds.), Machine Ethics* (pp. 254-284). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511978036.020
14. ANDERSON, S. (2011). *The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics. In M. Anderson & S. Anderson (Eds.), Machine Ethics* (pp. 285-296). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511978036.021
15. SUTRISNO, I. (2016). *A comprehensive review on intelligent surveillance systems. Communications in Science and Technology*. 1. 10.21924/cst.1.1.2016.7

16. ZUPAN, J. *Introduction to artificial neural network methods: what they are and how to use them*. Acta Chimica Slovenica, 1994, s. 327–352.
17. PAPERNOT, N. (2018). *A Marauder's Map of Security and Privacy in Machine Learning: An overview of current and future research directions for making machine learning secure and private*. AISec '18: Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security. 1-1. 10.1145/3270101.3270102.
18. PAPERNOT, N., McDaniel, P., SINHA, A., WELLMAN, M. (2018). *SoK: Security and Privacy in Machine Learning*. 399-414. 10.1109/EuroSP.2018.00035.
19. SMITH, R. (2012). *A Contemporary Look at Saltzer and Schroeder's 1975 Design Principles. Security & Privacy*, IEEE. 10. 20-25. 10.1109/MSP.2012.85.
20. PAPERNOT, N., SONG, S., MIRONOV, I., RAGHUNATHAN, A., TALWAR, K., ERLINGSSON, Ú., 2018, "*Scalable Private Learning with PATE*. Sixth International Conference on Learning Representations (ICLR 2018)
21. PAPERNOT, N., SONG, S., MIRONOV, I., RAGHUNATHAN, A., TALWAR, K., ERLINGSSON, Ú. 2018, "*Scalable Private Learning with PATE*. Sixth
22. STRATHERN, M. '*Improving Ratings'*: Audit in the British University System," European Review 1997, s. 305-321

Elektronické zdroje

1. MURPHY R., R., WOODS, D., D . *Beyond Asimov: The Three Laws of Responsible Robotics 2009* [online]. Ohio, 1541-1672/09/$26.00 © 2009 IEEE Published by the IEEE Computer Society. Dostupné z WWW: < https://www.researchgate.net/publication/224567023_Beyond_Asimov_The_Three_Laws_of_Responsible_Robotics>.
2. WINFIELD, A. (2018). *Experiments in Artificial Theory of Mind: From Safety to Story-Telling. Frontiers in Robotics and AI*. 5. 10.3389/frobt.2018.00075.
3. *Pojem metoda komparativní* [online]. Dostupné z WWW: < https://slovnik-cizich-slov.abz.cz/web.php/slovo/metoda-komparativni>.
4. ANDERSON, S. (2011). *The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics*. In M. Anderson & S. Anderson (Eds.), Machine Ethics (pp. 285-296) Cambridge: Cambridge University Press. doi:10.1017/CBO9780511978036.021
   Dostupné z WWW: https://www.cambridge.org/core/books/machine-ethics/unacceptability-of-asimovs-three-laws-of-robotics-as-a-basis-for-machine-ethics/D58C8BAD402DF52AD2785C17A68431EB
5. *Co je to BEZPEČNOSTNÍ TŘÍDA*? [online]. Dostupné z WWW: < https://www.bezpecnostni-dvere-mrize-kavan.cz/co-je-to-bezpecnostni-trida/>
6. *Honeywell* [online]. Dostupné z WWW: < https://www.honeywell.com/en-us/company/about-us >
7. *Biometrics and biometric data: What is it and is it secure?* [online]. Dostupné z WWW: < https://us.norton.com/internetsecurity-iot-biometrics-how-do-they-work-are-they-safe.html >
8. *Hra na Honěnou* [online]. Praha Dostupné z WWW: < https://www.databazeknih.cz/povidky/hra-na-honenou-444 >
9. *Sci-fi* [online]. Dostupné z WWW: < https://literaryterms.net/science-fiction/ >
10. *What is a neural network?* [online]. Dostupné z WWW: < https://www.techradar.com/news/what-is-a-neural-network >
11. *What is a qubit?* [online] Dostupné z WWW: < https://www.quantum-inspire.com/kbase/what-is-a-qubit/ >
12. *Terminator* [online] Dostupné z WWW: < https://www.thisisbarry.com/film/terminator-film-series-all-plots-explained/ >
13. *Premisa* [online]. Dostupné z WWW: < https://www.czechency.org/slovnik/LOGICK%C3%89%20VYPL%C3%9DV%C3%81N%C3%8D >
14. *Kauzalita* [online]. Dostupné z WWW: < https://encyklopedie.soc.cas.cz/w/Kauzalita >
15. *Entita* [online]. Praha Dostupné z WWW: < https://it-slovnik.cz/pojem/entita >
16. *What is Deep Learning?* [online]. Dostupné z WWW: < https://machinelearningmastery.com/what-is-deep-learning/ >
17. *Artificial Neural Network (ANN)* [online]. Dostupné z WWW: < https://www.investopedia.com/terms/a/artificial-neural-networks-ann.asp>

18. *feed-forward-back-propagation neural network* [online]. Dostupné z WWW: < https://www.researchgate.net/figure/Structure-of-the-feed-forward-back-propagation-neural-network-FBNN_fig1_336138883>
19. *Samoučící se neuronová síť - SOM, Kohonenovy mapy* [online]. Dostupné z WWW: < https://www.kiv.zcu.cz/studies/predmety/uir/NS/Samouc_NN2.pdf>
20. *Kohonen neural network* [online]. Dostupné z WWW: https://www.sciencedirect.com/science/article/abs/pii/S0003267096003157>
21. *Mechanické zábranné systémy (MZS)* [online]. Dostupné z WWW: https://www.security.cz/mechanicke-zabranne-systemy-mzs--2422.html
22. *high end* [online]. Dostupné z WWW:https://www.wordreference.com/encz/high-end
23. *Smart home* [online]. Dostupné z WWW: https://www.inels.cz/smarthome
24. *Elon Musk* [online]. Dostupné z WWW: < https://www.tesla.com/cs_CZ/elon-musk >
25. *Tesla* [online]. Dostupné z WWW: < https://www.tesla.com/cs_CZ/about >
26. *Machine learning* [online]. Dostupné z WWW: https://en.wikipedia.org/wiki/Machine_learning
27. *Goodhartovo pravidlo* [online]. Dostupné z WWW: https://cs.wikipedia.org/wiki/Goodhartovo_pravidlo
28. *Google* [online]. Dostupné z WWW: < https://www.computerhope.com/jargon/g/google.htm >
29. *Surveillance* [online]. Dostupné z WWW: < https://study.com/academy/lesson/what-is-surveillance-definition-systems-techniques.html >

Ostatní zdroje

Kromě výše uvedených zdrojů byly při zpracování bakalářské práce využity následující materiály:

- Konzultace se zástupci společnosti Deloitte, KPMG a Google
- Využití vývojářského prostředí Google a jejich neuronové sítě pro překlad některých textů
- Vývojové diagramy společnosti Google

Seznam zkratek

24/7	Určuje časové rozmezí 24 hodin 7 dní v týdnu	20
AHRC	Arts and Humanities Research Council	35
AI	artificial inteligence	27
AKU	akumulátorové nářadí	14
ANN	artificial neural network	39
as-is	používá se pro popis věci, nebo situace tak, jak aktuálně je (její aktuální stav)	11
BP	Backpropagation	46
BPFFNN	Backpropagation feed-forward neural network	47
BPNN	Backpropagation neural network	47
brute force	metoda síly, kdy se zkouší kombinace postupně jedna po druhé	38
CPU	Central processing unit - hlavní procesorová jednotka	57
DPH	daň z přidané hodnoty	52
end-to-end	forma - konec až konec - určuje komplexní testování/průběh od začátku až do konce	11
ESPRC	Engineering and Physical Sciences Research Council	35
EUCog	Society for Cognitive Systems	35
EZS	Elektronický zabezpečovací systém	15
FBNN	Front-backward neural network	42
FFBPNN	Feed-forward backpropagation neural network	46
FFNN	Feed-forward neural network	42
Haiku	druh japonské poezie	38
High-end	označení nejmodernějších technologií	54
IEEE	Institute of Electrical and Electronics Engineers	34
IZS	Integrovaný záchraný systém	13
KNN	k-nearest neighbours algorithm	46
ML	machine learning - strojové učení	56
MLP	Multi-level perceptron	46
MZS	mechanické zabezpečovací systémy	52
Neural AI smart home	chytrá domácnost s umělou inteligencí na principu neuronové sítě	55
PATE	Private Aggregation of Teacher Ensembles	59
PC	Personal computer = osobní počítač	14
RD	Rodinný dům	13
RFID	Radio Frequency Identification	22
sci-fi	science fiction	26
Smart Home	chytrá domácnost - označení bezpečnostních prvků s integrovanou umělou inteligencí nižšího řádu	55
SNN	Spiking neural networks	46
SOM	self organizing maps	46
SPZ	státní poznávací značka	22
SW	Software	21
UI	Umělá inteligence	21
Wi-Fi	wireless fidelity - bezdrátový přenos dat	50
YouTube	internetová stránka prezentující uživatelský video obsah	52