

VYSOKÁ ŠKOLA POLYTECHNICKÁ JIHLAVA

Aplikovaná informatika

IDENTIFIKACE OBRÁZKŮ VYTVOŘENÝCH UMĚLOU  
INTELIGENCÍ

Bakalářská práce

Autor práce: David Barzul

Vedoucí práce: doc. Dr. Ing. Jan Voráček, CSc.

Jihlava 2026

# Vysoká škola polytechnická Jihlava

Tolstého 16, 586 01 Jihlava

## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Autor práce: **David Barzul**

Studijní program: Aplikovaná informatika

Garant studijního programu: Ing. Lenka Kuklišová Pavelková, Ph.D.

Název práce: **Identifikace obrázků vytvořených umělou inteligencí**

Vedoucí práce: doc. Dr. Ing. Jan Voráček, CSc.

Cíl práce: Obecným cílem práce je ověření funkčnosti vybraného systému pro identifikaci obrázků, vytvořených umělou inteligencí. Naplněn bude prostřednictvím následujících konkrétních cílů: (1) Rešerše současného stavu poznání v oblasti strojové klasifikace obrázků, s důrazem na problematiku identifikace obrázků, generovaných prostřednictvím umělé inteligence. (2) Komparativní analýza volně dostupných nástrojů z této oblasti. (3) Výběr platformy, vhodné pro realizaci praktické části této práce. Pořízení relevantního datového souboru. Návrh, vyhodnocení a diskuse výsledků klasifikačních experimentů. (4) Zobecnění získaných poznatků a formulace doporučení pro další výzkum.

## Abstrakt

Bakalářská práce se zabývá způsobem, jak poznat, zda byl obrázek vytvořen umělou inteligencí. V posledních letech generativní modely rychle pokročily a dokážou vytvářet snímky, které vypadají jako skutečné fotografie. Původ obrázků vytvořených umělou inteligencí bývá často těžké určit. Uvedený stav představuje problém pro digitální pravdivost, ochranu autorských práv a potírání dezinformací.

Cílem práce je ověřit, zda vybraný systém dokáže spolehlivě rozpoznat obrázky vyrobené umělou inteligencí. Teoretická část popisuje současné metody strojového rozpoznávání obrazů a způsoby, jak odhalit obsah vygenerovaný AI. Analytická a praktická část porovná dostupné nástroje, vybere vhodnou platformu a provede klasifikační pokusy na připravené sadě dat. Výsledky pokusů se zhodnotí pomocí běžných metrik a probere se přínos i limity dosažených výsledků.

## Klíčová slova

Umělá inteligence; generování obrazů; detekce AI-generovaných obrázků; klasifikace obrazů; Decopy.ai

## Abstract

The bachelor thesis deals with how to recognize whether an image was created by artificial intelligence. In recent years, generative models have advanced rapidly and are capable of creating images that look like real photographs. The origin of images created by artificial intelligence is often difficult to determine. This situation poses a problem for digital trust, copyright protection, and combating misinformation.

The aim of this thesis is to verify whether the selected system can reliably recognize images produced by artificial intelligence. The theoretical part describes current methods of machine image recognition and ways to detect AI-generated content. The analytical and practical part compares available tools, selects a suitable platform, and performs classification experiments on a prepared data set. The results of the experiments will be evaluated using common metrics, and the benefits and limitations of the results achieved will be discussed.

## Keywords

Artificial intelligence; image generation; detection of AI-generated images; image classification; Decopy.ai

Prohlašuji, že předložená bakalářská práce je původní a zpracoval/a jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem v práci neporušil/a autorská práva (ve smyslu zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů, v platném znění, dále též „AZ“).

Byl/a jsem seznámen/a s tím, že na mou bakalářskou práci se plně vztahuje **AZ**, zejména § 60 (školní dílo).

Podle § 47b zákona o vysokých školách souhlasím se zveřejněním své práce podle Směrnice pro vedení, vypracování a zveřejňování závěrečných prací na VŠPJ, a to bez ohledu na výsledek obhajoby.

Beru na vědomí, že VŠPJ má právo na uzavření licenční smlouvy o užití mé bakalářské práce a prohlašuji, že **s o u h l a s í m** s případným užitím mé bakalářské práce (prodej, zapůjčení apod.).

Jsem si vědom/a toho, že užití své bakalářské práce či poskytnout licenci k jejímu využití mohu jen se souhlasem VŠPJ, která má právo ode mě požadovat přiměřený příspěvek na úhradu nákladů, vynaložených vysokou školou na vytvoření díla (až do jejich skutečné výše), z výdělku dosaženého v souvislosti s užitím díla či poskytnutím licence.

V Jihlavě dne Klikněte nebo klepněte sem a zadejte datum.

.....

Podpis studenta/ky

## Poděkování

Rád bych poděkoval vedoucímu mé bakalářské práce, doc. Dr. Ing. Janu Voráčkovi, CSc., za jeho pomoc, inspiraci a rady při psaní práce. Rovněž bych chtěl poděkovat rodině, přátelům a přítelkyni za podporu, trpělivost a motivaci během psaní práce.

## Obsah

<b>Seznam obrázků.....</b>	<b>7</b>
<b>Seznam tabulek .....</b>	<b>8</b>
<b>Seznam zkratk.....</b>	<b>9</b>
<b>Úvod .....</b>	<b>10</b>
<b>1 Teoretická část .....</b>	<b>11</b>
1.1 Umělá inteligence a generování obrázků .....	11
1.1.1 Hlavní typy generování obrázků.....	12
1.2 Klasifikace grafických dat.....	13
1.2.1 Pravidlová klasifikace obrázků.....	13
1.2.2 Statistická klasifikace obrázků.....	13
1.2.3 Metody klasifikace založené na distribuci.....	14
1.2.4 Metody klasifikace bez distribuce .....	15
1.3 Metody strojového rozpoznávání obrázků.....	15
1.3.1 Regionální konvoluční neuronová síť .....	16
1.3.2 Jednostupňové detekční architektury.....	17
1.4 Limity a výzvy detekce AI-generovaných obrázků .....	18
1.5 Komparativní analýza volně dostupných nástrojů .....	19
1.5.1 Limity detekce AI-generovaných obrázků v nástroji Decopy.ai .....	22
1.6 Vnitřní principy fungování nástroje Decopy.ai .....	22
1.7 Datové sady pro trénování klasifikátorů a identifikátorů obrázků .....	23
1.8 Datové sady pro testování identifikátorů obrázků .....	24
<b>2 Metodika .....</b>	<b>26</b>
2.1 Nástroj Decopy.ai.....	26
2.2 Způsob práce s daty a princip jejich zpracování .....	27
2.3 Návrh metodiky experimentu a stanovení metrik.....	28
<b>3 Praktická část .....</b>	<b>30</b>
3.1 Datová sada .....	30
3.2 Postup experimentu .....	31
3.3 Vyhodnocení výsledků .....	33
3.3.1 Vyhodnocení výsledků kategorie jídlo.....	33
3.3.2 Vyhodnocení výsledků kategorie lidé.....	34
3.3.3 Vyhodnocení výsledků kategorie zvířata.....	35
3.3.4 Vyhodnocení výsledků kategorie příroda.....	36
3.3.5 Vyhodnocení výsledků kategorie sport.....	37
3.4 Analýza chyb klasifikace.....	38
3.5 Diskuse výsledků.....	40
3.6 Shrnutí praktické části .....	40
<b>Závěr .....</b>	<b>41</b>
<b>Seznam použité literatury .....</b>	<b>43</b>
<b>Přílohy.....</b>	<b>45</b>

## Seznam obrázků

Obrázek 1: Diagram of a convolutional neural network (CNN) .....	14
Obrázek 2: Faster R-CNN.....	17
Obrázek 3: Úvodní stránka webu Decopy.ai .....	22
Obrázek 4: Možnosti měsíčních předplatných.....	27
Obrázek 5: Metriky matice záměny .....	28
Obrázek 6: Ukázka souboru dat .....	31
Obrázek 7: Příklad analýzy a výsledků nástroje .....	32
Obrázek 8: Příklad části původní tabulky a záznamu výsledků.....	32
Obrázek 9: Příklad nesprávné klasifikace obrázku (Skutečný jako AI) .....	38
Obrázek 10: Příklad nesprávné klasifikace obrázku (AI jako skutečný) .....	39

## Seznam tabulek

Tabulka 1: Komparativní analýza volně dostupných nástrojů .....	21
Tabulka 2: Výsledky z analýzy dat .....	33
Tabulka 3: Výsledky z analýzy dat .....	34
Tabulka 4: Výsledky z analýzy dat .....	35
Tabulka 5: Výsledky z analýzy dat .....	36
Tabulka 6: Výsledky z analýzy dat .....	37
Tabulka 7: Výsledky z analýzy dat .....	39

## Seznam zkratk

UI	Umělá inteligence
AI	Umělá inteligence
VAE	Variational AutoEncoder
CNN	Convolutional Neural Network (Konvoluční neuronové síť)
DM	Diffusion Models
GAN	Generative Adversarial Networks
VŠPJ	Vysoká škola polytechnická Jihlava
R-CNN	Region-Based Convolutional Neural Network
Fast R-CNN	Fast Region-Based Convolutional Neural Network
Faster R-CNN	Faster Region-Based Convolutional Neural Network
RPN	Region Proposal Network
YOLO	You Only Look Once
SSD	Single Shot MultiBox Detector
MV	Maximální věrohodnost

## Úvod

V posledních letech technologie umělé inteligence výrazně pokročily a silně ovlivňují vznik, úpravu i šíření digitálního obsahu. Nejvýrazněji se rozvíjí generování obrazových dat pomocí generativních modelů, které dokážou vytvářet vizuální výstupy působící jako skutečné fotografie. Obrázky vyrobené umělou inteligencí se už nyní využívají v mnoha oblastech, zároveň však přinášejí nové problémy s určením původu a ověřením důvěryhodnosti.

Určení, zda byl obrázek vytvořen algoritmem, je v současnosti velmi žádané. Rozlišení skutečné fotografie od synteticky vygenerovaného snímku je klíčové pro boj s dezinformacemi, ochranu autorských práv a zachování důvěry v digitální média. S rostoucí kvalitou výstupů generativních modelů se hranice mezi realitou a umělým obsahem stává téměř neviditelnou, proto roste poptávka po spolehlivých detekčních nástrojích.

Cílem bakalářské práce je ověřit, zda vybraný systém dokáže spolehlivě určit, že obrázek vytvořila umělá inteligence. Práce analyzuje metody strojového rozpoznávání obrazu, porovnává volně dostupné nástroje a v praxi prověřuje schopnost klasifikace obrazových dat. Přestože je práce technicky zaměřená, zabývá se rovněž etickými a společenskými dopady rozvoje umělé inteligence.

Bakalářská práce obsahuje tři hlavní části – teoretická část popisuje principy umělé inteligence při generování obrázků, způsoby klasifikace grafických dat, metody strojového rozpoznávání obrazů a přehled dostupných nástrojů pro detekci AI-generovaných obrázků. Metodická část práce obsahuje informace o vybraném nástroji, volbě způsobu práce s daty, návrhu metodiky a výběru metrik pro praktickou část práce. Praktická část vybírá, navrhuje a uskutečňuje klasifikační experiment a vyhodnocuje získané výsledky.

# 1 Teoretická část

Teoretická část bakalářské práce poskytuje odborný základ nezbytný pro pochopení problematiky identifikace obrázků vytvořených umělou inteligencí. Rozvoj technologií strojového učení a hlubokých neuronových sítí v posledních letech umožnil vznik generativních modelů, které jsou schopny vytvářet obrazová data s vysokou mírou realismu. Tyto modely zásadním způsobem ovlivnily způsob práce s digitálním obrazem a otevřely nové možnosti v oblasti tvorby vizuálního obsahu.

Současně s rozvojem generativních technologií roste potřeba metod, které dokáží rozlišit reálné obrazové záznamy od synteticky generovaných dat. Schopnost identifikace původu obrazů má význam nejen z technického hlediska, ale také v kontextu ochrany autorských práv, digitální důvěryhodnosti a prevence šíření dezinformací. Z tohoto důvodu se problematika strojového rozpoznávání a klasifikace obrazových dat stala předmětem intenzivního výzkumu.

Teoretická část se nejprve zaměřuje na základní principy umělé inteligence a generování obrazů, včetně přehledu nejvýznamnějších generativních modelů. Následně jsou představeny metody klasifikace grafických dat a způsoby extrakce obrazových příznaků. Závěrečná část teoretického přehledu se věnuje metodám strojového rozpoznávání obrázků, a to jak tradičním přístupům, tak moderním metodám založeným na hlubokém učení a dále limity nástrojů používaných k identifikaci fotografií vytvořených umělou inteligencí, vnitřní principy fungování vybraného nástroje a výběr sady dat pro praktickou část práce.

## 1.1 Umělá inteligence a generování obrázků

Umělá inteligence (Artificial Intelligence, UI, AI) je technologie, která umožňuje počítačům a strojům simulovat lidské učení, porozumění, řešení problémů, rozhodování, kreativitu a autonomii. Systémy UI se učí z dat, což jim umožňuje řešit složité problémy a jednoduché opakující se úkoly a postupem času zlepšovat své reakce. Umělá inteligence se dělí na základní podkategorie:

- **Strojové učení** – systém AI, který se učí na základě historických dat.
- **Hluboké učení** – modely, které využívají neuronové sítě a fungují podobně jako náš mozek.
- **Generativní AI** – modely hlubokého učení, které vytvářejí nový a originální materiál.

Umělé neuronové sítě napodobují proces rozpoznávání vzorů v mozku. Jeden z typů hlubokého učení, konvoluční neuronové sítě (Convolutional Neural Network, CNN), se specializuje na schopnost identifikovat objekty a vzory v datech. Neurony jsou specializovanou formou, která funguje podobným způsobem jako lidské oko. Nejsou sice tak složitá a komplexní jako lidský mozek, stroj dokáže rozpoznat obraz podobným způsobem, jakým vidí lidé. Je důležité si uvědomit, že stroj nevidí obrázky, ale pouze soubor čísel. Obrázek je rozdělen na pixely, které jsou reprezentovány čísly představujícími čáry, hrany a barvy.

K vytvoření obrázků generovaných umělou inteligencí model strojového učení prohledává miliony obrázků na internetu spolu s textem, který se k nim vztahuje. Algoritmy rozpoznávají trendy v obrázcích a textech a nakonec začnou odhadovat, které obrázky a texty k sobě pasují. Jakmile model dokáže předpovědět, jak by měl obrázek podle daného textu vypadat, může

vytvořit zcela nové obrázky od nuly na základě nové sady popisných textů, které uživatelé zadají do aplikace (Canva, 2025).

### 1.1.1 Hlavní typy generování obrázků

Jedním ze základních přístupů k generování obrazů je **učení modelů UI na velkých datových souborech**, které obsahují obrázky a odpovídající textové popisy.

V procesu učení model analyzuje statistické závislosti mezi vizuálními znaky a textovými popisy. Po dokončení učení, po obdržení textového dotazu (prompt), model neprovádí vyhledávání mezi dostupnými obrázky, ale generuje nový obrázek na základě naučených zákonitostí a vytváří výsledek, který odpovídá zadanému popisu.

**Variational AutoEncoder (VAE)** je druh autoenkodérů – neuronových sítí určených ke kompresi a obnově dat. Model se skládá ze dvou hlavních komponent: enkodéru a dekodéru.

- **Enkodér** převádí vstupní data do kompaktního skrytého zobrazení. Na rozdíl od klasických autoenkodérů VAE nevytváří pevné latentní zobrazení, ale parametry pravděpodobnostního rozdělení, ze kterého se vybírá skrytý vektor.
- **Dekodér** zase používá tento latentní vektor k obnovení vstupních dat. Učení modelu probíhá tak, aby se zmenšil rozdíl mezi původními a obnovenými daty a aby se zajistila regularizace latentního prostoru.

Díky pravděpodobnostnímu přístupu je VAE schopen nejen obnovit vstupní data, ale také generovat nové vzorky podobné těm, které byly obsaženy v trénovacím souboru. Pokud bychom měli použít analogii, VAE lze přirovnat k člověku, který si nezapamatuje každý obraz zvlášť, ale studuje obecné styly a formy, což mu umožňuje malovat nové obrazy v podobném stylu.

**Diffusion Models (DM)** jsou třída generativních modelů založených na postupném zašumění dat a učení zpětného procesu jejich obnovy. V průběhu učení se k obrázkům postupně přidává náhodný šum a model se učí předpovídat a odstraňovat tento šum v každé fázi. Po dokončení učení začíná generování nových obrazů náhodným šumem, který se postupně transformuje do strukturovaných dat. Tento přístup umožňuje modelu vytvářet nové vzory, které odpovídají rozložení trénovacích dat, bez ukládání nebo kopírování konkrétních příkladů.

Pokud dáme příklad, lze difúzní modely přirovnat k procesu vyvolávání fotografie: nejdříve je vidět pouze chaotické pozadí, ale s každým krokem se objevují stále jasnější detaily obrazu.

**Generative Adversarial Networks (GAN)** jsou třída generativních modelů, které se skládají ze dvou neuronových sítí: generátoru a diskriminátoru. Generátor vytváří nové vzorky dat, zatímco diskriminátor je hodnotí a snaží se odlišit generovaná data od skutečných. Učení GAN probíhá formou soutěžního procesu, ve kterém se generátor snaží vytvářet vzorky, které se co nejvíce blíží reálným datům, a diskriminátor je správně klasifikovat. Výsledkem této soutěže je, že se generátor postupně učí reprodukovat rozložení trénovacích dat a vytvářet nové, realistické příklady.

Pokud uvést příklad, pak umělec, který maluje kopie obrazů, se snaží oklamat kritika, a kritik se stává stále pozornějším. Výsledkem je, že umělec začíná vytvářet stále lepší obrazy.

## 1.2 Klasifikace grafických dat

Od raného dětství se člověk učí rozlišovat objekty okolního světa, zejména živé a neživé předměty, rostliny a zvířata. Díky neustálému pozorování a hromadění zkušeností se tento proces postupem času stává intuitivním.

Pro počítačové systémy není schopnost přirozená a vyžaduje speciální metody učení. Právě proto hraje důležitou roli klasifikace obrázků, která umožňuje počítači analyzovat vizuální informace, určovat příslušnost objektů k určitým kategoriím a interpretovat obsah obrázků.

Klasifikace grafických dat představuje klíčovou oblast počítačového vidění a strojového učení. Proces klasifikace obvykle zahrnuje extrakci významných obrazových příznaků, jejich následné zpracování a přiřazení vstupních dat do předem definovaných tříd. V průběhu vývoje byly navrženy různé přístupy, od jednoduchých pravidlových metod až po pokročilé modely hlubokého učení.

### 1.2.1 Pravidlová klasifikace obrázků

Pravidlová metoda je založena na přísně navrženém procesu sběru a označování obrázků podle konkrétního úkolu nebo cíle klasifikace. Tento proces provádějí ručně odborníci, kteří vybírají klíčové vlastnosti obrázku, které poskytují nejvíce vizuálních informací. Klasifikace obrázků na základě pravidel seskupuje podobné klastry pixelů do tříd pomocí těchto pravidel, která jsou založena na specializovaných znalostech. To také umožňuje provádět interpretovanou a přizpůsobitelnou klasifikaci bez použití složitých modelů strojového učení.

Mezi typické příklady pravidlové klasifikace patří prahování, detekce hran nebo šablonové porovnávání, kde jsou objekty identifikovány na základě předem stanovených vizuálních kritérií. Tyto metody jsou často využívány v úlohách s jasně definovanými podmínkami a omezeným počtem tříd. Výhodou pravidlového přístupu je vysoká transparentnost rozhodovacího procesu, díky níž lze snadno vysvětlit, proč byl konkrétní obraz zařazen do dané kategorie. Nevýhodou je naopak omezená schopnost přizpůsobení se variabilním nebo složitým datům, kde ručně definovaná pravidla nemusí dostatečně pokrýt všechny možné situace. Pravidlové metody jsou proto v praxi často kombinovány s pokročilejšími algoritmy, které zvyšují robustnost a přesnost klasifikace.

### 1.2.2 Statistická klasifikace obrázků

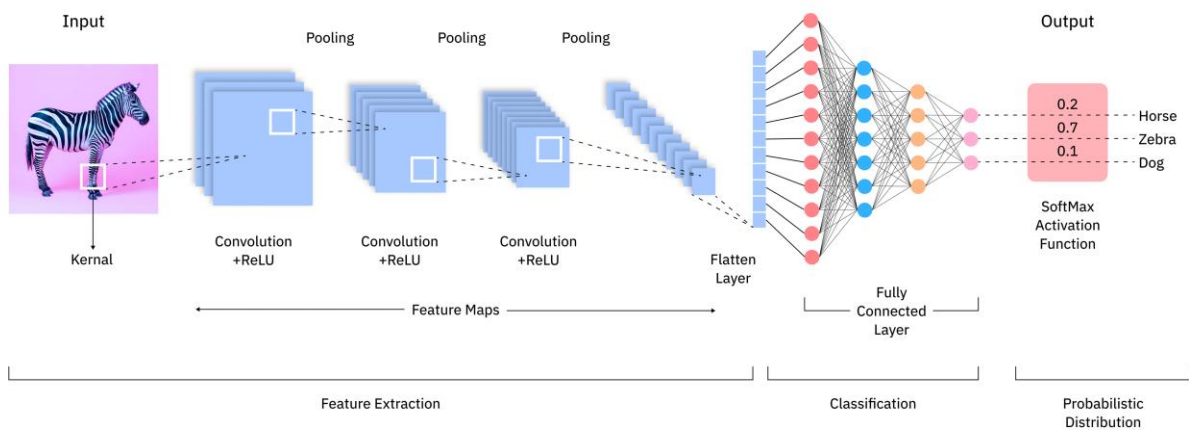
Statistický přístup ke klasifikaci je složitější než klasifikace obrazů založená na pravidlech. Statistická klasifikace obrázků se zaměřuje na automatické učení a identifikaci vzorů v obrázcích. Pro dosažení přesných výsledků je závislá na velkých označených datových sadách a pokročilých modelových architekturách, nejčastěji CNN.

Konvoluční neuronové sítě představují specifický typ hlubokých neuronových modelů navržených pro efektivní zpracování obrazových dat. Namísto plně propojených vrstev využívají CNN konvoluční vrstvy, které extrahují prostorové příznaky z obrazů pomocí souboru učitelých filtrů. Díky své schopnosti automaticky učit hierarchické reprezentace dat dokážou CNN nejen zachytit nízkourovňové rysy jako hrany nebo textury, ale i složitější vyšší úrovně obrazových struktur, což je klíčové pro úlohy klasifikace a detekce objektů (Moez Krichen, 2023).

Statistické metody nevyžadují ruční definování pravidel, jelikož modely se učí přímo z trénovacích dat na základě pravděpodobnostních vztahů mezi vstupními znaky a výstupními třídami. Mezi hlavní výhody tohoto přístupu patří vysoká přesnost a schopnost generalizace na nová data. Nevýhodou je vyšší výpočetní náročnost a nižší interpretovatelnost výsledků ve srovnání s pravidlovými metodami. Statistická klasifikace se proto nejčastěji využívá v úlohách, kde je k dispozici dostatek dat, například v rozpoznávání objektů, obličejů nebo při detekci obrazů generovaných umělou inteligencí.

Popis obrázku:

Obrázek znázorňuje základní architekturu konvoluční neuronové sítě pro klasifikaci obrazových dat. Vstupní obraz je zpracován konvolučními vrstvami, které extrahují důležité vizuální příznaky, přičemž pooling vrstvy snižují rozměr dat a zvyšují odolnost vůči drobným změnám v obraze. Po fázi extrakce příznaků jsou data převedena do jednorozměrného vektoru a zpracována plně propojenými vrstvami. Výstupní vrstva se softmax aktivací funkcí určuje pravděpodobnostní rozdělení tříd a výslednou klasifikaci obrazu.



**Obrázek 1: Diagram of a convolutional neural network (CNN)**

*zdroj: What is image classification? (IBM, 2025)*

### 1.2.3 Metody klasifikace založené na distribuci

Metody založené na distribuci závisí na jasně definovaných předpokladech ohledně statistických charakteristik obrazových dat. Metody jako maximální věrohodnost (MV) a bayesovské klasifikátory analyzují rozdělení pravděpodobností hodnot pixelů nebo extrahovaných znaků pro určení příslušnosti ke třídě. Při klasifikaci obrazů MV přiřadí každý pixel do třídy, jejíž statistický model nejlépe odpovídá pozorovaným datům. Bayesovská klasifikace naopak používá Bayesovu větu k odhadu pravděpodobnosti, že obraz patří do určité třídy, přičemž využívá jak předchozí informace, tak pozorovaná data. Bayesova věta umožňuje změnit podmíněné pravděpodobnosti kombinováním předchozích pravděpodobností tříd s pravděpodobností pozorovaných znaků, aby se určil nejpravděpodobnější třída pro danou oblast obrazu. Obecně všechny přístupy vyžadují statistické modely pro každou třídu a klasifikují pixely nebo segmenty tím, že odhadují pravděpodobnost jejich příslušnosti ke každé třídě na základě těchto modelů.

Výhodou distribučních metod je jejich matematická srozumitelnost a relativně nízká výpočetní náročnost. Díky explicitnímu modelování pravděpodobnostních rozdělení jsou výsledky klasifikace dobře interpretovatelné. Nevýhodou je citlivost na porušení předpokladů o rozdělení

dat, například normality, a omezená schopnost pracovat s velmi složitými nebo vysoce nelineárními daty. Právě proto metody založené na distribuci často uplatňují v úlohách s homogenními daty, například v dálkovém průzkumu Země nebo lékařském zobrazování, a v moderních systémech bývají nahrazovány nebo doplňovány metodami strojového učení.

Například studie „Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks“ se zaměřuje na detekci obrazů generovaných pomocí GAN modelů s využitím hlubokých neuronových sítí. Autoři v této práci navrhují přístup založený na extrakci pixelových příznaků, které jsou následně zpracovány pomocí konvoluční neuronové sítě. Model je trénován na datasetu obsahujícím jak reálné fotografie, tak synteticky generované obrazy, přičemž cílem je naučit síť rozpoznávat charakteristické artefakty vznikající během generování obrazů (Martin-Rodriguez a kol., 2023).

#### 1.2.4 Metody klasifikace bez distribuce

CNN nabízejí moderní přístup k klasifikaci obrázků, který není závislý na rozdělení, a to prostřednictvím učení znaků přímo z dat, nikoli na základě předem definovaných statistických předpokladů. CNN se skládají z několika vrstev, které postupně identifikují znaky obrazu, od jednoduchých vzorů po složitější struktury. Používají operace jako konvoluce a spojování. Skládání je matematický proces používaný k extrakci znaků ze vstupních obrázků pomocí filtru nebo jádra, které se pohybuje po datech. Slučování také používá operaci filtrování vstupních dat, ale na rozdíl od skládání nezahrnuje učení vážených parametrů. Ačkoli učení CNN vyžaduje velké sady dat s popisky a značné výpočetní zdroje, tento přístup často poskytuje mnohem vyšší přesnost díky schopnosti automaticky učit se hierarchické vlastnosti z nezpracovaných obrazových dat.

Metody bez distribučních předpokladů se vyznačují vysokou flexibilitou a schopností pracovat s nelineárními vztahy mezi obrazovými znaky. Díky tomu jsou vhodné pro analýzu komplexních vizuálních struktur, kde klasické statistické přístupy selhávají. Proto metody klasifikace bez distribuce nacházejí široké uplatnění v moderních systémech počítačového vidění, například při automatickém rozpoznávání objektů, analýze videa nebo identifikaci synteticky generovaného obsahu. Nevýhodou je nižší transparentnost rozhodovacího procesu a obtížnější interpretace výsledků. Přes tato omezení představují metody klasifikace bez distribuce v současnosti dominantní směr vývoje v oblasti rozpoznávání obrazů a tvoří základ většiny pokročilých klasifikačních systémů.

### 1.3 Metody strojového rozpoznávání obrázků

Současné metody strojového rozpoznávání obrázků využívají hluboké neuronové sítě k automatické extrakci příznaků přímo z obrazových dat, čímž překonávají omezení tradičních přístupů založených na ručně navržených znacích. Konvoluční neuronové sítě se staly základem moderních detekčních a klasifikačních architektur, mezi které patří například R-CNN, Fast R-CNN, Faster R-CNN nebo jednostupňové detekční modely typu SSD a YOLO. Tyto architektury se liší zejména způsobem generování kandidátních oblastí a rychlostí zpracování obrazu, což ovlivňuje jejich vhodnost pro různé aplikační scénáře. Významným aspektem hodnocení uvedených

metod je dosažená přesnost, výpočetní náročnost a schopnost generalizace na různorodá obrazová data, což hraje klíčovou roli při aplikaci detekčních modelů v reálných podmínkách i při analýze obrazů generovaných umělou inteligencí.

### 1.3.1 Regionální konvoluční neuronová síť

**R-CNN** (Region-Based Convolutional Neural Network) je dvoustupňový detektor objektů, který nejprve generuje přibližně 2 000 návrhů oblastí pro každý obrázek. Navrhované oblasti jsou zmenšeny na pevnou velikost a poté zpracovány prostřednictvím samostatných sítí pro extrakci a klasifikaci znaků. Každé oblasti je přiřazeno skóre spolehlivosti na základě její předpovězené třídy. Oblasti, které mají vysokou překrývající se hodnotu (IoU) s oblastmi s vyšším skóre, jsou následně potlačeny. Zbývající oblasti s vysokou spolehlivostí, které se nepřekrývají, tvoří konečný výstup modelu.

Nevýhodou původní architektury R-CNN je vysoká výpočetní náročnost, jelikož každý region obrazu je zpracováván samostatně konvoluční neuronovou sítí, což vede k pomalejšímu zpracování ve srovnání s novějšími přístupy. (Ross Girshick a kol., 2014).

**Fast R-CNN** představuje významné zlepšení oproti původnímu modelu R-CNN v oblasti detekce a rozpoznávání objektů v obrazech. Hlavním cílem této metody je zvýšit rychlost zpracování a zároveň zlepšit přesnost klasifikace.

Na rozdíl od R-CNN, který provádí konvoluční výpočty pro každý navržený region zvlášť, Fast R-CNN zpracovává celý vstupní obraz jedním průchodem konvoluční neuronovou sítí. Výsledkem je sdílená konvoluční mapa příznaků, ze které jsou následně extrahovány oblasti zájmu (Ross Girshick, 2015).

Pro převod oblastí s různými rozměry do jednotného tvaru Fast R-CNN využívá Region of Interest Pooling vrstvu, která umožňuje efektivní zpracování regionů bez nutnosti opakovaných výpočtů, čímž výrazně snižuje výpočetní náročnost celého systému. Následně jsou extrahované příznaky předány plně propojeným vrstvám, které provádějí klasifikaci objektů a zároveň regresi ohraničujících boxů.

Další výhodou Fast R-CNN je společné učení klasifikace i lokalizace objektů v rámci jednoho modelu pomocí vícesložkové ztrátové funkce. Architektura Fast R-CNN přináší lepší konvergenci při trénování a vyšší přesnost než předchozí přístupy. Přestože Fast R-CNN stále využívá externí algoritmy pro generování návrhů regionů, představuje důležitý krok směrem k efektivnějším a praktičtějším systémům strojového rozpoznávání obrazů.

#### **Faster R-CNN**

Faster R-CNN představuje další vývojovou generaci R-CNN architektur, která integruje vlastní modul pro generování regionálních návrhů přímo do neuronové sítě, čímž odstraňuje potřebu externích algoritmů pro tvorbu regionů zájmu. Klíčovým prvkem Faster R-CNN je tzv. Region Proposal Network (RPN), která je trénována společně s detekční sítí a generuje regionální návrhy ve formě ohraničujících boxů přímo z konvolučních map příznaků, což značně zrychluje proces detekce (Shaoqing Ren, 2016).

RPN generuje návrhy oblastí zájmu pomocí posuvného okna nad konvoluční mapou příznaků a predikuje možné ohraničující boxy spolu s pravděpodobností, že daná oblast obsahuje objekt.

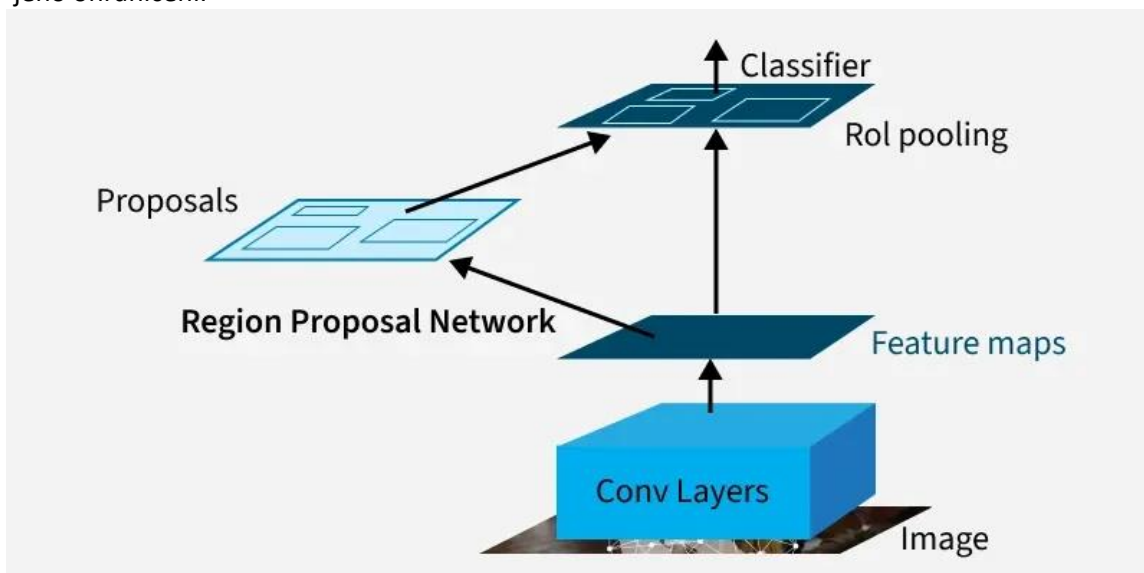
Návrhy jsou následně předány detekční síti, která provádí klasifikaci objektů a zpřesnění jejich lokalizace. Díky sdílení výpočtů mezi RPN a detekční částí je celý proces výrazně efektivnější a rychlejší než u předchozích přístupů.

Faster R-CNN dosahuje vysoké přesnosti při zachování přijatelné výpočetní náročnosti, což umožňuje jeho využití i v praktických aplikacích. Tento model se stal základem pro mnoho dalších detekčních architektur a významně ovlivnil další vývoj v oblasti strojového rozpoznávání obrazů.

Přestože Faster R-CNN přináší významné zrychlení procesu detekce, stále se jedná o dvoustupňový detekční model, což znamená, že z hlediska čisté rychlosti zpracování může být pomalejší než některé jednostupňové detektory (Shaoqing Ren, 2016).

Popis obrázku:

Obrázek znázorňuje architekturu detekčního modelu Faster R-CNN určeného pro lokalizaci a klasifikaci objektů v obraze. Vstupní obraz je nejprve zpracován konvolučními vrstvami, které extrahují mapy příznaků. Na tyto mapy navazuje Region Proposal Network, jež generuje návrhy oblastí s výskytem objektů, které jsou dále normalizovány pomocí Region of Interest Pooling. Výsledné reprezentace jsou následně využity klasifikátorem k určení třídy objektu a zpřesnění jeho ohraničení.



**Obrázek 2: Faster R-CNN**

*zdroj: Faster R-CNN | ML (GeeksforGeeks, 2025)*

### 1.3.2 Jednostupňové detekční architektury

Jednostupňové detektory jsou navrženy tak, aby prováděly detekci objektů přímo nad celým obrazem bez potřeby generování regionálních návrhů. Tento přístup integruje extrakci příznaků, lokalizaci objektů a jejich klasifikaci do jednoho sjednoceného modelu, čímž dochází ke snížení výpočetní náročnosti a umožnění zpracování v reálném čase (Muhammad Yaseen, 2025).

Typickými zástupci jednostupňových detektorů jsou modely jako SSD nebo rodina YOLO, které využívají konvoluční neuronové sítě k simultánní predikci tříd a ohraničujících boxů. Tyto architektury dosahují vysoké rychlosti zpracování a jsou proto vhodné zejména pro aplikace vyžadující real-time detekci, přestože mohou v některých případech vykazovat nižší přesnost než dvoustupňové přístupy (Muhammad Yaseen, 2025).

**YOLO (You Only Look Once)** je rodina jednostupňových architektur pro detekci objektů postavených na Darknetu, open-source frameworku pro konvoluční neuronové sítě. YOLO bylo představeno v roce 2016 a bylo navrženo s důrazem na rychlost, díky čemuž je velmi vhodné pro detekci objektů v reálném čase a často je považováno za nejmodernější detektor. Na rozdíl od R-CNN, které zpracovává extrahované oblasti obrazu prostřednictvím více sítí pro extrakci a klasifikaci znaků, YOLO integruje tyto úkoly do jedné sítě. Zatímco R-CNN generuje přibližně 2 000 návrhů oblastí na jeden obrázek, YOLO předpovídá méně než 100 ohraničujících rámečků. Výsledkem je, že YOLO pracuje výrazně rychleji a produkuje méně falešných pozitivních výsledků v pozadí, i když má tendenci vykazovat vyšší chybu lokalizace. Od svého prvního vydání prošlo YOLO četnými aktualizacemi s průběžnými vylepšeními zaměřenými jak na rychlost, tak na přesnost detekce.

**SSD (Single Shot MultiBox Detector)** představuje jednostupňovou detekční architekturu, která kombinuje lokalizaci objektů a jejich klasifikaci v rámci jediného průchodu neuronovou sítí. Na rozdíl od dvoustupňových metod SSD nevyužívá samostatnou fázi návrhu oblastí zájmu, ale přímo predikuje ohraničující boxy a příslušné třídy objektů nad celým obrazem. Detekce je prováděna pomocí předem definovaných kotevních boxů různých měřítek a poměrů stran, které umožňují zachytit objekty různých velikostí. Architektura SSD využívá konvoluční neuronovou síť jako extraktor příznaků a provádí detekci na více úrovních sítě.

Hlavní výhodou SSD je vysoká rychlost zpracování, která umožňuje použití v aplikacích pracujících v reálném čase. Nevýhodou může být nižší přesnost při detekci velmi malých objektů ve srovnání s dvoustupňovými architekturami. SSD se využívá například v bezpečnostních systémech, autonomních vozidlech, analýze videa nebo mobilních aplikacích.

## 1.4 Limity a výzvy detekce AI-generovaných obrázků

Rostoucí schopnosti generativních modelů zásadním způsobem mění podobu digitálního obrazového obsahu. Moderní generativní systémy jsou schopny produkovat obrazy s vysokou vizuální kvalitou, detaily a přirozenou strukturou, což ztěžuje jejich automatickou identifikaci. Detekce obrazů vytvořených umělou inteligencí se proto stává stále složitější úlohou, která vyžaduje neustálé přizpůsobování používaných metod rychle se vyvíjejícím generativním technologiím (Adrian Lokner a kol., 2024).

Jedním z důležitých omezení současných detekčních přístupů je jejich omezená schopnost generalizace. Detekční modely jsou obvykle trénovány na konkrétních typech syntetických obrazů a často vykazují vysokou úspěšnost pouze v prostředí, které odpovídá trénovacím datům. Při aplikaci na obrazy vytvořené jinými generativními modely nebo novějšími architekturami může docházet k výraznému poklesu přesnosti.

Další významnou výzvou je schopnost generovaných obrazů vyhýbat se detekci. Generativní modely mohou být cíleně upravovány tak, aby minimalizovaly artefakty, na kterých jsou detekční algoritmy založeny. Jednoduché postprodukční operace, jako je změna rozlišení, komprese nebo úprava barevného prostoru, mohou výrazně narušit detekční mechanismy. Existují rovněž postupy založené na adversariálních útocích, které umožňují vytvářet obrazy navržené přímo s cílem obcházet detekční systémy.

Tradiční forenzní metody se často spoléhají na fyzikální a digitální artefakty pocházející z reálného snímacího procesu, jako je senzorový šum, charakteristické obrazové stopy nebo metadata. Tyto signály mohou být u obrazů generovaných umělou inteligencí zcela nepřítomné, maskované nebo uměle napodobeny, čímž výrazně klesá jejich vypovídací hodnota v kontextu autentifikace. S nástupem moderních generativních technik tradiční forenzní důkazy ztrácejí svoji spolehlivost, což podtrhuje potřebu vývoje nových metod automatizované analýzy, jako jsou přístupy založené na hlubokém učení (Luisa Verdoliva, 2020).

Dalším problémem je dostupnost a kvalita datových sad určených pro trénování a testování detekčních modelů. Spolehlivá detekce vyžaduje rozsáhlé, aktuální a vyvážené datové soubory zahrnující široké spektrum generativních technik. Nedostatečná reprezentace nových modelů nebo nevyváženost mezi reálnými a syntetickými obrazy může vést ke zkreslení výsledků a omezené použitelnosti detekčních systémů v reálných podmínkách.

Významnou roli hraje také interpretovatelnost detekčních metod. Mnoho moderních přístupů založených na hlubokém učení funguje jako „černé skříňky“, u nichž je obtížné vysvětlit konkrétní důvody výsledného rozhodnutí. Nedostatek transparentnosti může snižovat důvěru v detekční systémy, zejména v oblastech, kde mají výsledky právní nebo společenské důsledky.

Zvyšující se realističnost syntetického obsahu, rychlý vývoj generativních modelů a možnosti obcházení detekce ukazují na nutnost kombinovat více přístupů, včetně hlubokého učení, frekvenční analýzy a digitální forenziky. Tyto limity zároveň zdůrazňují potřebu dalšího výzkumu a průběžné aktualizace detekčních metod, aby bylo možné reagovat na dynamický vývoj v oblasti generativní umělé inteligence.

## 1.5 Komparativní analýza volně dostupných nástrojů

V oblasti generování a detekce obrazů vytvořených umělou inteligencí existuje široká škála volně dostupných nástrojů, které se liší principem fungování, přesností i způsobem využití. Komparativní analýza umožňuje systematicky porovnat vlastnosti jednotlivých řešení a identifikovat jejich silné a slabé stránky v kontextu praktického využití. Takový přístup je nezbytný pro výběr vhodného nástroje pro experimentální ověření v praktické části práce. V rámci komparativní analýzy byly vybrány pět volně dostupných nástrojů zaměřených na detekci obrazů vytvořených umělou inteligencí, konkrétně Decopy.ai, Illuminarty, Hive Moderation, AI or Not a TruthScan. Výběr nástrojů vychází z jejich dostupnosti, zaměření na detekci syntetického obsahu a možnosti praktického využití v rámci experimentální části práce.

Nástroj **Decopy.ai** představuje řešení založené na hlubokém učení, které analyzuje obrazová data pomocí vzorů v pixelech, strukturách a metadatech. Systém byl trénován na rozsáhlé datové sadě obsahující miliony obrazů generovaných různými modely, jako jsou Midjourney, Stable Diffusion nebo DALL-E. Výstupem je klasifikace obrazu doplněná o pravděpodobnostní hodnocení, přičemž nástroj umožňuje rychlé a uživatelsky jednoduché ověření autenticity obrazu. Výhodou řešení je dostupnost bez registrace a rychlost zpracování; nevýhodou zůstává závislost na trénovacích datech a omezená transparentnost rozhodovacího procesu.

Nástroj **Illuminarty** využívá kombinaci metod počítačového vidění a analýzy obrazu, přičemž poskytuje detailnější výstup než běžné klasifikátory. Kromě pravděpodobnosti AI generování dokáže identifikovat pravděpodobný zdroj generativního modelu a lokalizovat části obrazu,

které vykazují znaky syntetického původu. Výhodou přístupu je vyšší interpretovatelnost výsledků; nevýhodou může být složitější práce s výstupy a delší doba analýzy.

Nástroj **Hive Moderation** představuje komerční řešení využívající neuronové sítě pro detekci syntetického obsahu v obrazech i videích. Systém je navržen pro škálovatelné nasazení v online prostředí a poskytuje klasifikaci založenou na pravděpodobnostním hodnocení. Výhodou řešení je robustnost a širší využitelnost; nevýhodou může být omezený přístup k detailům fungování modelu.

Nástroj **AI or Not** je online nástroj zaměřený přímo na detekci AI-generovaných obrázků. Poskytuje jednoduché a rychlé vyhodnocení obrazu s pravděpodobnostním skóre, přičemž je vhodný pro rychlou kontrolu jednotlivých souborů. Výhodou je dostupnost a intuitivní rozhraní; nevýhodou je nižší interpretovatelnost výsledků, nízká přesnost a omezená možnost dávkového zpracování větších datasetů.

Nástroj **TruthScan AI Image Detector** je pokročilý nástroj pro detekci AI-generovaných obrazů z různých generátorů, včetně Midjourney, DALL-E a Stable Diffusion. Nástroj poskytuje klasifikaci obrazu a pravděpodobnostní skóre, což zvyšuje interpretovatelnost výsledků. Výhodou je vysoká přesnost a možnost analyzovat větší sady dat; nevýhodou je částečně omezený přístup pro veřejnost a nízká transparentnost.

Srovnání uvedených nástrojů ukazuje rozdíly zejména v míře přesnosti, srozumitelnosti a uživatelské přívětivosti. Nástroje založené na hlubokém učení dosahují vysoké rychlosti a automatizace, avšak často postrádají vysvětlitelnost výsledků. Naopak přístupy kombinující více analytických metod poskytují detailnější informace, ale za cenu vyšší složitosti. Výzkumy zaměřené na hodnocení podobných nástrojů zároveň poukazují na skutečnost, že žádný detekční systém nedosahuje stoprocentní spolehlivosti a výsledky mohou být ovlivněny typem obrazových dat i použitým generativním modelem.

Tabulka 1: Komparativní analýza volně dostupných nástrojů

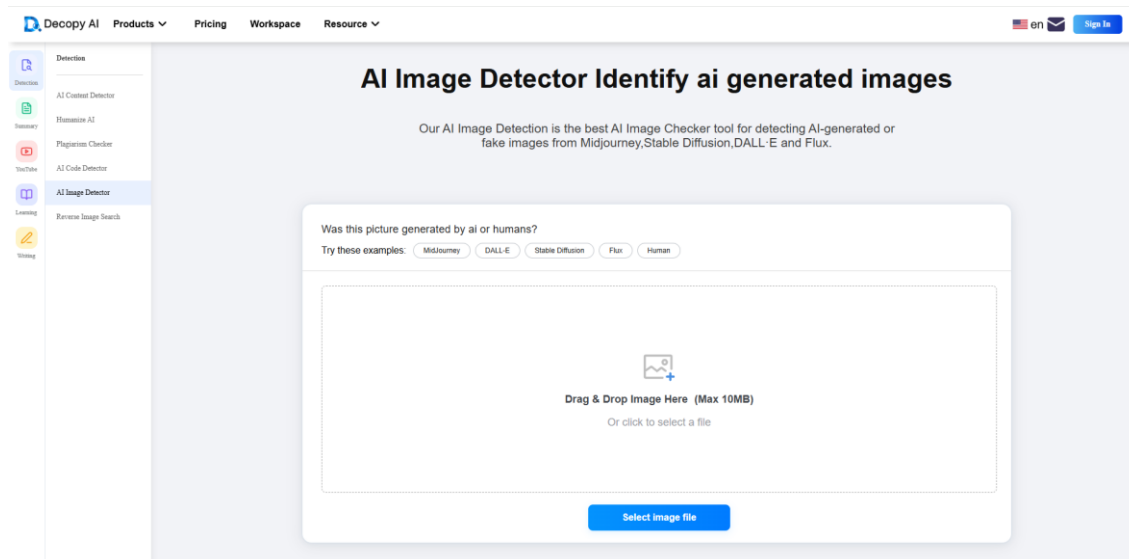
Nástroj	Přesnost	Rychlost	Interpretovatelnost	Uživatelská přívětivost	Automatizace	Transparentnost
<b>Decopy.ai</b>	Vysoká	Rychlá	Vysoká (vysvětlení výstupu, pravděpodobnosti)	Velmi jednoduché	Omezená	Částečná
<b>Illuminarty</b>	Střední	Střední	Střední (vizualizace, lokalizace)	Střední	Omezená	Střední
<b>Hive Moderation</b>	Vysoká	Velmi rychlá	Nízká	Velmi jednoduché	Vysoká	Nízká
<b>AI or Not</b>	Střední	Rychlá	Nízká	Velmi jednoduché	Nízká	Nízká
<b>TruthScan</b>	Vysoká	Střední	Střední (pravděpodobnosti)	Střední	Vysoká	Střední

*Zdroj: vlastní zpracování*

Pro praktickou část práce byl zvolen nástroj **Decopy.ai**, který slouží k analýze obrazů a určení jejich původu, konkrétně k detekci synteticky generovaných obrázků. Volba tohoto nástroje je podpořena několika klíčovými argumenty:

- Prvním důvodem je simulace reálného použití. Decopy.ai je nástroj, který je dostupný pro běžné uživatele bez potřeby programátorských znalostí či instalace složitých knihoven.
- Druhým důvodem je dostupnost a jednoduchost použití. Využití Python notebooků a platform typu Hugging Face by sice umožnilo větší flexibilitu, automatizaci a možnost ladění modelů, nicméně implementace takových řešení by vyžadovala hlubší znalosti programování, přípravu rozsáhlých datasetů a vlastní trénink modelů. Tím by se hlavní cíl práce – experimentálně zhodnotit stávající detekční nástroje – posunul spíše k vývoji modelu než k analýze dostupných řešení.
- Nakonec, nástroj Decopy.ai je sice komerční a jeho vnitřní architektura není uživateli přístupná, nicméně při výstupu poskytuje popis obrazu a odůvodnění klasifikace, čímž nabízí určitou míru transparentnosti. Z tohoto důvodu jej nelze považovat za typický black-box model, protože uživatel získává vysvětlení, proč byl obrázek označen jako syntetický či reálný. Tato vlastnost usnadňuje interpretaci výsledků a podporuje objektivní hodnocení schopnosti nástroje identifikovat AI-generované obrazy.

Celkově použití Decopy.ai odpovídá hlavnímu cíli práce: **experimentálně ověřit schopnost dostupného nástroje identifikovat AI-generované obrazy v různorodém datasetu**, přičemž zohledňuje reálné podmínky nasazení a dostupnost pro uživatele.



**Obrázek 3: Úvodní stránka webu Decopy.ai**  
 zdroj: zdroj: Vlastní zpracování (2026)

### 1.5.1 Limity detekce AI-generovaných obrázků v nástroji Decopy.ai

Nástroj Decopy.ai, přestože využívá pokročilé metody strojového učení, vykazuje limity, které ovlivňují jeho schopnost spolehlivě rozlišovat mezi reálnými a synteticky generovanými obrazy. Jedním z hlavních omezení je závislost modelu na charakteru dat, na kterých byl trénován. V případě obrázků vytvořených méně běžnými generativními modely nebo novějšími verzemi existujících nástrojů může docházet ke snížení přesnosti klasifikace, protože model nemusí být schopen správně identifikovat jejich specifické znaky.

Dalším významným limitem je citlivost na kvalitu vstupních dat. Při zpracování obrázků, které byly upraveny kompresí, změnou rozlišení nebo jinými úpravami, může dojít ke zkreslení výsledků. Takové zásahy mohou narušit jemné charakteristiky obrazu, na jejichž základě Decopy.ai provádí analýzu, což může vést k nesprávné klasifikaci.

Omezení představuje také práce s částečně upravenými obrazy. V případech, kdy je reálná fotografie doplněna o prvky vytvořené umělou inteligencí, může nástroj vykazovat nižší přesnost, protože obraz obsahuje kombinaci reálných a syntetických znaků. Takové situace jsou z hlediska klasifikace složitější a mohou vést k nejednoznačným výsledkům.

## 1.6 Vnitřní principy fungování nástroje Decopy.ai

Nástroj Decopy.ai je založen na využití hlubokého učení pro analýzu obrazových dat a identifikaci znaků typických pro synteticky generované obrazy. Jádrem systému jsou konvoluční neuronové sítě, které umožňují automatickou extrakci příznaků přímo z pixelové reprezentace obrazu. Tyto sítě pracují hierarchicky, přičemž v počátečních vrstvách dochází k detekci základních struktur,

jako jsou hrany a textury, zatímco v hlubších vrstvách jsou identifikovány komplexnější vzory a globální charakteristiky obrazu.

Proces klasifikace vychází z trénování modelu na rozsáhlých datových sadách obsahujících jak reálné fotografie, tak obrazy vytvořené různými generativními modely. Během trénování dochází k optimalizaci váhových parametrů neuronové sítě prostřednictvím minimalizace ztrátové funkce, která vyjadřuje rozdíl mezi predikovanou a skutečnou třídou. Takto natrénovaný model je následně schopen přiřadit novému vstupnímu obrazu pravděpodobnost příslušnosti k jednotlivým třídám.

Důležitou součástí vnitřního fungování je analýza jemných statistických a frekvenčních charakteristik obrazu. Synteticky generované obrazy mohou vykazovat specifické vzory v rozložení pixelů nebo odlišnosti ve frekvenční doméně, které vznikají v důsledku procesu generování. Model využívá tyto odchylky jako diskriminační znaky pro rozlišení mezi reálným a uměle vytvořeným obsahem.

Dalším aspektem je detekce vizuálních artefaktů, které mohou vznikat při generování obrazu. Jedná se například o nepravidelnosti v texturách, nekonzistentní detaily nebo deformace objektů. Přestože tyto znaky nemusí být vždy zřejmé při vizuálním posouzení člověkem, neuronová síť je schopna je zachytit a zahrnout do rozhodovacího procesu.

Výsledkem analýzy je pravděpodobnostní skóre, které vyjadřuje míru jistoty klasifikace. Takový výstup umožňuje kvantifikovat rozhodnutí modelu a využít jej při dalším vyhodnocení pomocí klasifikačních metrik. Vnitřní principy fungování nástroje Decopy.ai tak kombinují pokročilé metody z oblasti hlubokého učení, statistické analýzy a detekce obrazových artefaktů, což umožňuje efektivní identifikaci syntetického obrazového obsahu.

## 1.7 Datové sady pro trénování klasifikátorů a identifikátorů obrázků

Datové sady představují klíčovou součást vývoje, trénování a hodnocení systémů pro strojové rozpoznávání obrazu. Kvalita, velikost a struktura dat ovlivňují, jak dobře jsou modely schopny generalizovat a poskytovat spolehlivé výsledky. V oblasti počítačového vidění se standardizované datové sady a benchmarky používají k objektivnímu srovnání různých metod a algoritmů.

Jednou z nejznámějších obecných datových sad je ImageNet, která obsahuje miliony označených obrázků rozdělených do tisíců kategorií. ImageNet sehrál důležitou roli v rozvoji hlubokého učení a konvolučních neuronových sítí tím, že poskytl dostatek dat pro trénování složitých modelů. Dalším důležitým datovým souborem je Microsoft COCO, který se zaměřuje nejen na klasifikaci objektů, ale také na jejich lokalizaci a segmentaci. COCO obsahuje realistické obrázky s více objekty v různých kontextech, což činí úkoly náročnějšími a blíže situacím v reálném světě.

S rostoucím významem generativní umělé inteligence vznikají také specializované datové sady zaměřené na rozlišování reálných a synteticky generovaných obrazů. Datové soubory obvykle kombinují fotografie pořízené skutečnými snímači zařízeními s obrazy vytvořenými generativními modely. Uvedené datové sady umožňují trénování a testování klasifikátorů určených k identifikaci obsahu vytvořeného umělou inteligencí a hrají klíčovou roli při hodnocení účinnosti detekčních metod.

Významným aspektem při práci s datovými sadami je vyváženost tříd. Nevyvážená data, kde jedna třída výrazně převažuje nad ostatními, mohou vést ke zkresleným výsledkům a snížení schopnosti modelu správně rozpoznávat méně zastoupené kategorie. V kontextu detekce obrazů generovaných umělou inteligencí může nevyváženost způsobit zvýšený výskyt falešně pozitivních nebo falešně negativních klasifikací.

V rámci výzkumu zaměřeného na identifikaci obrazů vytvořených umělou inteligencí je kladen důraz na využívání specializovaných datových sad a nástrojů které jsou cíleně navrženy pro rozlišování mezi reálnými a synteticky generovanými obrazy. Obecné datové sady určené pro klasifikaci objektů nebo scén často neobsahují charakteristiky typické pro AI-generovaný obsah, což může vést k omezené vypovídací schopnosti experimentů v oblasti detekce syntetických obrazů. Zaměření na specializovaná data proto umožňuje přesnější hodnocení schopností klasifikačních modelů v kontextu řešené problematiky.

## 1.8 Datové sady pro testování identifikátorů obrázků

Pro účely praktické části byla vytvořena vlastní rozsáhlá datová sada obsahující celkem 1000 obrazů, přičemž 500 představuje obrazy vytvořené umělou inteligencí a 500 tvoří reálné fotografie pořízené lidmi. Takové vyvážené rozdělení tříd je nezbytné pro objektivní vyhodnocení klasifikačních metod a minimalizuje vliv nerovnováhy dat na výsledky experimentu.

Obrazy byly získány z několika volně dostupných datových sad, především prostřednictvím platformy Hugging Face, která poskytuje přístup k širokému spektru syntetických i reálných obrazových dat. Syntetické obrazy pocházejí z různých generativních modelů, což umožňuje zahrnout do testů různé typy uměle vytvořeného obsahu, zatímco reálné fotografie byly vybírány tak, aby odpovídaly tematickým kategoriím a vizuálním vlastnostem syntetických obrázků.

Datová sada byla rozdělena do pěti tematických kategorií: Lidé, sport, zvířata, jídlo a příroda, přičemž každá kategorie obsahuje 200 obrazů (100 reálných a 100 AI generovaných). Tento přístup umožňuje testovat schopnost detekčních nástrojů pracovat s různými typy obsahu a analyzovat, zda je klasifikace citlivá na konkrétní tematiku či vizuální charakteristiky.

Výběr takového datasetu byl motivován několika důvody:

- **Relevance pro reálné scénáře:** Zahrnutí běžných kategorií odpovídá typům obrazů, se kterými se uživatelé setkávají v praxi, což zvyšuje hodnotu experimentu.
- **Různorodost syntetického obsahu:** Použití více generativních modelů umožňuje testovat detekční nástroje na různých variantách AI obrazů.
- **Vyváženost a kontrola:** Každá kategorie obsahuje stejný počet reálných i AI generovaných obrázků, což zajišťuje spravedlivé hodnocení klasifikátorů.
- **Srovnatelnost a náročnost úlohy:** Tematická shoda mezi reálnými a AI obrazy eliminuje možnost jednoduchého rozhodování na základě vizuálních stereotypů a nutí detekční nástroje analyzovat hlubší charakteristiky snímků.

Výhodou této datové sady je vysoká kontrola nad obsahem, možnost systematického testování různých typů scén a realistické podmínky pro hodnocení detekčních algoritmů. Omezením

zůstává fakt, že dataset stále nepokrývá všechny možné varianty obrazů generovaných AI, nicméně poskytuje dostatečně rozsáhlý a tematicky rozmanitý základ pro experimentální část práce.

## 2 Metodika

Cílem metodické části práce je představit a vysvětlit postup realizace praktické části bakalářské práce. Kapitola popisuje zvolený přístup k řešení problematiky identifikace obrazů vytvořených umělou inteligencí a vymezuje jednotlivé kroky, které budou v rámci experimentu provedeny. Součástí kapitoly jsou informace o vybraném nástroji, způsobu práce s daty a principech jejich zpracování. Dále je popsán návrh metodiky experimentu a stanovení metrik pro vyhodnocení úspěšnosti klasifikace. Uvedené postupy vytvářejí základ pro realizaci praktické části práce a umožňují systematické ověření funkčnosti vybraných detekčních nástrojů.

### 2.1 Nástroj Decopy.ai

Nástroj Decopy.ai představuje systém určený pro detekci obrazů vytvořených umělou inteligencí, který využívá pokročilé metody strojového učení a analýzy obrazových dat. Základní princip spočívá v identifikaci statistických a vizuálních vzorů, které se liší mezi reálnými fotografiemi a synteticky generovanými obrazy. Detekce probíhá prostřednictvím modelů trénovaných na rozsáhlých datových sadách obsahujících obrazy vytvořené různými generativními modely.

Detektor obrázků generovaných umělou inteligencí je bezplatný nástroj pro detekci obrázků vytvořených umělou inteligencí. Datový soubor použitý k trénování našeho detektoru obrázků generovaných umělou inteligencí obsahuje přibližně 10 milionů obrázků, včetně umělých obrázků vygenerovaných různými modely umělé inteligence, jako jsou Midjourney, Stable Diffusion, DALL-E a Flux. Pro zlepšení procesu trénování a testování byl tento datový soubor dále optimalizován, aby náš detektor obrázků generovaných umělou inteligencí mohl efektivně a přesně určit, zda je obrázek generován umělou inteligencí, identifikovat umělecká díla vytvořená umělou inteligencí a odhalit falešné avatary nebo složené fotografie. Pomůže vám určit pravost obrázku. (ZINGKODE LTD, 2026)

Výstupem systému je pravděpodobnostní skóre, které vyjadřuje míru jistoty, s jakou byl obraz klasifikován jako generovaný umělou inteligencí nebo jako reálný. Takový přístup umožňuje flexibilnější interpretaci výsledků a poskytuje uživateli lepší přehled o spolehlivosti rozhodnutí.

#### 2.1.1 Předplatné nástroje Decopy.ai

Pro realizaci praktické části práce bylo nutné využít placenou verzi nástroje Decopy.ai, protože bezplatná verze umožňuje zpracovat pouze 10 obrázků bez registrace a s registrací dalších 20. Vzhledem k tomu, že datová sada obsahuje celkem 1000 obrázků, bylo nezbytné zvolit takový tarif, který umožní analyzovat celý dataset bez omezení.

Pro tento účel bylo zvoleno měsíční předplatné Lite, které poskytuje 1100 analýz (tzv. quotas) za jedno zúčtovací období. Tento limit byl dostatečný pro otestování všech obrazů v datové sadě a zároveň ponechal menší rezervu pro případné opakované testování nebo kontrolu výsledků.

Cena zvoleného předplatného činila 6,99 USD (145.31 Kč), což představuje relativně nízké náklady vzhledem k rozsahu provedeného experimentu. Volba tohoto tarifu byla motivována snahou minimalizovat náklady při zachování možnosti realizovat kompletní experimentální analýzu.

Použití placené verze nástroje zároveň umožnilo stabilní a nepřerušovaný průběh experimentu, což je důležité pro získání konzistentních a spolehlivých výsledků. Tento krok byl tedy nezbytný pro úspěšnou realizaci praktické části práce.

## Subscription Plans & Credits

The screenshot displays a pricing page with two tabs: 'Plans' and 'Premium Credits'. Below the tabs are four subscription cards:

- Free:** \$0 /mo. Monthly Usage: 20 quotas. Premium Credits: 20 credits/mo. Features: Limited basic features, A few trials of premium features. Button: Current Plan.
- Lite:** \$6.99 /mo. Monthly Usage: 1100 quotas. Premium Credits: 800 credits/mo. Credit rate: \$0.0087/credit. Features: More basic features, Get more credits, access more premium features, Can purchase premium credits (never expire). Button: Subscribe Now.
- Basic:** \$14.99 /mo. Monthly Usage: 3000 quotas. Premium Credits: 2000 credits/mo. Credit rate: \$0.0075/credit. Features: More basic features, Get more credits, access more premium features, Can purchase premium credits (never expire). Button: Subscribe Now.
- Premium:** \$29.99 /mo. Monthly Usage: Unlimited. Premium Credits: 5000 credits/mo. Credit rate: \$0.006/credit. Features: Unlimited basic features, Get more credits, access more premium features, Can purchase premium credits (never expire), Unlimited storage space (coming soon). Button: Subscribe Now.

Obrázek 4: Možnosti měsíčních předplatných

zdroj: Vlastní zpracování (2026)

## 2.2 Způsob práce s daty a princip jejich zpracování

Zpracování dat v rámci praktické části práce vychází ze systematického postupu, který zahrnuje přípravu datové sady, její organizaci, samotnou klasifikaci a následné vyhodnocení výsledků. Každý obraz v datové sadě je opatřen odpovídajícím označením určujícím jeho skutečnou třídu, tedy zda se jedná o reálnou fotografii nebo obraz vytvořený umělou inteligencí. Takto připravená data slouží jako referenční základ pro posouzení správnosti klasifikace.

V první fázi dochází k organizaci datové sady, která je rozdělena podle jednotlivých tříd. Obrazy jsou ukládány v jednotném formátu a odpovídající kvalitě, aby nedocházelo k ovlivnění výsledků vlivem technických parametrů. Důraz je kladen na zachování původních vlastností obrazů bez dodatečných úprav, které by mohly zkreslit proces detekce.

Samotné zpracování probíhá prostřednictvím nástroje Decopy.ai. Každý obraz je postupně analyzován a nástroj přiřazuje výstup ve formě klasifikace a pravděpodobnostního skóre. Výsledky jsou zaznamenávány do strukturované tabulky, která obsahuje informace o skutečné třídě a přiřazenému výsledku.

Na základě získaných dat je následně provedeno vyhodnocení pomocí matice záměn. Z hodnot true positive, true negative, false positive a false negative jsou vypočteny základní metriky klasifikace, jako je accuracy, precision, recall a F1-score. Takový postup umožňuje objektivní posouzení výkonosti detekčního nástroje.

Zvolený způsob práce s daty zajišťuje přehlednost, reprodukovatelnost a systematickosti celého experimentu. Jednotný postup zpracování umožňuje porovnání výsledků a identifikaci

případných nedostatků klasifikace, což vytváří základ pro následnou analýzu a diskusi dosažených výsledků.

## 2.3 Návrh metodiky experimentu a stanovení metrik

Návrh metodiky experimentu je zaměřen na ověření schopnosti vybraného nástroje identifikovat obrazy vytvořené umělou inteligencí a rozlišit je od reálných fotografií. Experiment vychází z principu binární klasifikace, při které jsou obrazy rozdělovány do dvou tříd, konkrétně na **synteticky generované obrazy** a **reálné fotografie**. Metodika navazuje na poznatky uvedené v teoretické části práce, zejména v oblasti metod strojového rozpoznávání obrazů a limitů detekce AI-generovaného obsahu.

Základem experimentu je vlastní datová sada obsahující vyvážený počet obrazů obou tříd. Každý obraz je opatřen odpovídajícím označením, které slouží jako referenční hodnota pro následné vyhodnocení. V rámci experimentu dochází k postupnému zpracování jednotlivých obrazů pomocí zvoleného detekčního nástroje, přičemž pro každý obraz je zaznamenán výstup klasifikace. Výsledky jsou systematicky ukládány do strukturované podoby, která umožňuje jejich další analýzu.

Součástí metodiky je také zajištění konzistentních podmínek testování. Všechny obrazy jsou zpracovávány ve stejném formátu a bez dodatečných úprav, aby nedošlo k ovlivnění výsledků. Postup zpracování je jednotný pro všechny položky datové sady, což zajišťuje reprodukovatelnost experimentu a srovnatelnost výsledků.

Vyhodnocení úspěšnosti klasifikace je založeno na využití **matice záměn**, která umožňuje přehledně zachytit správné a nesprávné klasifikace. Na základě této matice jsou určeny hodnoty true positive, true negative, false positive a false negative. Tyto hodnoty tvoří základ pro výpočet vybraných metrik, které poskytují komplexní pohled na výkonnost klasifikačního systému.

	Predicted Positive	Predicted Negative	
Actual Positive	<b>TP</b> <i>True Positive</i>	<b>FN</b> <i>False Negative</i>	<b>Sensitivity</b> $\frac{TP}{(TP + FN)}$
Actual Negative	<b>FP</b> <i>False Positive</i>	<b>TN</b> <i>True Negative</i>	<b>Specificity</b> $\frac{TN}{(TN + FP)}$
	<b>Precision</b> $\frac{TP}{(TP + FP)}$	<b>Negative Predictive Value</b> $\frac{TN}{(TN + FN)}$	<b>Accuracy</b> $\frac{TP + TN}{(TP + TN + FP + FN)}$

**Obrázek 5: Metriky matice záměny**

zdroj: Understanding the Confusion Matrix | by Dany W. (Medium, 2020)

Mezi hlavní metriky patří:

- **Přesnost** (accuracy), která vyjadřuje podíl správně klasifikovaných obrazů z celkového počtu.

$$Accuracy = \frac{True\ Positives(TP) + True\ Negatives(TN)}{Total\ samples\ (TP + TN + FP + FN)}$$

- **Preciznost** (precision), která hodnotí přesnost detekce syntetických obrazů vzhledem k počtu všech obrazů označených jako syntetické.

$$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Positives(FP)}$$

- **Citlivost** (recall) vyjadřuje schopnost systému správně identifikovat všechny syntetické obrazy v datové sadě.

$$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)}$$

- **F1-score** představuje harmonický průměr hodnot precision a recall a poskytuje vyvážené hodnocení klasifikačního výkonu.

$$F1 - score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}$$

Použití **matice záměny** je vhodným způsobem, jak vyhodnotit klasifikační model, protože rozčleňuje různé typy správných a nesprávných rozhodnutí. Při identifikaci obrázků vytvořených umělou inteligencí je důležité nejen znát celkovou přesnost, ale také pochopit, jak k chybám dochází. Matice záměny ukazuje, kolik syntetických obrázků bylo správně identifikováno, kolik bylo nesprávně označeno jako skutečné a kolik skutečných fotografií bylo omylem klasifikováno jako syntetické. To pomáhá analyzovat chování klasifikačního nástroje a poskytuje základ pro interpretaci výsledků experimentu.

Pro hodnocení výkonu detekčních nástrojů byly zvoleny metriky **accuracy, precision, recall a F1 score**, přičemž matice záměn slouží jako základ pro výpočet všech uvedených ukazatelů.

## 3 Praktická část

Praktická část bakalářské práce navazuje na poznatky uvedené v teoretické a analytické části a zaměřuje se na ověření schopnosti detekčních nástrojů identifikovat obrazy vytvořené umělou inteligencí. Hlavním cílem je experimentálně posoudit úspěšnost klasifikace obrazových dat a ověřit funkčnost vybraného nástroje v podmínkách odpovídajících reálnému použití.

V rámci praktické části je využit nástroj Decopy.ai, který slouží k analýze obrazů a určení jejich původu. Experiment je založen na klasifikaci obrazů do dvou tříd, konkrétně na synteticky generované obrazy a reálné fotografie. Pro realizaci experimentu je použita vlastní datová sada obsahující vyvážený počet obrazů obou kategorií, což umožňuje objektivní vyhodnocení výsledků.

Součástí praktické části je popis postupu experimentu, způsobu zpracování dat a vyhodnocení výsledků pomocí klasifikačních metrik. Důraz je kladen na analýzu správných i chybných klasifikací, která umožňuje identifikovat silné a slabé stránky zvoleného nástroje. Výsledky experimentu jsou následně interpretovány v kontextu teoretických poznatků a slouží jako základ pro formulaci závěrů práce.

### 3.1 Datová sada

Pro účely praktické části byla využita vlastní datová sada zaměřená na klasifikaci obrazů podle jejich původu. Datová sada obsahuje celkem 1000 obrazů, z nichž 500 představuje obrazy vytvořené umělou inteligencí a 500 tvoří reálné fotografie pořízené člověkem. Takové rozdělení zajišťuje vyváženost tříd, která je důležitá pro objektivní vyhodnocení klasifikačního experimentu.

Reálné fotografie byly vybírány s důrazem na podobnost scén s obrazy vytvořenými umělou inteligencí. Datový soubor je rozdělen do pěti kategorií: lidé, jídlo, sport, příroda a zvířata, které jsou dále rozděleny na dvě části: reálné obrázky a obrázky vytvořené umělou inteligencí. Tematická podobnost mezi jednotlivými třídami zvyšuje náročnost klasifikační úlohy a omezuje možnost rozhodování na základě jednoduchých vizuálních rozdílů.

Všechny obrazy byly zachovány v původní kvalitě bez dodatečných úprav, aby nedošlo k ovlivnění výsledků experimentu. Takový přístup umožňuje realistické ověření schopnosti detekčního nástroje pracovat s běžnými obrazovými daty. Vytvořená datová sada poskytuje vhodný základ pro realizaci experimentu a umožňuje následné vyhodnocení výsledků pomocí standardních klasifikačních metrik.

Součástí zpracování dat je tabulka, která obsahuje přehled všech obrazů zařazených do experimentu a výsledků jejich klasifikace. Tabulka zahrnuje informace o jednotlivých obrazech, konkrétně jejich pořadové číslo, skutečný typ obrazu a výsledek klasifikace získaný pomocí detekčního nástroje. Typ obrazu určuje, zda se jedná o reálnou fotografii (Real) nebo obraz vytvořený umělou inteligencí (AI), zatímco položka rozhodnutí vyjadřuje klasifikaci provedenou nástrojem.

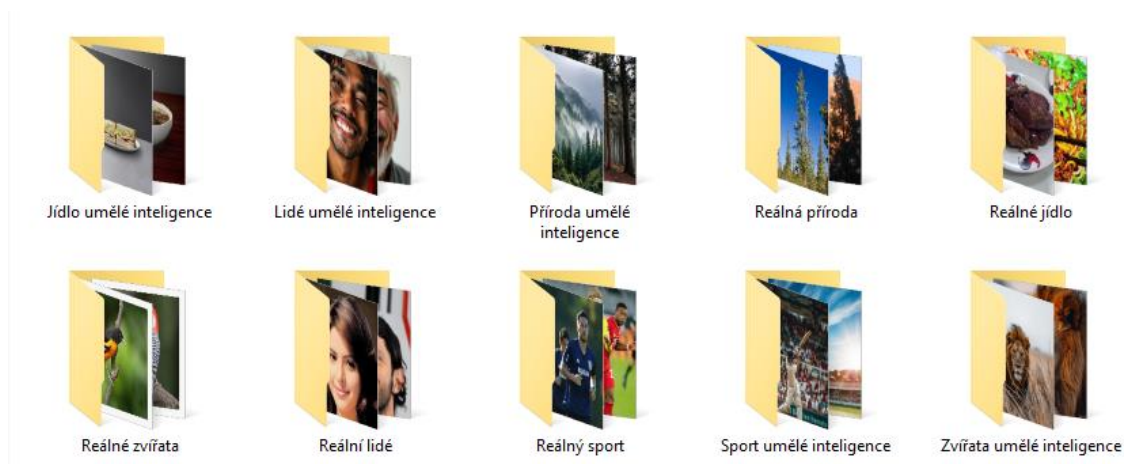
Takto strukturovaná tabulka umožňuje přehledné porovnání skutečných hodnot s predikovanými výsledky a slouží jako základ pro další vyhodnocení experimentu. Na jejím základě je možné sestavit matici záměn a následně vypočítat jednotlivé klasifikační metriky.

Evidence všech obrazů v jedné tabulce zároveň zajišťuje transparentnost a reprodukovatelnost experimentu.

### 3.2 Postup experimentu

Postup experimentu byl navržen s cílem systematicky ověřit schopnost detekčního nástroje rozlišovat mezi reálnými a synteticky generovanými obrazy. Experiment vychází z připravené datové sady a probíhá v několika na sebe navazujících krocích, které zajišťují jednotný a opakovatelný způsob zpracování dat.

V první fázi došlo k přípravě datové sady, která byla rozdělena podle jednotlivých tříd na obrazy vytvořené umělou inteligencí a reálné fotografie. Každý obraz byl opatřen odpovídajícím označením určujícím jeho skutečný původ. Následně byly všechny obrazy připraveny ve vhodném formátu pro zpracování v detekčním nástroji, přičemž nebyly prováděny žádné dodatečné úpravy, které by mohly ovlivnit výsledky klasifikace.



**Obrázek 6: Ukázka souboru dat**

*zdroj: Vlastní zpracování (2026)*

Ve druhé fázi probíhala samotná klasifikace obrazů pomocí nástroje Decopy.ai. Každý obraz byl postupně nahrán do nástroje, který provedl analýzu a přiřadil výstup ve formě klasifikace, tedy označení obrazu jako reálného nebo vytvořeného umělou inteligencí.



V závěrečné fázi došlo ke zpracování získaných výsledků, které byly využity pro sestavení matice záměn a výpočet klasifikačních metrik. Jednotný postup zpracování všech obrazů zajistil konzistentní podmínky experimentu a umožnil objektivní posouzení výkonnosti detekčního nástroje.

### 3.3 Vyhodnocení výsledků

Vyhodnocení výsledků experimentu je založeno na porovnání skutečných tříd obrazů s výsledky klasifikace získanými pomocí detekčního nástroje. Pro přehledné zpracování výsledků byla využita matice záměn, která umožňuje zachytit správné i nesprávné klasifikace v rámci binární úlohy rozlišování mezi reálnými a synteticky generovanými obrazy.

Na základě zaznamenaných dat byly určeny čtyři základní hodnoty, konkrétně true positive, true negative, false positive a false negative. Hodnota true positive představuje počet obrazů vytvořených umělou inteligencí, které byly správně klasifikovány jako syntetické. Hodnota true negative vyjadřuje počet reálných fotografií správně označených jako reálné. False positive představuje případy, kdy byla reálná fotografie nesprávně klasifikována jako obraz vytvořený umělou inteligencí, zatímco false negative označuje syntetické obrazy chybně klasifikované jako reálné.

#### 3.3.1 Vyhodnocení výsledků kategorie jídlo

Tabulka 2: Výsledky z analýzy dat

	Predikce Pozitivní	Predikce Negativní
Predikce Pozitivní	True Positive = 78	False Negative = 22
Predikce Negativní	False Positive = 8	True Negative = 92

Zdroj: vlastní zpracování

Na základě těchto hodnot byly vypočteny základní metriky klasifikace. Metrika accuracy vyjadřuje celkovou úspěšnost klasifikace jako podíl správně klasifikovaných obrazů z celkového počtu. Precision hodnotí přesnost identifikace syntetických obrazů a udává, jak velká část obrazů označených jako syntetické je skutečně správně klasifikována. Recall vyjadřuje schopnost detekovat všechny syntetické obrazy v datové sadě. F1-score představuje kombinovanou metriku, která zohledňuje jak precision, tak recall a poskytuje vyvážené hodnocení klasifikačního výkonu.

Výsledky získané z matice záměny:

- $$Accuracy = \frac{True\ Positives(TP) + True\ Negatives(TN)}{Total\ samples\ (TP + TN + FP + FN)} = \frac{78 + 92}{78 + 92 + 8 + 22} = \frac{170}{200} = 0.85$$
- $$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Positives(FP)} = \frac{78}{78 + 8} = \frac{78}{86} \approx 0.907$$
- $$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)} = \frac{78}{78 + 22} = \frac{78}{100} = 0.78$$

- $F1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} = 2 \times \frac{(0.907 \times 0.78)}{(0.907 + 0.78)} \approx 0.832$

V kategorii **jídlo** dosahuje detekční nástroj hodnoty **accuracy = 0,85**, což znamená, že správně klasifikuje přibližně 85 % obrazů, zatímco zhruba 15 % případů je vyhodnoceno chybně.

Hodnota **precision = 0,91** naznačuje, že většina obrazů označených jako AI generované skutečně spadá do této kategorie, a počet falešně pozitivních výsledků je tedy poměrně nízký.

Naopak hodnota **recall = 0,78** ukazuje, že nástroj nedokáže zachytit všechny syntetické obrazy – přibližně pětina z nich zůstává nerozpoznána a je klasifikována jako reálná. Tento výsledek poukazuje na obtížnost detekce zejména u vizuálně kvalitních a realistických AI obrazů jídla.

Hodnota **F1-score = 0,83** potvrzuje relativně vyvážený výkon modelu, který kombinuje poměrně vysokou přesnost s nižší schopností zachytit všechny pozitivní případy.

Z pohledu klasifikace reálných obrazů je nástroj poměrně spolehlivý, protože jen malá část fotografií byla chybně označena jako syntetická. U AI generovaných obrazů je však patrné, že jejich detekce představuje větší problém, což souvisí s rostoucí kvalitou generativních modelů a jejich schopností napodobit reálné vizuální charakteristiky.

### 3.3.2 Vyhodnocení výsledků kategorie lidé

Tabulka 3: Výsledky z analýzy dat

	Predikce Pozitivní	Predikce Negativní
Predikce Pozitivní	True Positive = 74	False Negative = 26
Predikce Negativní	False Positive = 6	True Negative = 94

Zdroj: vlastní zpracování

Výsledky získané z matice záměny:

- $Accuracy = \frac{True\ Positives(TP) + True\ Negatives(TN)}{Total\ samples\ (TP + TN + FP + FN)} = \frac{74 + 94}{74 + 94 + 6 + 26} = \frac{168}{200} = 0.84$

- $Precision = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Positives(FP)} = \frac{74}{74 + 6} = \frac{74}{80} = 0.925$

- $Recall = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)} = \frac{74}{74 + 26} = \frac{74}{100} = 0.74$

- $F1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} = 2 \times \frac{(0.925 \times 0.74)}{(0.925 + 0.74)} \approx 0.822$

V kategorii **lidé** dosahuje detekční nástroj hodnoty **accuracy = 0,84**, což znamená, že správně klasifikuje přibližně 84 % obrazů, zatímco zbývajících 16 % případů je vyhodnoceno nesprávně.

Hodnota **precision = 0,93** ukazuje, že většina obrazů označených jako AI generované je skutečně syntetického původu, což svědčí o nízkém počtu falešně pozitivních klasifikací. Model je tedy poměrně spolehlivý při potvrzování AI obrazů.

Na druhou stranu **recall = 0,74** naznačuje, že nástroj nezachytí všechny AI generované obrazy, přičemž přibližně čtvrtina z nich je chybně označena jako reálná. Tento výsledek poukazuje na obtížnost detekce zejména u realistických portrétů a scén obsahujících lidi.

Hodnota **F1-score = 0,82** potvrzuje celkově vyvážený výkon modelu, i když s určitým omezením v oblasti úplnosti detekce.

Z pohledu reálných obrazů je klasifikace poměrně přesná, protože pouze malý počet fotografií byl nesprávně označen jako syntetický. U AI generovaných obrazů se však opět projevuje slabší schopnost modelu zachytit všechny případy, což naznačuje, že obrazy lidí patří mezi náročnější kategorie z hlediska detekce.

### 3.3.3 Vyhodnocení výsledků kategorie zvířata

Tabulka 4: Výsledky z analýzy dat

	Predikce Pozitivní	Predikce Negativní
Predikce Pozitivní	True Positive = 86	False Negative = 14
Predikce Negativní	False Positive = 10	True Negative = 90

Zdroj: vlastní zpracování

Výsledky získané z matice záměny:

- $$Accuracy = \frac{True\ Positives(TP) + True\ Negatives(TN)}{Total\ samples\ (TP + TN + FP + FN)} = \frac{86 + 90}{86 + 90 + 10 + 14} = \frac{176}{200} = 0.88$$
- $$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Positives(FP)} = \frac{86}{86 + 10} = \frac{86}{96} \approx 0.896$$
- $$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)} = \frac{86}{86 + 14} = \frac{86}{100} = 0.86$$
- $$F1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} = 2 \times \frac{(0.896 \times 0.86)}{(0.896 + 0.86)} \approx 0.878$$

V kategorii **zvířata** dosahuje detekční nástroj hodnoty **accuracy = 0,88**, což představuje nejvyšší celkovou úspěšnost ze zatím analyzovaných kategorií. Model tedy správně klasifikuje přibližně 88 % obrazů, zatímco chybovost se pohybuje kolem 12 %.

Hodnota **precision ≈ 0,90** ukazuje, že většina obrazů označených jako AI generované je skutečně syntetického původu, i když počet falešně pozitivních případů je o něco vyšší než u kategorií jídlo a lidé.

Na rozdíl od předchozích kategorií dosahuje model v této oblasti také vyšší hodnoty **recall = 0,86**, což znamená, že dokáže zachytit většinu AI generovaných obrazů. Ve srovnání s kategoriemi lidí

(recall = 0,74) a jídlo (recall = 0,78) jde o výrazné zlepšení schopnosti detekce syntetického obsahu.

Tomu odpovídá i hodnota **F1-score**  $\approx$  **0,88**, která potvrzuje vyvážený a celkově lepší výkon modelu v této kategorii.

Z výsledků vyplývá, že obrazy zvířat jsou pro detekční nástroj méně problematické než obrazy lidí nebo jídla. Lze předpokládat, že syntetické obrazy zvířat obsahují více rozpoznatelných artefaktů nebo charakteristických znaků, které model dokáže lépe identifikovat. Naopak u realistických portrétů či scén s jídlem mohou být tyto rozdíly méně výrazné, což ztěžuje jejich detekci.

### 3.3.4 Vyhodnocení výsledků kategorie příroda

Tabulka 5: Výsledky z analýzy dat

	Predikce Pozitivní	Predikce Negativní
Predikce Pozitivní	True Positive = 82	False Negative = 18
Predikce Negativní	False Positive = 4	True Negative = 96

Zdroj: vlastní zpracování

Výsledky získané z matice záměny:

- $$Accuracy = \frac{True\ Positives(TP) + True\ Negatives(TN)}{Total\ samples\ (TP + TN + FP + FN)} = \frac{82 + 96}{82 + 96 + 4 + 18} = \frac{178}{200} = 0.89$$
- $$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Positives(FP)} = \frac{82}{82 + 4} = \frac{82}{86} \approx 0.953$$
- $$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)} = \frac{82}{82 + 18} = \frac{82}{100} = 0.82$$
- $$F1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} = 2 \times \frac{(0.953 \times 0.82)}{(0.953 + 0.82)} \approx 0.882$$

V kategorii příroda dosahuje detekční nástroj hodnoty **accuracy** = **0,89**, což představuje nejlepší celkový výkon mezi doposud analyzovanými kategoriemi. Model tedy správně klasifikuje přibližně 89 % obrazů, zatímco chybovost činí jen 11 %.

Hodnota **precision**  $\approx$  **0,95** ukazuje velmi vysokou přesnost – většina obrazů označených jako AI generované je skutečně syntetického původu, a falešně pozitivní klasifikace jsou minimální (pouze 4 z 200).

Hodnota **recall** = **0,82** naznačuje, že detekční nástroj zachytí většinu AI generovaných obrazů, přičemž menší část (18 obrazů z 100) zůstává nerozpoznána. Tento výsledek je lepší než u kategorií lidé (recall = 0,74) a jídlo (recall = 0,78), ale mírně horší než u kategorie zvířata (recall = 0,86).

Hodnota **F1-score**  $\approx$  **0,88** potvrzuje vyvážený výkon modelu a naznačuje, že kombinace vysoké přesnosti a solidní schopnosti zachytit AI obrazy činí tuto kategorii nejsnáze identifikovatelnou.

Ve srovnání s předchozími kategoriemi se ukazuje, že syntetické obrazy přírody jsou relativně dobře rozpoznatelné, pravděpodobně díky charakteristickým strukturám, texturám a vzorům, které model dokáže spolehlivě analyzovat. Kategorie příroda tak představuje pro detekční nástroj nejméně náročnou oblast z hlediska rozlišení reálného a AI obsahu.

### 3.3.5 Vyhodnocení výsledků kategorie sport

Tabulka 6: Výsledky z analýzy dat

	Predikce Pozitivní	Predikce Negativní
Predikce Pozitivní	True Positive = 84	False Negative = 16
Predikce Negativní	False Positive = 9	True Negative = 91

Zdroj: vlastní zpracování

Výsledky získané z matice záměny:

- $$Accuracy = \frac{True\ Positives(TP) + True\ Negatives(TN)}{Total\ samples\ (TP + TN + FP + FN)} = \frac{84 + 91}{84 + 91 + 9 + 16} = \frac{175}{200} = 0.875$$
- $$Precision = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Positives(FP)} = \frac{84}{84 + 9} = \frac{84}{93} \approx 0.903$$
- $$Recall = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)} = \frac{84}{84 + 16} = \frac{84}{100} = 0.84$$
- $$F1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} = 2 \times \frac{(0.903 \times 0.84)}{(0.903 + 0.84)} \approx 0.870$$

V kategorii sport dosahuje detekční nástroj hodnoty **accuracy = 0,875**, což ukazuje, že model správně klasifikuje přibližně 88 % obrazů, zatímco zhruba 12 % je vyhodnoceno chybně.

Hodnota **precision ≈ 0,90** znamená, že většina obrazů označených jako AI generované skutečně patří do této kategorie, a počet falešně pozitivních klasifikací je relativně nízký (9 z 200).

Hodnota **recall = 0,84** naznačuje, že model dokáže zachytit většinu AI generovaných obrazů, přičemž menší část (16 obrazů z 100) zůstává nerozpoznána. Tento výsledek je lepší než u kategorií lidé (recall = 0,74) a jídlo (recall = 0,78), ale mírně horší než u zvířata (recall = 0,86) a příroda (recall = 0,82).

Hodnota **F1-score ≈ 0,87** potvrzuje vyvážený výkon modelu a dobré spojení mezi přesností a schopností zachytit AI obrazy.

Ve srovnání s ostatními kategoriemi se ukazuje, že obrazy sportovních scén jsou detekčním nástrojem relativně dobře rozpoznatelné, i když drobné vizuální variace a dynamika scén mohou ztížit úplnou detekci. Kategorie sport tedy patří mezi oblasti, kde je nástroj spolehlivý, ale stále existuje prostor pro zlepšení, zejména při zachycení všech syntetických obrazů.

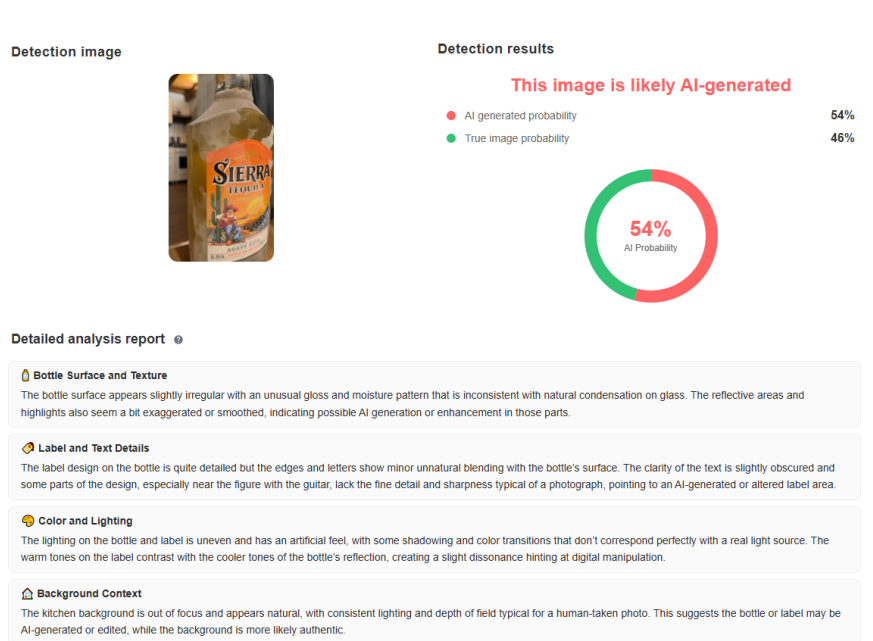
Použití uvedených metrik umožňuje komplexní posouzení výkonnosti detekčního nástroje. Získané hodnoty poskytují nejen informaci o celkové úspěšnosti klasifikace, ale také o schopnosti

nástroje správně identifikovat jednotlivé třídy a minimalizovat chyby. Výsledky slouží jako základ pro následnou analýzu a diskusi, ve které budou podrobněji rozebrány příčiny správných i nesprávných klasifikací.

### 3.4 Analýza chyb klasifikace

Součástí vyhodnocení experimentu je podrobná analýza chyb klasifikace, přičemž se zaměřujeme na případy, kdy byly obrázky zařazeny do nesprávných kategorií. Tyto chyby se v matici záměny projevují jako false positive a true negative. Analýzou těchto chyb můžeme lépe pochopit limity detekčního nástroje a zjistit, kdy má tendenci obrázky nesprávně klasifikovat.

V případě false positive dochází k situaci, kdy je reálná fotografie nesprávně označena jako obraz vytvořený umělou inteligencí. Takové chyby se mohou vyskytovat u obrazů s neobvyklými vizuálními charakteristikami, například při specifickém osvětlení, vysoké míře úprav nebo netypické kompozici. Detekční nástroj může takové vlastnosti vyhodnotit jako znaky syntetického obrazu, což vede k chybnému rozhodnutí. Na příkladu tohoto obrázku vidíme, že nástroj přiřadil skutečnému obrázku nesprávnou klasifikaci, protože jej označil za vytvořený umělou inteligencí. Na obrázku je zobrazena zmrazená láhev s kondenzací na povrchu skla a hlavním důkazem nástroje, že se jedná o obrázek vytvořený umělou inteligencí, bylo to, že povrch láhve byl nerovný a nekonzistentní s podivným odleskem a vlhkostí na povrchu láhve, což není typické pro běžnou kondenzaci na skle.

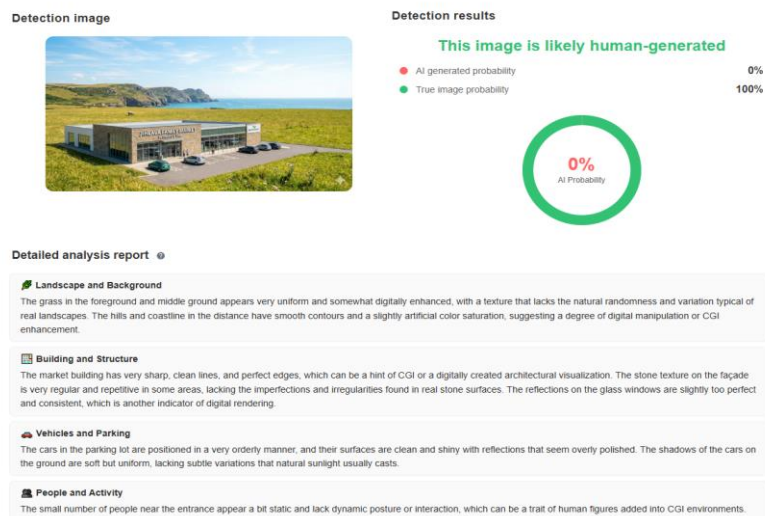


**Obrázek 9: Příklad nesprávné klasifikace obrázku (Skutečný jako AI)**

*zdroj: Vlastní zpracování (2026)*

False negative představují případy, kdy je obraz vytvořený umělou inteligencí nesprávně klasifikován jako reálný. Tyto chyby jsou často způsobeny vysokou kvalitou generovaných obrazů, které obsahují minimum viditelných artefaktů a vykazují realistické detaily. V takových situacích je pro detekční nástroj obtížné odlišit syntetický obsah od skutečné fotografie, zejména pokud obraz odpovídá běžným scénám z reálného života. Na příkladu obrázku níže vidíme, že nástroj jej označil za skutečný, ačkoli byl vytvořen umělou inteligencí. Nástroj vycházel

především z předpokladu, že tento obrázek je s největší pravděpodobností skutečný, avšak upravený grafickými nástroji.



**Obrázek 10: Příklad nesprávné klasifikace obrázku (AI jako skutečný)**

*zdroj: Vlastní zpracování (2026)*

Pro hodnocení schopnosti detekčního nástroje rozlišovat mezi reálnými a AI generovanými obrazy byly vyhodnoceny všechny kategorie z datové sady: **lidé, jídlo, zvířata, příroda a sport**. V tabulce jsou uvedeny hlavní klasifikační metriky pro každou kategorii: accuracy, precision, recall a F1-score.

**Tabulka 7: Výsledky z analýzy dat**

Kategorie	Accuracy	Precision	Recall	F1-score
Lidé	0,84	0,925	0,74	0,822
Jídlo	0,85	0,907	0,78	0,832
Zvířata	0,88	0,896	0,86	0,878
Příroda	0,89	0,953	0,82	0,882
Sport	0,875	0,903	0,84	0,870

*Zdroj: vlastní zpracování*

Analýza výsledků ukazuje několik zajímavých trendů:

- Nejvyšší **accuracy** byla dosažena u kategorií **příroda (0,89)** a **zvířata (0,88)**, což naznačuje, že model dokáže spolehlivě rozlišovat AI obrazy v přírodních a zvířecích scénách.
- Kategorie **lidé** vykazuje nejnižší **recall (0,74)**, což znamená, že detekční nástroj nezachytí přibližně čtvrtinu AI generovaných portrétů a lidských scén. To je pravděpodobně způsobeno vysokou vizuální podobností syntetických a reálných fotografií lidí.
- **Precision** je obecně velmi vysoká napříč všemi kategoriemi (0,90–0,95), což naznačuje, že falešně pozitivní klasifikace jsou minimální a model je konzervativní při označování obrazů jako AI generovaných.

- Hodnoty **F1-score** ukazují, že celkový výkon modelu je nejvyváženější u kategorií **příroda (0,882)** a **zvířata (0,878)**, zatímco kategorie **lidé (0,82)** a **jídlo (0,832)** jsou o něco náročnější na detekci.

### 3.5 Diskuse výsledků

Analýza výkonu detekčního nástroje napříč pěti kategoriemi dat ukazuje, že schopnost modelu odlišit reálné a AI generované obrazy se liší v závislosti na typu scény a charakteristických vizuálních prvcích. Kategorie zvířata a příroda vykazují nejvyšší hodnoty accuracy a F1-score, což naznačuje, že model dokáže efektivně identifikovat typické artefakty generativních algoritmů, jako jsou pravidelné textury, opakující se vzory nebo drobné geometrické odchylky.

Naopak kategorie lidé a jídlo představují pro nástroj větší výzvu. V případě portrétů nebo scén s jídelními kompozicemi jsou vizuální rozdíly mezi syntetickým a reálným obsahem méně výrazné, což se odráží v nižších hodnotách recall a mírně nižších F1-score. To naznačuje, že model někdy nezachytí všechny AI generované obrazy, pravděpodobně kvůli vysoké realistické kvalitě generativních výstupů a variabilitě scén. Kategorie sport vykazuje stabilní výkon, který je mezi extrémy, a ukazuje, že dynamické scény jsou detekčně zvládnutelné, ale mohou obsahovat určité obtížné vizuální prvky.

Celkově výsledky potvrzují, že detekční nástroje založené na hlubokém učení poskytují solidní výsledky napříč kategoriemi, přičemž největší riziko nesprávné klasifikace hrozí u scén s vysokou vizuální podobností reálných a AI generovaných obrazů. Tyto poznatky podtrhují důležitost vyváženého a rozmanitého tréninkového datasetu, který zahrnuje širokou škálu scén a kategorií, aby nástroj byl schopen zvládnout i obtížné případy.

### 3.6 Shrnutí praktické části

Praktická část bakalářské práce byla zaměřena na experimentální ověření schopnosti detekčního nástroje identifikovat obrazy vytvořené umělou inteligencí. Na základě připravené datové sady obsahující reálné a synteticky generované obrazy byl proveden klasifikační experiment, jehož cílem bylo posoudit úspěšnost rozlišování mezi jednotlivými třídami.

V rámci experimentu byla provedena systematická klasifikace všech obrazů a výsledky byly zaznamenány do přehledné struktury umožňující jejich další zpracování. Na základě těchto dat byla sestavena matice záměn a vypočteny základní metriky klasifikace, které poskytly ucelený pohled na výkonnost zvoleného nástroje.

Analýza výsledků ukázala, že nástroj dosahuje dobré celkové přesnosti, avšak vykazuje omezení zejména v oblasti detekce všech syntetických obrazů. Identifikované chyby klasifikace a jejich následná analýza poskytly důležité poznatky o fungování nástroje a vlivu charakteru dat na výsledky.

Získané poznatky tvoří základ pro závěrečnou část práce, ve které budou shrnuty hlavní přínosy, omezení a doporučení pro další výzkum v oblasti detekce obrazů vytvořených umělou inteligencí.

## Závěr

Bakalářská práce byla zaměřena na problematiku identifikace obrazů vytvořených umělou inteligencí, která v současnosti představuje významné téma v oblasti digitálních technologií. Rychlý rozvoj generativních modelů umožňuje vytvářet vizuálně realistický obsah, jehož rozlišení od reálných fotografií je stále obtížnější. Takový vývoj přináší nejen technologické možnosti, ale také výzvy v oblasti důvěryhodnosti digitálního obsahu, ochrany autorských práv a šíření dezinformací.

V teoretické části práce byly popsány základní principy umělé inteligence a generování obrazů, metody klasifikace grafických dat a přístupy strojového rozpoznávání obrazů. Pozornost byla věnována zejména moderním metodám založeným na hlubokém učení, které tvoří základ současných detekčních nástrojů. Součástí teoretické části byla také analýza limitů detekce AI-generovaných obrazů a významu datových sad pro trénování a testování klasifikačních modelů.

Analytická část práce se zaměřila na přehled a porovnání vybraných volně dostupných nástrojů určených pro detekci syntetického obrazového obsahu. Na základě stanovených kritérií byl jako vhodný nástroj pro praktickou část zvolen Decopy.ai, který umožňuje rychlou a přehlednou analýzu obrazů bez nutnosti složité implementace.

Praktická část byla zaměřena na experimentální ověření schopnosti zvoleného nástroje rozlišovat mezi reálnými a synteticky generovanými obrazy. Pro účely experimentu byla vytvořena vlastní datová sada obsahující vyvážený počet obrazů různých kategorií. Klasifikace byla provedena systematickým způsobem a výsledky byly vyhodnoceny pomocí matice záměn a základních klasifikačních metrik.

Získané výsledky ukázaly, že detekční nástroj dosahuje poměrně vysoké celkové přesnosti, avšak vykazuje určité nedostatky zejména v oblasti zachycení všech syntetických obrazů. Analýza chyb klasifikace potvrdila, že moderní generativní modely produkují obrazy s vysokou mírou realismu, což výrazně ztěžuje jejich detekci. Výsledky zároveň poukazují na skutečnost, že úspěšnost klasifikace je ovlivněna nejen kvalitou modelu, ale také charakterem vstupních dat.

Práce rovněž poukázala na širší souvislosti problematiky, zejména na otázky etiky a důvěryhodnosti digitálního obsahu. Schopnost rozlišovat mezi reálným a syntetickým obsahem se stává klíčovou nejen v technologické oblasti, ale také v kontextu společnosti, kde může mít zásadní dopad na vnímání reality a šíření informací.

Na základě provedené analýzy lze konstatovat, že detekce AI-generovaných obrazů představuje komplexní problém, který není možné zcela vyřešit pomocí jednoho nástroje nebo přístupu. Z hlediska dalšího výzkumu se nabízí několik směrů rozvoje. Jednou z možností je využití rozsáhlejších a rozmanitějších datových sad, které by zahrnovaly širší spektrum scén, stylů a typů obrazů. Takový přístup by umožnil přesnější vyhodnocení detekčních nástrojů a lepší pochopení jejich schopností v reálných podmínkách. Další možností je zaměření na specializované datové sady určené přímo pro detekci syntetického obsahu, které obsahují detailně anotovaná data a reflektují aktuální vývoj generativních technologií.

Významný potenciál představuje také kombinace více detekčních přístupů. Využití více nástrojů nebo modelů současně může vést ke zvýšení přesnosti klasifikace, protože jednotlivé přístupy mohou zachytit odlišné charakteristiky obrazových dat. Alternativně lze uvažovat o návrhu

vlastního klasifikačního modelu, který by byl trénován na specificky připravené datové sadě odpovídající konkrétním požadavkům.

Bakalářská práce splnila stanovené cíle, zejména v oblasti ověření funkčnosti vybraného detekčního nástroje a analýzy jeho výsledků. Získané poznatky mohou sloužit jako základ pro další výzkum v oblasti identifikace syntetického obrazového obsahu a přispět k lepšímu pochopení této problematiky.

## Seznam použité literatury

- COLLEGE OF SAINT BENEDICT AND SAINT JOHN'S UNIVERSITY. *Artificial Intelligence and Images*. Online. 2025, aktualizováno 10.24.2025. Dostupné z: <https://guides.csbsju.edu/c.php?g=1297123&p=10164978>. [cit. 2025-11-21].
- GEEKSFORGEES. *Understanding the Confusion Matrix in Machine Learning*. Online. 2026, aktualizováno 01.23.2026. Dostupné z: <https://decopy.ai/ai-image-detector/>. [cit. 2026-03-07].
- GIRSHICK, Ross. Fast R-CNN. Online. In: *Fast R-CNN*. Microsoft Research, 2015, s. 1440 - 1448. Dostupné z: [https://www.cv-foundation.org/openaccess/content\\_iccv\\_2015/papers/Girshick\\_Fast\\_R-CNN\\_ICCV\\_2015\\_paper.pdf](https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Girshick_Fast_R-CNN_ICCV_2015_paper.pdf). [cit. 2026-01-13].
- HIVE. *They said it couldn't be solved. We solved it*. Online. 2026. Dostupné z: <https://hivemoderation.com>. [cit. 2026-03-07].
- IBM. *What are diffusion models?* Online. What are diffusion models? 2023. Dostupné z: <https://www.ibm.com/think/topics/diffusion-models#186915248>. [cit. 2026-01-13].
- IBM. *What is artificial intelligence (AI)?* Online. 2024. Dostupné z: <https://www.ibm.com/think/topics/artificial-intelligence>. [cit. 2025-11-21].
- IBM. *What is image classification?* Online. What is image classification? Dostupné z: <https://www.ibm.com/think/topics/image-classification>. [cit. 2026-01-13].
- IBM. *What is image recognition?* Online. What is image recognition? Dostupné z: <https://www.ibm.com/think/topics/image-recognition>. [cit. 2026-01-13].
- IBM. *What is object detection?* Online. What is object detection? Dostupné z: <https://www.ibm.com/think/topics/object-detection>. [cit. 2026-01-13].
- ILLUMINARTY. *Is an AI behind your image?* Online. 2026. Dostupné z: <https://illuminarty.ai/en/>. [cit. 2026-03-07].
- KRICHEN, Moez. Convolutional Neural Networks: A Survey. Online. *Artificial Intelligence Models, Tools and Applications with A Social and Semantic Impact*. 2023, vol. 2023, no. 12, article 12, s. 42. ISSN 2073-431X. Dostupné z: <https://www.mdpi.com/2073-431X/12/8/151>. [cit. 2026-03-16].
- MARTIN-RODRIGUEZ, Fernando; GARCIA-MOJON, Rocio a FERNANDEZ-BARCIELA, Monica. Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks. Online. *Image Processing and Pattern Recognition Based on Deep Learning—2nd Edition*. 2023, vol. 2023, no. 12, article 23, s. 1-15. ISSN 1424-8220. Dostupné z: <https://www.mdpi.com/1424-8220/23/22/9037>. [cit. 2026-03-16].
- MOHAMMED, Samiyaa. *Architecture review: Two-stage and one-stage object detection*. Online. V12. Elsevier, 2025. Dostupné z: <https://www.sciencedirect.com/science/article/pii/S2773186325001100>. [cit. 2026-01-14].
- REN, Shaoqing; HE, Kaiming; GIRSHICK, Ross a SUN, Jian. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. Online. V3. 2016. Dostupné z: <https://arxiv.org/pdf/1506.01497>. [cit. 2026-01-13].

MEDIUM. *Understanding the Confusion Matrix*. Online. 2020. Dostupné z: <https://medium.com/@danyal.wainstein1/understanding-the-confusion-matrix-b9bc45ba2679>. [cit. 2026-03-07].

VERDOLIVA, Luisa. Media Forensics and DeepFakes: an overview. Online. In: *Journal of Selected Topics in Signal Processing*. Vol. 14, no. 5. IEEE, 2020, s. 910 - 932. Dostupné z: <https://arxiv.org/pdf/2001.06564>. [cit. 2025-11-21].

WANG, Sheng-Yu; WANG, Oliver; ZHANG, Richard; OWENS, Andrew a EFROS, Alexei. *CNN-generated images are surprisingly easy to spot... for now*. Online. V2. CVPR 2020, 2020. Dostupné z: <https://arxiv.org/pdf/1912.11035>. [cit. 2026-01-13].

ZINGKODE LTD. *AI Image Detector Identify ai generated images*. Online. 2026. Dostupné z: <https://decopy.ai/ai-image-detector/>. [cit. 2026-03-07].

## **Přílohy**

Do příloh jsem při elektronickém odevzdání bakalářské práce nahrál dataset.