

# Možnosti použití AI v práci CGI Generalisty

BcA. Martin Siuda

---

Diplomová práce  
2023



Univerzita Tomáše Bati ve Zlíně  
Fakulta multimediálních komunikací

---

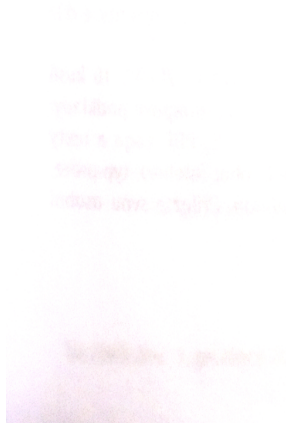
Univerzita Tomáše Bati ve Zlíně  
Fakulta multimediálních komunikací  
Ateliér Audiovize

Akademický rok: 2022/2023

# ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: BcA. Martin Siuda  
Osobní číslo: K21274  
Studijní program: N0211P310005 Teorie a praxe audiovizuální tvorby  
Specializace: Vizuální efekty  
Forma studia: Prezenční  
Téma práce: 1. Teoretická část: Limity využití AI v práci CGI Generalisty  
2. Praktická část: Významná triková spolupráce na magisterském filmu (nezbytná délka závislá na náročnosti technologie po konzultaci a schválení Výrobní komisí AAV), nebo VFX na souboru audiovizuálních děl, oficiálně schváleného před odevzdáním Výrobní komisí AAV ve výstupní kvalitě uvedené ve Výrobní knize AAV, v minimální délce 5 minut, nebo samostatné audiovizuální dílo založené na využití VFX v délce odpovídající námětu a náročnosti technologie v minimální délce 4 minuty. Varianta musí být schválena před odevzdáním Výrobní komisí ateliéru Audiovizuální tvorba. viz Zásady pro vypracování



## Zásady pro vypracování

### 1. Teoretická část:

Rozsah práce: minimálně 30 normostran textu bez započítání obsahu, rejstříku a obrazových příloh.

Formální podoba: Jednotná formální úprava teoretické části práce, její uložení a zpřístupnění se řídí aktuální verzí příslušné směrnice rektora. Student odevzdává 1 ks fyzické (tištěné) práce v pevné vazbě. Tištěná verze práce obsahuje originální „Zadání DP/BP“ včetně příslušných podpisů a studentem podepsané Prohlášení o původnosti práce. Práce v elektronické podobě obsahuje nascanované „Zadání DP/BP“ se všemi formálními náležitostmi a také nepodepsané Prohlášení studenta o původnosti práce. Plný text elektronické verze ve formátu PDF/A a případné přílohy (zkomprimované do jednoho zip souboru) student odevzdá nahráním do IS/STAG a do příslušné složky na NAS-AAV (viz níže).

Pokyny k vypracování: prostudujte a analyzujte dostupné materiály z profesního hlediska a formulujte závěry a získané vědomosti do podoby akademického/odborného textu.

### 2. Praktická část:

Přípustné varianty praktické části:

- 1) Významná triková spolupráce na magisterském filmu (nezbytná délka závislá na náročnosti technologie po konzultaci a schválení Výrobní komisí AAV).
- 2) VFX na souboru audiovizuálních děl, oficiálně schváleného před odevzdáním Výrobní komisí AAV ve výstupní kvalitě uvedené ve Výrobní knize AAV, v minimální délce 5 minut.
- 3) Samostatné audiovizuální dílo založené na využití VFX v délce odpovídající námětu a náročnosti technologie v minimální délce 4 minuty. Varianta musí být schválena před odevzdáním Výrobní komisí ateliéru Audiovizuální tvorba.

Další požadované materiály praktické části:

- a) VFX breakdown („making-of“, „behind-the-scenes“) k předloženému audiovizuálnímu dílu. (var. 1, 2, 3).
- b) Písemná explikace z pohledu dané specializace. Minimální rozsah 2 normostrany (var. 1, 2, 3). c) Anotace (var. 1, 2, 3).
- d) Storyboard, případně animatik (var. 1, 2, 3).
- e) Štábová listina (var. 1, 2).

V případě, že je dílo autorským počinem nebo není součástí praktické části SZZ studenta Produkce, je nutné dodržet doložení požadovaných materiálů a-h dle zadání specializace Produkce. Tato data odevzdává za projekt vždy jeden člověk. Nezbytná je konzultace s vedením AAV.

Všechny odevzdávané materiály musí splňovat vnitřní technické normy dle Výrobní knihy AAV pro odevzdávání prací a musí být řádně popsány (jméno, název, logo fakulty, formát, rozlišení). Součástí závěrečné práce je vytištěný a podepsaný formulář „Údaje o diplomové práci studenta“.

### Uložení na NAS:

Ve složce na NAS-AAV, označené „Bakalářská / Magisterská práce“ uložte:

1. Teoretickou práci ve formátu PDF/A a případné přílohy (zkomprimované do jednoho zip souboru) dle specifikací výše.
2. Vytvořte podsložku Praktická práce, která bude obsahovat materiály částí a- h. Řádně nazvaný film/absolventské dílo odevzdávejte ve formátech splňujících vnitřní technické normy AAV pro odevzdávání prací.
3. Vytvořte podsložku s názvem Katalog, která bude obsahovat „Podklady pro katalog FMK UTB ve Zlíně“: 10 kusů obrazové dokumentace praktické části závěrečné práce pro využití v publikacích FMK. Formát pro bitmapové podklady: JPEG, barevný prostor RGB, rozlišení 300 dpi, 250 mm delší strana. Formáty pro vektory: AI, EPS, PDF. Loga a texty v křivkách. V samostatném textovém souboru uveďte jméno a příjmení, login do portálu UTB, obor (ateliér), typ práce, přesný název práce v češtině i v angličtině, rok obhajoby, osobní e-mail, osobní web, telefon. Přiložte svou osobní fotografii v tiskovém rozlišení.

Rozsah diplomové práce: **viz Zásady pro vypracování**  
Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam doporučené literatury:

RICHARD E., Williams. The Animators Survival Kit. ISBN 9780571238347.  
BIRN, Jeremy. Digital Lighting and Rendering. 2. New Riders PTG. ISBN 9780571238347.  
YOT, Richard. Light for Visual Artists: Understanding & Using Light in Art & Design. Laurence King Publishing; Reprint edition. ISBN 185669660X.  
GRAGE, Pierre. Inside VFX: An Insider's View Into The Visual Effects And Film Business. 2. CreateSpace Independent Publishing Platform, 2014. ISBN 1503349241.

Vedoucí teoretické části: **MgA. Marcel Legindi**  
Ateliér Audiovize  
Vedoucí praktické části: **MgA. Marcel Legindi**  
Ateliér Audiovize

Datum zadání diplomové práce: **1. prosince 2022**  
Termín odevzdání diplomové práce: **19. května 2023**

---

**Mgr. Josef Kocourek, Ph.D.**  
děkan

---

**MgA. Irena Kocí, Ph.D.**  
vedoucí ateliéru

**Ve Zlíně dne 1. prosince 2022**



## PROHLÁŠENÍ AUTORA BAKALÁŘSKÉ / DIPLOMOVÉ PRÁCE

### Beru na vědomí, že

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

### Prohlašuji, že:

- jsem na bakalářské/diplomové práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.

Ve Zlíně dne: .....

Jméno a příjmení studenta: .....

podpis studenta

## **ABSTRAKT**

Tato diplomová práce se zabývá zkoumáním možností, které má generalista v rámci nástrojů založených na umělé inteligenci. Práce má dvě části, teoretickou a praktickou. V teoretické části se budu věnovat základům nutným k pochopení konceptů AI, vysvětlení pojmů a technologií spolu se srovnáním konvenčních metod s novými možnostmi AI. Dále poté srovnám oblasti, ve kterých vzniká AI řešení nahrazující určité chybějící expertízy, kterými generalista oproti specialistům většinou nedisponuje, nebo disponuje jen v omezeném rozsahu. Ve druhé části tyto teze a tvrzení podložím na praktickém experimentu a případové studii, kdy budu mapovat jednotlivé kroky a srovnávat AI metody s konvenčními postupy.

Cíl práce je prozkoumat a definovat rozsah možností, kterými CGI generalista dnes může v rámci AI disponovat, zjistit, jestli mohou zvýšit jeho efektivitu a případně najít jejich slabá místa.

Klíčová slova: umělá inteligence, generalista, CGI, ai, VFX, komunikace, workflow, problemsolving, analýza, nástroje, pipeline, test, postupy

## **ABSTRACT**

This diploma thesis focuses on exploring the possibilities that a generalist has within tools based on artificial intelligence. The work consists of two parts: theoretical and practical. In the theoretical part, I will cover the basics necessary for understanding AI concepts, explaining terms and technologies, along with comparing conventional methods with new AI possibilities. Furthermore, I will compare areas where AI solutions emerge to replace certain missing expertise, which generalists usually lack or possess only to a limited extent, compared to specialists. In the second part, I will support these theses and statements with a practical experiment and case study, mapping individual steps and comparing AI methods with conventional procedures.

The aim of the work is to explore and define the range of possibilities that a CGI generalist can access within AI today, determine if they can increase their efficiency, and potentially identify their weaknesses.

Keywords: artificial intelligence, generalist, CGI, ai, VFX, communication, workflow, problemsolving, analysis, tools, pipeline, test, procedures

Prohlašuji, že odevzdaná verze bakalářské/diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

# OBSAH

<b>ÚVOD.....</b>	<b>11</b>
<b>I TEORETICKÁ ČÁST .....</b>	<b>12</b>
<b>1 HISTORICKÝ KONTEXT AI.....</b>	<b>13</b>
1.1 PRVNÍ AI TEORIE Z HLEDISKA VĚDY A TECHNOLOGIE.....	14
1.1.1 Komunikace pomocí symbolů.....	14
<b>2 AI JAKO NÁSTROJ V CGI TVORBĚ .....</b>	<b>16</b>
2.1 CGI VE FILMU, BRUTE FORCE METODA.....	16
2.1.1 Specialisté a technologické bariéry .....	17
2.2 AI, DEEP LEARNING, MACHINE LEARNING. HIERARCHIE POJMŮ.....	19
2.2.1 AI - Artificial Inteligence.....	19
2.2.2 Machine Learning .....	19
2.2.3 Strukturovaná data.....	19
2.2.4 Latentní prostor .....	20
2.2.5 Deep learning .....	20
2.2.6 Neuronová síť.....	20
2.2.7 Typy Machine learning algoritmů.....	21
2.2.8 Supervised Learning.....	21
2.2.9 Unsupervised Learning .....	21
2.2.10 Reinsforcement Learning .....	22
2.3 MOŽNOSTI VYUŽITÍ VE VFX .....	22
2.3.1 Generování obrazu .....	22
2.3.2 Konvoluční filtr .....	23
2.3.3 Metoda Stabilní Difuze .....	24
2.3.4 Prompt a jeho optimalizace .....	25
2.3.5 Datasetsy – tréninková data .....	26
2.3.6 AI jako nástroj pro optimalizaci HW výkonu .....	27
2.4 PROGRESIVNÍ KROKY VS NAHODILÁ ITERACE (INTUITIVNÍ SKOKY).....	28
2.5 ČAS .....	29
2.5.1 Repetitivní činnosti v tvorbě .....	30
2.6 GENERALISTA/SPECIALISTA.....	30
2.7 AKTUÁLNÍ ALTERNATIVNÍ AI METODY VE VFX INDUSTRY .....	31
2.7.1 Rotoscoping.....	31
2.7.2 Retuše .....	33
2.7.3 Concept art, matte painting .....	34
2.7.4 Charakterová Animace .....	36
2.7.5 Motion Capture .....	36
2.7.6 Texturování, de-lighting a tvorba materiálů.....	38
2.7.7 Depht mapy .....	40
2.7.8 Optimalizace výkonu.....	42
2.7.9 Simulace (Fluid, volume, particles) .....	44

2.7.10	Facial replacement, de-aging.....	45
2.7.11	3D scanning.....	48
2.7.12	3D modelování.....	51
2.7.13	Psaní scriptů.....	52
<b>3</b>	<b>OMEZENÍ AI NÁSTROJŮ.....</b>	<b>54</b>
3.1	KONTROLA.....	54
3.1.1	Omezené UI.....	54
3.1.2	Flickering.....	54
3.1.3	Trénování vlastních modelů.....	55
3.1.4	Datové přenosy.....	55
<b>II</b>	<b>PRAKTICKÁ ČÁST.....</b>	<b>56</b>
<b>4</b>	<b>ZRYCHLENÍ PRÁCE CGI GENERALISTY POMOCÍ AI NÁSTROJŮ .....</b>	<b>57</b>
4.1	ZADÁNÍ EXPERIMENTU.....	57
4.1.1	Cíl experimentu.....	57
4.2	METODOLOGIE EXPERIMENTU.....	57
4.2.1	Parametry měření.....	57
4.2.2	Slabiny a odchylky experimentu.....	57
4.3	PŘÍPADOVÉ STUDIE.....	58
4.3.1	Rendering: AI post process motion blur.....	58
4.3.2	Rotoscoping.....	60
4.3.3	Retušování.....	62
4.3.4	AI upscale obrazu.....	64
4.3.5	Tvorba posteru – snaha kontrolovat výstup generativního AI Midjourney.....	65
	<b>ZÁVĚR .....</b>	<b>67</b>
	<b>SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK.....</b>	<b>71</b>
	<b>SEZNAM OBRÁZKŮ .....</b>	<b>72</b>
	<b>SEZNAM TABULEK.....</b>	<b>75</b>



## ÚVOD

Odvětví vizuálních efektů a počítačem generované grafiky obecně se vyvíjí neuvěřitelnou rychlostí. Od dob, kdy bylo nutné výtvarníků pro výrobu reálných kulis, malířů pro vytváření dokreslovaných rozšiřujících scén na sklo, specialistů na pyrotechniku a efekty uvnitř kamery se obor postupně přesunul do doby, kdy tyto profese postupně vytlačily jejich digitální alternativy. Ruční práce se stala spíše raritou, a většina triků a figlů vzniká v postprodukcí, kde na proces dohlíží teamy specialistů zaměřujících se na jednotlivé digitální složky vizuálních efektů. U menších projektů se nezdá, že team těchto specialistů nahradí několik málo široce zaměřujících se generalistů, kteří jsou díky komplexnímu mixu znalostí a schopností schopni odbavovat méně komplexní projekty za využití menších nákladů. Díky postupujícímu vývoji technologií se stále rozšiřuje výčet specializací, které je schopen pokrýt jeden člověk. Stále zde však existují oblasti, které vyžadují určitou dávku řemeslné zdatnosti a nelze je jednoduše nahradit pomocí konvenčních metod.

V této práci se pokusím zaměřit na možnosti AI, které můžou generalisté využít jako multiplikátory svých schopností, a konkurovat specializovaně zaměřeným teamům ve velkých studiích. Pokusím se zde zmapovat a otestovat oblasti, ve kterých lze AI využít a odhalit případné nedostatky, které z tohoto využití mohou vyplývat.

## **I. TEORETICKÁ ČÁST**

## 1 HISTORICKÝ KONTEXT AI

Téma umělé inteligence provází lidstvo již nějakou dobu. První dochované odkazy naznačující dlouhodobou touhu člověka pochopit principy myšlení, a vytvořit vlastní verzi tohoto složitého systému sahají jak jinak než do Antického Řecka. Tehdejší filozofové se intenzivně zabývali zkoumáním myšlenkového procesu, který popisovali jako mechanickou manipulaci se symboly. Myšlenka vycházela z toho, z toho, že lidské uvažování je možné kvantifikovat.

Umělá inteligence byla také součástí mytologie, kdy Zeus vytvořil měděného obra Talóse aby střežil pro krále ostrov Kréta. Ve středověku se velmi osobitě stejným tématem zabývali alchymisté. Švýcarský alchymista Paracelsus ve své knize Nature of Things popisuje postup, jak vytvořit umělého člověka. Postup spočívá ve smíchání koňského trusu a spermatu ze kterého po čtyřiceti dnech vyroste umělá bytost. I v Česku existuje legenda, ve které na konci šestnáctého století pražský rabín Jehuda Low ben Becalel vytvořil hlíněného obra golema, aby chránil židovskou čtvrť před útoky křesťanů. V dnešní novodobé historii se fascinace umělou inteligencí přetavila do popkultury, kdy vznikají příběhy o strojích, které svou inteligencí předčí lidstvo a způsobí jeho pád. Skynet který odpálí jaderné rakety a spustí soudný den, Matrix, ve kterém jsou lidé zotročeni stroji a slouží jim jako zdroj energie a podobně<sup>1</sup>.



Obrázek 1 Talós (2023, Midjourney)

---

<sup>1</sup> McCorduck, P. Machines who think: a personal inquiry into the history and prospects of artificial intelligence. místo neznámé : MA: A.K. Peters, 2004. ISBN 1568812051.

## 1.1 První AI teorie z hlediska vědy a technologie

V sedmnáctém století Gottfried Wilhelm Leibniz, Thomas Hobbes a René Descartes zkoumali možnost, převedení racionálních myšlenek do systému podobném matematice či geometrii.

Tito Filozofové položili svou hypotézou základní kámen výzkumu umělé inteligence. Na jejich myšlenky později navázal Allen Newell a Herbert Alexander Simon, kteří je také dále rozvedli v rámci své hypotézy PSSH (Physical symbol system hypothesis). PSSH říká, že pokud definujeme inteligenci skrze schopnost, pracovat se symbolovým systémem, mohou být stroje po splnění této podmínky také považovány za inteligentní. PSSH definuje několik druhů symbolů, jako Logické (a, nebo, pro všechny x platí y...), Algebraické (+, -, x, nebo číselné jednotky nutné pro práci s matematickými vzorci), nebo například počítačové symboly (ano/ne – 1/0).

-Inteligentní lidské myšlení: symboly zakódovány v našem mozku, vzorce a výrazy jsou myšlenky, procesy jsou mentální operace myšlení.

-Umělá inteligence: symboly jsou data, vzorce a výrazy jsou vztahy mezi daty, procesy jsou manipulace s daty.

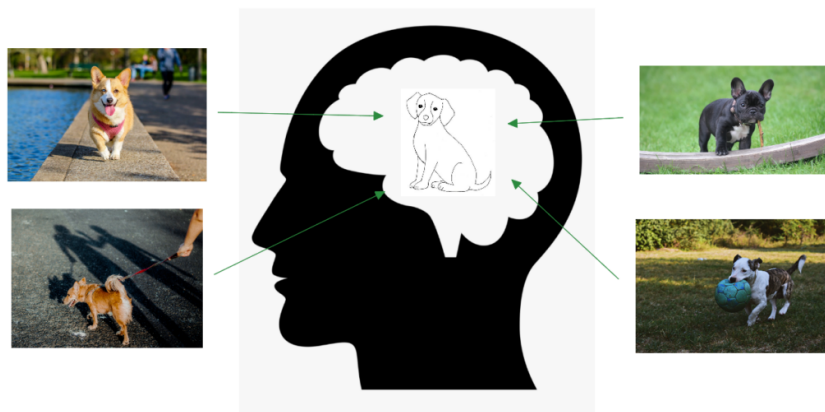
Na tuto hypotézu navazuje další (David Touretzky a Dean Pomerleau), která rozlišuje symboly a signály. Tato verze hypotézy ukazuje, že jakýkoli symbol je možné převést do signálu reprezentovaného binárním kódem (1/0 – ano/ne), a je tedy možné symboly digitalizovat, a vytvářet tak (teoreticky) digitální simulace čehokoliv včetně chování živých organismů.

Pro nás jsou symboly reference objektů ve fyzickém světě. Můžeme použít například slovo židle, které reprezentuje určitou věc, kterou známe z fyzického světa. Slova můžeme řetězit za sebe, a vytvářet tak vzorec, který tvoří komplexnější symboly.

### 1.1.1 Komunikace pomocí symbolů

Slovo židle není pro komunikanta (Osoba přijímající sdělení) jasný symbol. Je to pro něj více či méně přesná oblast v latentním prostoru, který s určitou přesností utvoří představu v mozku komunikanta. Představí si židli, ale to je stále poměrně obsáhlý pojem. Židle jsou různé, jde tak o to, vytvořit řetěžený shluk symbolů, který tuto představu zúží. Například "Červená plastová houpací židle pro děti". Díky komunikaci a řetězení symbolů se prostor chyby v mysli komunikanta postupně zmenšuje. V dalších kapitolách této práce začne být

podobnost mezi principem mezilidské komunikace se způsobem kterým pracují AI algoritmy zřetelnější.<sup>2</sup>



Obrázek 2 Symbolová komunikace

---

<sup>2</sup> Deacon, Terrence W. The Symbolic Species: The Co-evolution of Language and the Brain. místo neznámé : W. W. Norton & Company, 1998. 0393317544.

## 2 AI JAKO NÁSTROJ V CGI TVORBĚ

### 2.1 CGI ve filmu, brute force metoda

Od doby první myšlenky chytrých strojů se generaci od generace výkon počítačů i složitost výpočtů značně posouvaly, a v Osmdesátých letech dvacátého století se začaly počítače využívat pro generování filmové grafiky (CGI). První Snímek, kde bylo CGI použito byl film *Westworld* (1973). Zde se jednalo o dnes poměrně jednoduchý trik rozpixelování obrazu. Za tímto efektem je ale jistě na svou dobu velice složitý výpočet, který nám bude sloužit jako odrazový můstek k vysvětlení toho, jak současné programy na vizuální efekty zpracovávají data a jakým způsobem tvoří nová výstupní, (většinou obrazová) data.



Obrázek 3 Efekt rozpixelování z filmu *Westworld* (1973, Michael Crichton)

Kdybych měl tento efekt – rozpixelování digitálního obrazu přeložit do kroků, které vyjádřím slovy, jednalo by se o proces, kde program vybere podle zadání každý  $x$ -tý pixel, poté identifikuje odstín pixelu na barevném spektru, a poté tuto informaci vyšle ke zpracování do výstupu, kde zobrazí  $x$ -krát redukované množství pixelů které je průměrem hodnot těchto pixelů.

Za tento efekt můžeme v grafických editorech řetězit další efekty s tím, že kdykoliv se má zpracovat nový snímek, musí hardware pomocí hrubé síly spočítat řetězec jednotlivých úkonů.

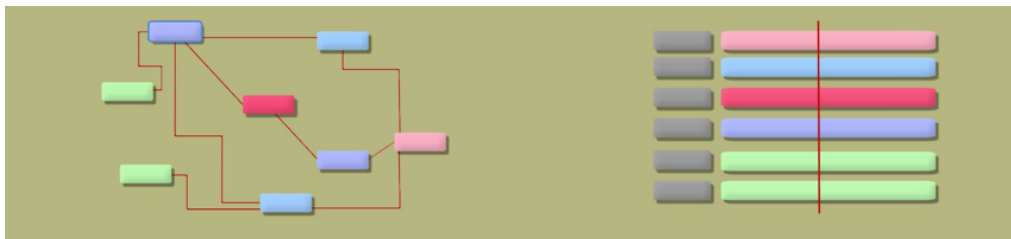
Postupem času se technologie zdokonalují a hrubý výkon počítačů stále roste, podobně jako se zefektivňuje způsob řešení problémů. Využívá se například grafická akcelerace, nebo paralelní řešení několika problémů současně, přesto při rostoucí náročnosti operací



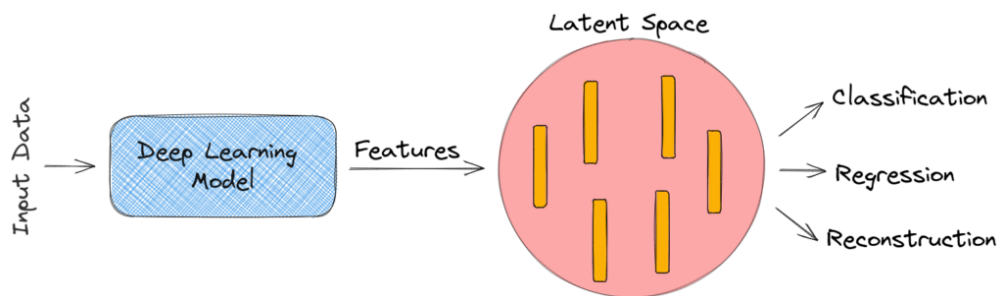
exponenciálně roste čas potřebný na jejich provedení, protože je nutné všechny kroky opakovaně provést.

Podobný princip platí v CGI pro jakýkoliv proces. Současné softwary na tvorbu 2D i 3D generované grafiky řeší úkoly totožným způsobem, ať se jedná o path tracing paprsků světla v 3D prostoru při renderu, simulaci kapalin nebo přidání chromatické aberace či vinětace do výsledného obrazu.

Existují různé způsoby, jakými je možné řetězit a optimalizovat úkony, aby byl proces efektivnější – nodové či layered systémy nebo cachování neměnných fragmentů .



Obrázek 5 Srovnání schématu nodového a layered systému



Obrázek 4 Schéma deep learning procesu

Všechny programy musí jednotlivé kroky při změně opakovaně vypočítávat a nejsou schopny je intuitivně přeskakovat a tvořit nematematické predikce jako lidská mysl nebo AI softwary kterým se budu věnovat v dalších částech práce.<sup>3</sup>

### 2.1.1 Specialisté a technologické bariéry

Od dob, kdy specializovaní umělci ručně dokreslovali sety na sklo, které se poté napozicovalo před kameru, aby vznikl dojem rozdílného prostoru, se díky CGI technologiím výroba přesunula do uzavřených studií, kde dotváření efektů probíhá za pomoci sofistikovaných grafických editorů na výkonném hardwaru. Tyto editory násobně

<sup>3</sup> Brinkmann, Ron. The Art and Science of Digital Compositing: Techniques for Visual Effects, Animation and Motion Graphics. místo neznámé : Morgan Kaufmann, 2008. 0123706386.

zefektivnily proces výroby triků a přidaly množství nových možností od dnes banálního zálohování kroků a využití CTRL+Z pro krok zpět, po složité simulační softwary umožňující snímek za snímkem vytvořit masivní simulace explozí, davů, digitální reprezentace v podstatě čehokoliv myslitelného.

Tyto možnosti ale přidaly nový aspekt. Každým dalším softwarem vzniká bariéra, komplexní filtr, mezi umělcem a dílem, kterou musí umělec-specialista překlenout svou znalostí. Například ovládat tužku je technicky mnohem snadnější než ovládat 3d animační software. Umělec – specialista tak musí pokrýt nové spektrum znalostí, aby využil tuto nabízenou efektivitu, jejíž růst nějakým způsobem koreluje s automatizací jednotlivých kroků.



Obrázek 6 Malování dokreslovačky na sklo ve filmu Star Wars VI (1983, George Lucas)

V současnosti je software a hardware sofistikovanou kalkulačkou, pomocí které může její operátor tvořit výsledný obraz. VFX specialista využívá svých znalostí k tomu, aby klikal na správná tlačítka a progresivně tvořil výsledný obrazový výstup. Stroj nezná například způsob, jakým se pohybuje zebra, a tak je nutný přítomný specialista animátor, který tuto znalost pokrývá. Specialisté zadávají stroji dílčí úkoly, protože stroj sám nezná zadání ani limity, ve kterých má „tvořit“. Animátor tedy musí obsáhnout jak znalosti softwaru, tak znalosti charakteristických pohybů zebry a pravidel s animací souvisejících.

Limity dnes využívaných konvenčních softwarů v kontextu této práce jsou tedy dva: nejsou schopny přeskakovat logické kroky, musí je tudíž opakovaně počítat pomocí matematických algoritmů a hrubého výkonu počítače. A druhý, který s tím nepřímo souvisí a vychází z jejich matematické podstaty – nejsou schopny se držet abstraktních, na matematiku těžko převoditelných pravidel (kompozice, anatomie, způsob pohybu...).

## 2.2 AI, deep learning, machine learning. Hierarchie pojmů

Dříve, než se podrobněji zaměřím na využití AI metod ve VFX, je nutné si definovat několik pojmů a vztahů mezi nimi.

### 2.2.1 AI - Artificial Intelligence

AI je pojem zahrnující druhy softwaru, které jsou schopny jednat autonomně.

### 2.2.2 Machine Learning

Machine Learning je způsob, kterým je možné řešit problémy na základě automatického zpracovávání dat a standartních algoritmů. Pracuje se strukturovanými daty jako gender, rasa, barva, velikost. Machine learning probíhá na základě zobecňování/generalizace. V případě, že chceme stroj naučit rozpoznávat jablka od hrušek, stroj si musí pomoci trénovacími daty hrušek a jablek najít parametry, které tyto dva druhy ovoce od sebe odlišují. Jako rozdílné parametry může sloužit například barva a tvar. Většina jablek je červených, a hrušky bývají oproti jablkům šišaté. Na graf si stroj tedy vytyčí dva parametry: červenost a šišatost a objekty podle nich rozřazuje do správných směrů. Parametrů může být téměř nekonečně mnoho, a tak nám vzniká více rozměrný graf.

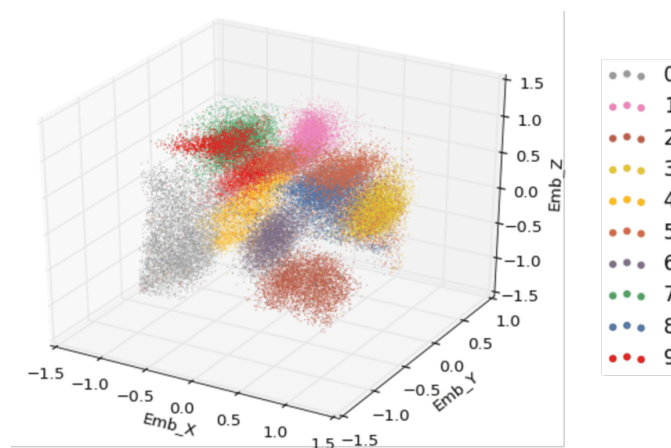
Pro praktický příklad, například Google používá technologie Machine learningu, ke správnému odhadnutí uživatele pomocí jeho geografické polohy a dalších dat a správnému vyhledávání, zpracovávání biometrických dat a vyvozování zdravotního stavu třeba u nositelné elektroniky apod. Jedná se tedy v nadsázce o velice pokročilou samostatně se regulující statistiku. Ta na základě vstupních dat umísťuje parametry těchto dat v latentním prostoru a na základě souřadnic vyvozuje soudy. Je to abstraktní síť na sebe navázaných korelací.

### 2.2.3 Strukturovaná data

Strukturovaná data jsou data, která mají jasně dané atributy. Nejčastěji existují ve formě seznamů, popisů a jasně daných informačních atributů. Nestrukturovaná data jsou naopak data vyžadující nějakou formu pre-processingu (úprav dat). Například texty nebo obrázky.

### 2.2.4 Latentní prostor

Latentní prostor je koncept, který ve strojovém učení označuje skrytý význam dat. V latentním prostoru se data zobrazují jako bod v prostoru s více dimenzemi, kde každá dimenze odpovídá určitému atributu nebo vlastnosti. Tyto atributy mohou být například barva, tvar nebo textura v případě obrázků, nebo hodnoty v případě tabulky s číselnými daty. Tyto atributy jsou zobrazeny jako souřadnice v latentním prostoru, což umožňuje porovnávat a klasifikovat příklady na základě jejich vzájemných vztahů a podobností.



Obrázek 7 Latentní prostor

### 2.2.5 Deep learning

Deep learning je podmnožinou Machine learning algoritmů. Součástí deep learning programů jsou mnohvrstevnaté neuronové sítě. Deep learning se používá v případech, kdy je potřeba abstraktnějších procesů (např. rozpoznávání obrazu, zvuku). Dokáže se opakovaním sám zdokonalovat (Unsupervised learning).

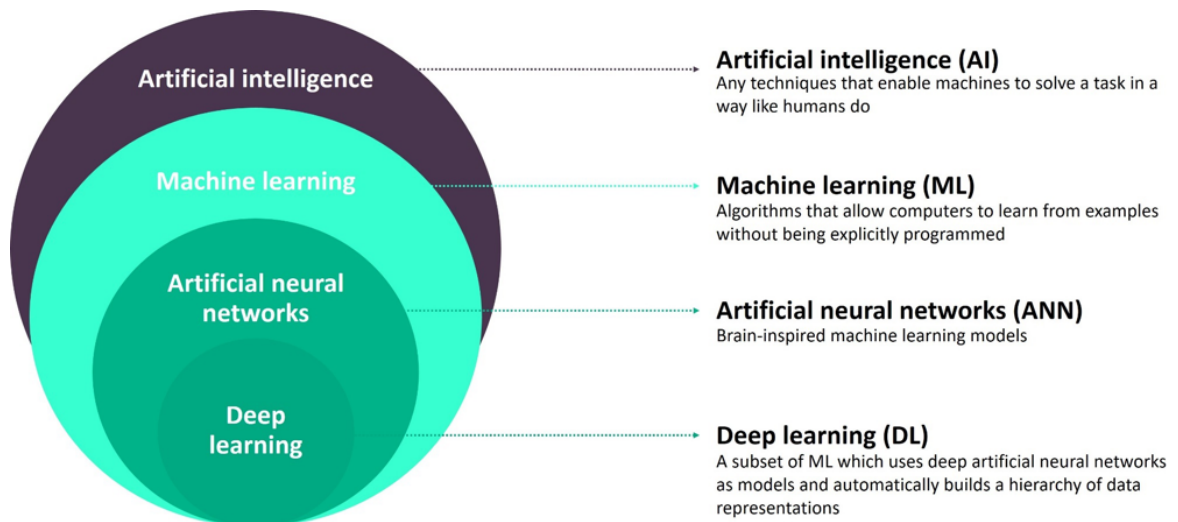
Ne všechny korelace jsou známkou základní kauzální souvislosti. Umělá inteligence obecně pracuje s velkým množstvím komplikovaných korelací, díky jejichž vzájemnému propojení je pravděpodobnost, že udělá chybu v nepřímé úměrnosti k jejich množství. Chyba jako taková ale nikdy není vyloučena absolutně.

### 2.2.6 Neuronová síť

Neuronová síť je posloupnost navzájem propojených vrstev neuronů/uzlů. Zjednodušeně všechny vstupy vstupují do všech vrstev neuronů, které vstupují do výstupní hodnoty dalšího neuronu až do výstupu. Síť hledá v prostoru chyby místo, které produkuje nejmenší chybu. Aby síť fungovala, musí mít pro vstupy vytrénovaný model na trénovacích datech. V praxi

to může fungovat tak, že model pozná na obrázku v první verzi vegetaci, v druhé vrstvě rozpozná že se jedná o strom, a v další vrstvě určí s určitou pravděpodobností druh stromu (prostor chyby se zmenšuje – od vegetace přes strom k druhu).

### 2.2.7 Typy Machine learning algoritmů



Obrázek 8 Hierarchie AI pojmů

Další členění Machine learning algoritmů je na Supervised Learning (učení pod dohledem), Unsupervised Learning (učení bez dohledu) a Reinforcement Learning (zpětnovazební učení)<sup>4</sup>.

### 2.2.8 Supervised Learning

Supervised Learning potřebuje operátora. Používá se při učení modelů na základě označených dat. Označenými daty jsou tomto kontextu například datasey obrázků, které obsahují popisky toho, co na nich je.

### 2.2.9 Unsupervised Learning

Unsupervised Learning využívá neoznačených dat k trénování modelů, nepotřebuje operátora. Tvoří vzorce a asociace a pomocí korelací vyvozuje kauzality (rozdíl mezi korelací a kauzalitou je že korelace může být náhodná, zatímco v kauzalitě je vždy příčina a následek).

<sup>4</sup>GOODFELLOW, Ian, BENGIO, Yoshua a COURVILLE, Aaron. Deep learning. Cambridge, Massachusetts: MIT Press, 2017

### 2.2.10 Reinforcement Learning

Reinforcement Learning provádí akce pomocí metody pokus-omyl na základě kterých se učí, jak dojít k cíli. Pomocí algoritmů učících se z předchozích výsledků rozhoduje o tom, jakou akci provede příště. Výhodou strojů při machine learningu je, že jsou schopny oproti člověku velice rychle zpracovávat data, vytrénovat model obvykle netrvá desetiletí, ale spíše nižší desítky hodin. U komplexních modelů je ale potřeba velmi silného HW.<sup>5</sup>

## 2.3 Možnosti využití ve VFX

Pokud je hezký obraz, či uvěřitelný vizuální efekt zobecněn na soubor slovně těžko definovatelných, avšak stále platných pravidel (světlost, kulatost, barva...), které je možné v určitém rámci a v určité například společenské kultuře definovat, je AI ideální nástroj na generování hezkých obrázků. V tomto rámci by zavedení AI procesů do tvorby zcela změnilo paradigma nejen vizuálních efektů, ale tvorby obecně.

### 2.3.1 Generování obrazu

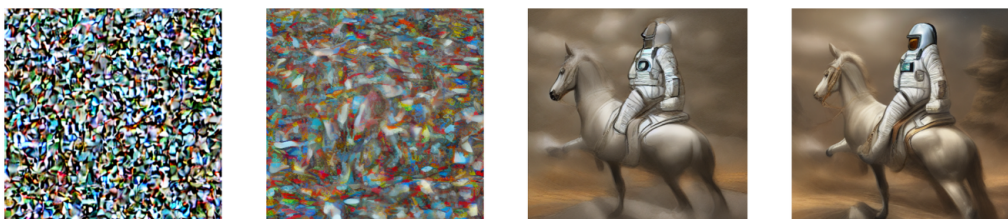
Před generování obrazu se modely v neuronové síti učí rozpoznávat a reprodukovat určité vzory a struktury, které se vyskytují v poskytnutých tréninkových datech. Například model může být trénován na obrázcích lidí, aby se naučil rozpoznávat rysy lidské tváře a reprodukovat je na nových obrázcích. Generování obrazu tedy nejdříve předcházelo jeho rozpoznávání. V roce 2015 se na internetu vyskytl online nástroj na obarvování fotografií, který pomocí neuronové sítě rozpoznal v pixelech objekty, osvětlení, a pixelům přiřadil barevnou hodnotu. Technologické demo položilo základ pro další vývoj. Tím byl upscaling obrazu, který využíval neuronovou síť pro zvyšování rozlišení a dopočítávání prázdných oblastí mezi pixely. AI gigapixel umí rozpoznat, je-li na fotografii v nízkém rozlišení například les, a tomuto shluku pixelů pak dodat patřičné detaily v podobě dalších doplňujících pixelů.

---

<sup>5</sup> Julianna Delua, SME, IBM Analytics, Data Science/Machine Learning. IBM.com. [Online] 12. Březen 2021. <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>.



Od hledání vzorců v datech po generování dat ze vzorců - celý proces rozpoznávání se podařilo přetvořit v generování pomocí šumu. Firma Open AI vydala v roce 2021 první verzi vytrénovaného softwaru Dalle. Tomuto softwaru předcházelo více nezávislých a méně komplexních prototypů na generování obrazových jednotlivostí, jako například krajin, nebo tváří. Rozdíl je hlavně v objemu tréninkových dat. Obrázky progresivně vznikají za pomoci náhodně vygenerovaného šumu. V tomto šumu se v iterujícím se procesu AI snaží najít vzory navázané na prompt (tento pojem vysvětlím později). Kvalita výsledku se odvíjí od množství tréninkových dat, ale i od počtu iterací, i když od určité úrovně začne rozdíl ve výsledku dramaticky klesat. V každé další iteraci, která načítá šum z té předchozí se šum zpřesní, a chyba v zobrazení se postupně snižuje. Protože je seed původního šumu vygenerovaný náhodně, vznikne při každém novém pokusu se stejným promptem jiný výsledek.



Obrázek 9 Fáze generování obrázku ze šumu. (2021, OpenAI)

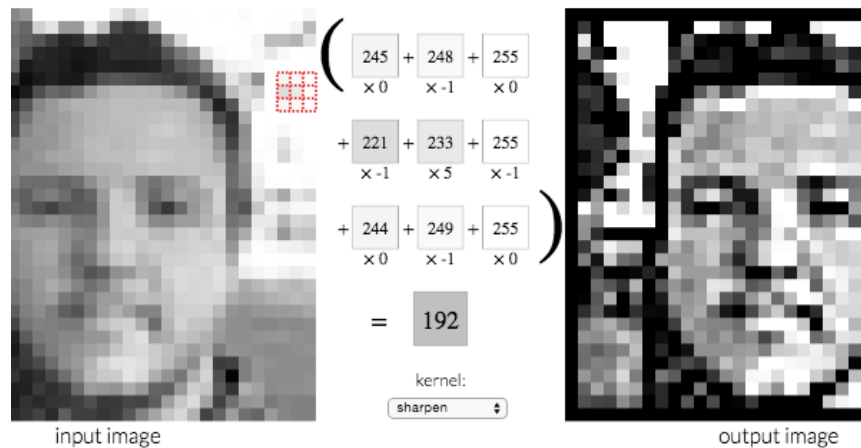
### 2.3.2 Konvoluční filtr

Všechny tyto procesy (rozpoznávání i generování) se obvykle provádí pomocí konvolučního filtru.

Konvoluční filtr je matematická operace, která je aplikována na vstupní obrázek, aby z něj extrahovala vlastnosti. V kontextu deep learningu se konvoluční filtry často používají ke zpracování dat s mřížkovou strukturou, kde detekují specifické vzory, jako jsou hrany, rohy nebo jiné tvary. Každý konvoluční filtr má sadu vah, které určují, jakým způsobem přemění vstupní data. Tyto váhy se neuronová síť učí během trénovacího procesu, aby maximalizovala svou schopnost extrahovat nejužitečnější vlastnosti z vstupních dat.

V kontextu VFX se může jednat například právě o obrázky.

Konvoluční filtry pracují tak, že filtr posouvají po vstupním obrázku, po jednom pixelu, a počítají skalární součin mezi hodnotami filtru a vstupním obrázkem na každé pozici. Toto produkuje dvourozměrnou aktivační mapu, která se poté předá přes ne-lineární aktivační funkci, jako je například funkce ReLU, aby se do sítě zavedla nelineárnost a zlepšila její schopnost učit se složité vzory.



Obrázek 10 Příklad funkce sharpen založené na konvolučním filtru

V zásadě jsou konvoluční filtry silným nástrojem v deep learningu a jsou široce používány v různých aplikacích, včetně klasifikace obrázků, detekce objektů a segmentace obrázků<sup>6</sup>.

### 2.3.3 Metoda Stabilní Difuze

AI modely používané softwaru Stable diffusion (Stability AI), Dalle2 (OpenAI) i Midjourney (Midjourney) fungují na společném principu stabilní difuze. Difuze je nahodilý rozptyl původní struktury do neuspořádaného šumu. Touto metodou se olabelovaná vstupní tréninková data (například z datasetu LIAON5B který obsahuje pět bilionů obrázků párovaných s textem) rozptylují do nahodilého šumu. AI se v každé další iteraci najít v obraze původní struktury, které přiřazuje k vektorům vytyčeným z labelů (strom, pes, temný atd.). Pomocí tohoto rozpoznávání vzorců vznikají v neuronové síti asociace. Podstata

<sup>6</sup>GATYS, Leon A., ECKER, Alexander S. a BETHGE, Matthias. Image style transfer using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[online].2016,s.2414-2423[cit.2023-04-09].Dostupnéz: [https://openaccess.thecvf.com/content\\_cvpr\\_2016/html/Gatys\\_Image\\_Style\\_Transfer\\_CVPR\\_2016\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2016/html/Gatys_Image_Style_Transfer_CVPR_2016_paper.html)

obrázkových generátorů spočívá v obrácení tohoto procesu. Ten funguje podobně jako starší AI denoisery nebo upscalery, ale zároveň hledá v šumu struktury vyplývající z obsahu textového promptu. Algoritmus totiž zná způsob, jakým se obrázek měnil na šum, a tak dokáže ze zdánlivě neuspořádaného šumu odvodit obrázek. Původní šum tak přizpůsobuje předchozím rozptýleným datům z tréninkového procesu. Výstup je pokaždé jiný, protože v šumu je implementován náhodný seed (nahodilá variace rozmístění zrn v šumu), který se v defaultu při každém opakování mění.

### 2.3.4 Prompt a jeho optimalizace

Prompt je textové rozhraní, skrz které může uživatel komunikovat s AI platformou. Na slova v textu se pomocí vektorů v latentním prostoru (matematický mnohadimenzionální prostor) napojí parametry (kulatost, lesklost, červenost, chlupatost atd.) Jedná se o vzorec korelací, které zahrnují vlastnosti, pro které nemusí ani existovat v lidském jazyce slova, ale pomocí velkého množství trénovacích dat se v neuronové síti vytvoří abstraktní analogie a asociace. Prompt na tyto data odkazuje a zpětně z nich vzniká výstup. Prompt může vypadat například takto: „*Photography of red hood dressed young girl sitting on big wolf in deep forest, fairytale, night, film lighting, hyperrealistic, artistic, cinematic*“. Každé slovo odkazuje k různým parametrům v latentním prostoru, které tvoří funkční asociaci a pomocí řetězení slov se tato asociační rovnice stává konkrétnější<sup>7</sup>.



Obrázek 11 červená karkulka vygenerovaná pomocí Midjourney (2023, Midjourney)

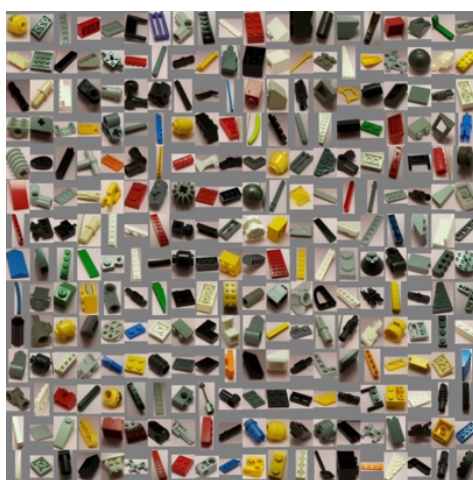
<sup>7</sup>LEWIS, Nick. How to Write an Awesome Stable Diffusion Prompt. How-To Geek [online]. 2021 [cit. 2022-04-23]. Dostupné z: <https://www.howtogeek.com/833169/how-to-write-an-awesome-stable-diffusion-prompt/>

Prompt odkazuje na parametry. V současné době se mi jako nejučinnější varianta ukázalo organizovat prompt způsobem uváděným na stránkách midjourney. Začít tím co chceme vygenerovat, přidat popisná upřesňující slova a prompt zakončit popisem stylizace (využit rozšířený pojem nebo jméno umělce). Midjourney přidává možnost dalších upřesňujících příkazů, jako například `-v (1-5)` pro definování verze kterou má využít, nebo například `-ar 16:9` pro poměr stran 16:9.

Možností je, nechat si i prompt vygenerovat za pomoci umělé inteligence ChatGPT. Ta bohužel ale využívá neaktuální data z internetu (2021). Od té doby se celé odvětví značně posunulo. Při využití placené verze GPT-4 je ale možné vkládat vlastní informace, a tak například nakopírovat aktuální dokumentaci cílového generátoru a nechat GPT-4 pracovat na základě aktuálních paradigmat.

### 2.3.5 Datasetsy – tréninková data

Data se do AI softwarů sbírají z internetu, například v případě obrazových generativních AI se nejvíce využívají indexované labelované obrázky. Jedná se o obrázky, ke kterým je připojený nějaký textový popis. Nejčastějším zdrojem podobných obrázků jsou fotobanky, novinové servery, umělecká portfolia jako artstation apod. Datasetsy na vytrénování vlastního specifického modelu je také možné nakoupit přes služby třetích stran jako například LegoBricks což je dataset více než dvanácti tisíc fotografií legových kostek<sup>8</sup>.



Obrázek 12 Dataset LegoBricks  
(2020, LegoBricks)

---

<sup>8</sup>HAZELZET,Joost.LegoBrickImages[online].Kaggle,2018[cit.2023-04-09].Dostupné:  
<https://www.kaggle.com/datasets/joosthazelzet/lego-brick-images>

### 2.3.6 AI jako nástroj pro optimalizaci HW výkonu

Zde se jedná spíše o testovací modely, ne o použitelné verze softwarů, ale potenciál je velký. Například v oblasti simulací, ke kterým jsou dnes potřeba výkonné počítače, a i tak je práce s nimi velmi časově náročná.

V současné době se věrné přírodní a fyzikální fenomény (kolize objektů, částicové systémy, simulace kapalin, volumetrické simulace, voxelové programovatelné growth systémy atd.) v pohyblivém cgi vytváří pomocí simulací. Tento postup je fyzikálně přesný, avšak náročný na čas potažmo hw výkon počítače.

Limitace HW výkonu se projevují především v pomalém výpočtu simulace a v omezené schopnosti zpracovat velká množství dat. To může vést k chybám v simulaci nebo při snaze o optimalizaci k nedostatečně detailnímu vykreslení scény, což může zase negativně ovlivnit kvalitu a realističnost výsledku.

Celý proces je exponenciálně pomalejší u velkých scén, kde je potřeba nasimulovat více různých na sebe působících prvků. Většinou se jedná o rozsáhlé simulací destrukcí měst, interakce s rozbouřeným mořem apod. Filmový divák přitom nevnímá přesnost z hlediska výpočtu, ale spíše z obrazové a kinetické věrohodnosti. Tato náročnost vychází z faktu, že v rámci konvenčních metod se vše přepočítává matematicky pomocí hrubého výkonu o kterém jsem již psal v jedné z předchozích kapitol.

V souvislosti s možnostmi AI existuje na tomto poli velký potenciál pro vývojový posun. Pomocí AI by bylo možné výpočet těchto scén podstatně zjednodušit. Například vytrénovat model na datech vycházejících ze simulací ohně, včetně jeho dílčích vlastností jako jakým způsobem interaguje s překážkou či okolním větrem. Nebylo by potom nutné oheň přesně simulovat každý snímek, ale AI algoritmus by přizpůsobil oheň prostředí, a jeho interakci a chování by odhadl za pomoci predikcí. Pro uživatele by se toho příliš nezměnilo, ale umožnilo by to další výkonnostní skoky šetřící čas. V současné době totiž růst efektivity pracovních stanic pro CGI zpomaluje a čas, který náročné simulace zaberou ve svém důsledku přímo ovlivňuje kvalitu výstupu.

Tématu optimalizace HW výkonu se ještě dotknu v kapitolách o interpolaci snímků, upscalingu, denoiserů, a vývoji alternativních simulačních nástrojů.

## 2.4 Progresivní kroky vs nahodilá iterace (intuitivní skoky)

Nejpodstatnější současný rozdíl v tvorbě jednotlivých prvků pomocí konvenčních metod a při využití umělé inteligence ze kterého plyne i způsob práce bych rozdělil dvěma termíny.

Progresivní kroky a nahodilá iterace.

Zde vycházím z přednášky Valentine Kozina na VFX Festivalu, která proběhla v době covidu online formou. Valentine Kozin je Lead Technical Artist Rare studia a ve své praxi se zabývá ve své podstatě přesně tím, o čem pojednávám v této práci. A sice využití AI v rámci CGI<sup>9</sup>.

Umělec tvořící nějaké dílo (Animovanou scénu/vizuální efekt, matte-paint apod.) většinou progresivně postupuje směrem k cíli. V jednotlivých krocích se stále více přibližuje výsledku, až přijde moment, kdy je dílo „hotovo“.

Abych se dostal z fáze A do fáze E, musí člověk nejdříve projít fáze B, C a D. Při cestě k cíli se většinou vyskytují různorodé překážky a problémy, které musí člověk vyřešit či obejít, za pomoci svých znalostí, zkušeností a nástrojů. Je to proces progresivních kroků.

Umělá inteligence je ale nástrojová sada jiné kategorie, jsou to nástroje, které k nalezení řešení nevyžadují znalost dílčích oblastí. Stačí znát výchozí a finální body A a E a AI obstará zbytek. Oproti původnímu konceptu pro člověka postupně odpadá nutnost znalosti dílčích specializovaných metod.

Tyto AI nástroje pracují pro průměrného generalistu jako black box, který progresivní kroky směrem k výsledku přeskakuje, a nabízí přímý výstup. Je zde jen vstup, a výstup mezi kterými leží latentní prostor se svými těžko definovatelnými procesy.

To je zároveň i nevýhoda tohoto nástroje. Generalista, využívající umělou inteligenci nemá možnost do latentního prostoru nijak zasahovat, a jeho moc je tak omezena na předefinování vstupů (promptů) tak, aby se výstup co nejvíce podobal tomu, co je jeho reálný cíl.

K této finální fázi, kterou jsem po vzoru Kozina zobecnil písmenem E většinou nevede hned první pokus – iterace, a tak musí přijít opakování, kdy generalista redefinuje prompt generující výstup. Vzniká tak množství nahodilých iterací. Tento fakt ilustruji dále v praktické části za pomoci experimentu. Samozřejmě v praxi je možné tyto postupy

---

<sup>9</sup>Kozin, Valentine. Záznam online konference. [Online] 23. prosinec 2020.

[https://www.youtube.com/watch?v=miLIwQ7yPkA&ab\\_channel=EscapeStudios](https://www.youtube.com/watch?v=miLIwQ7yPkA&ab_channel=EscapeStudios).



kombinovat s konvenčními metodami, ale je dobré mít tuto modularitu omezující funkcionalitu AI na paměti pro lepší představu o tom, kde se hodí tyto nástroje využít, a kde je lepší prozatím zůstat u konvenčních metod, popřípadě zvolit hybridní přístup.

## 2.5 Čas

Čas je důležitý parametr při volbě nástrojů.

Komplexnost výstupu přímo ovlivňuje jak čas nutný na jeho realizaci, tak odborné pokrytí, které musí generalista zvládnout, popřípadě si na některé části procesu najmout specialistu.

V kontextu komplexního VFX projektu zahrnujícího mnoho odborností, popřípadě dalších lidských zdrojů je pravděpodobnost vzniku chyby, jejíž oprava zvýší finální časovou náročnost větší. A čím je projekt složitější, tím více chyb vzniká a je nutno více opakování kroků na jejich odhalení a opravení. Překážky, které je nutné v průběhu práce překonávat žádají rozdílná řešení. Přehled a out of the box myšlení, kterým by měl generalista oplývat urychluje překonávání těchto překážek.

Hlavní limitací člověka při plnění jednotlivých kroků v tvorbě je v konečném důsledku čas. Ten je možný částečně zkrátit pomocí jeho schopností a zkušeností. Pokud série kroků v činnosti trvá například hodinu, je možnost zhodnotit výstup a zpracovat změny jednou za hodinu. Pokud je časová dotace na finální výstup deset hodin, bude mít tvůrce pouze deset pokusů na to zhodnotit svou práci. Proto je velký rozdíl, pokud člověk najde správnou cestu na druhý pokus, nebo na desátý pokus.

Pokusem je nyní myšlena situace, kdy je nutno opakovat nějaký proces, například z důvodu odhalení chyby.



Obrázek 13 Srovnání času postupů retuší fotografií – vlevo původní fotografie, uprostřed retuš v Photoshopu 2 hodiny, vpravo Dalle 2 retuš 10 minut. (2023, Veronika Bočková)

V tomto kontextu může být AI velice nápomocna, protože zpracovat úpravy jí trvá v podstatě zanedbatelný čas.

Tvůrčí svoboda, kterou mi nástroje poskytují je pro mě osobně dána tím, jak rychle mi trvá dostat se z bodu A do bodu E a jak moc bod E nakonec odpovídá mé představě toho, jak by měl vypadat<sup>10</sup>.

### 2.5.1 Repetitivní činnosti v tvorbě

V rámci procesu tvorby se vyskytuje množství repetitivních procesů, které nemají žádný reálný přínos v podobě zvýšené kontroly nad výsledkem. V tomto ohledu vidím v AI nástrojích, kterým se budu věnovat v další části práce velký potenciál. Potenciál pro vychýlení parametru chybovosti a snížení vyplývajícího času ve prospěch tvůrce. Díky těmto nástrojům bude postupně odpadat nutnost porozumění dílčím pravidlům a díky tomu bude docházet k opětovnému ztenčování technologicko-znalostní bariéry mezi autorem a dílem.

## 2.6 Generalista/specialista

VFX generalista je osoba, která má široké znalosti a dovednosti v oblasti vizuálních efektů a může se podílet na různých fázích vývoje VFX projektu. Tento typ profesionála má schopnost pracovat na široké škále úkolů, včetně modelování, texturování, animace, renderingu a compositingu. Generalisté jsou velmi užiteční v situacích, kdy je třeba rychle reagovat na změny nebo když je třeba pracovat na projektech s omezeným rozpočtem nebo časovým harmonogramem.

VFX specialisté jsou naopak odborníci v určité oblasti VFX, například v modelování, texturování nebo compositingu, zaměřují se na konkrétní oblast a mohou být díky tomu velmi zkušení ve svém oboru. To jim teoreticky umožňuje pracovat na složitějších a detailnějších projektech v týmu složeném z dalších specialistů.

Mezi výhody VFX generalistů patří jejich široká znalost a dovednosti, které jim umožňují pracovat na různých projektech a rychle se přizpůsobit změnám. Díky svým širokým znalostem jsou generalisti schopni odbavovat projekty v menších týmech nebo i jako jednotlivci. Navíc díky propojování jednotlivých oborů může být generalista schopen hledat rychlejší cesty, které ho dovedou ke stejným výsledkům.

---

<sup>10</sup>NEURALCONCEPT.TheIterativeDesignProcess: A step-by-step guide & the role of deep learning [online]. [cit. 2023-04-09]. Dostupné z: <https://www.neuralconcept.com/post/the-iterative-design-process-a-step-by-step-guide-the-role-of-deep-learning>

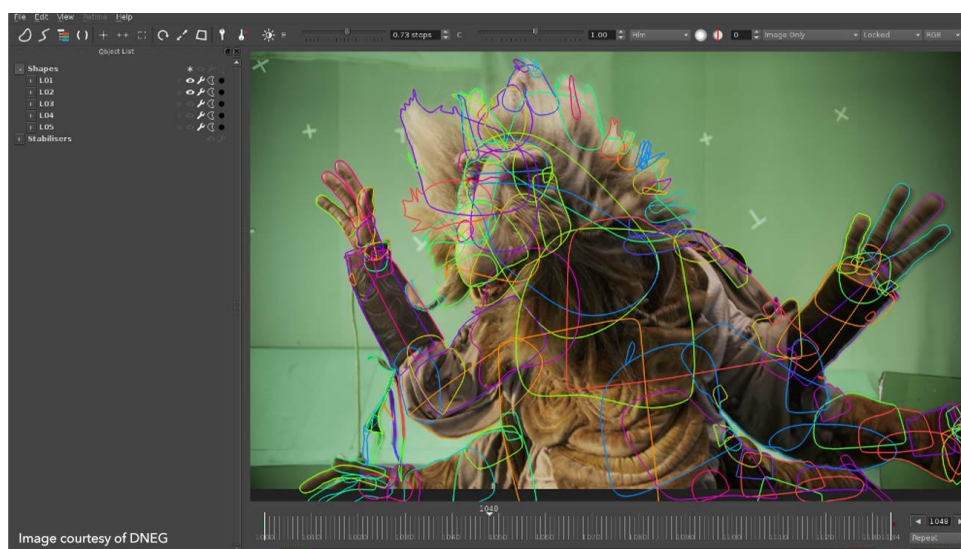
David Epstein ve své knize *Range: Why Generalists Triumph in a Specialized World* popisuje několik důvodů, proč se v konečném důsledku v pravidlech dnešního světa vyplatí být raději generalistou než specialistou. V prostředí, které je komplexní, a zároveň vysoce proměnlivé dokáží generalisté díky silně vybudovanému obecnému přehledu lépe reagovat, propojovat oblasti a hledat alternativní řešení na místech, kde je specialisté pro své úzce specifikované znalosti nevidí. Generalisté si svým způsobem práce budují schopnost aplikovat znalosti a schopnosti na nové a odlišné situace a kontexty. Navíc pokud jde o finální dílo, je generalista schopen zastat více/všechny fáze procesu, a hotové dílo má v naší společnosti vždy větší hodnotu než jedna určitá jeho část vyžadující další dopracování<sup>11</sup>.

Navíc v situaci, kdy se rozvíjejí AI nástroje se zdá, že některé specializace postupně ztrácí svůj účel. Těmto „ohroženým“ specializacím se budu věnovat v další kapitole.

## 2.7 Aktuální alternativní AI metody ve VFX industry

### 2.7.1 Rotoscoping

Rotoscoping je postprodukční technika, při které se ručně vytvářejí masky kolem objektů nebo osob ve videu, aby se mohly později oddělovat od pozadí. Díky této technice je možné objekty nebo postavy zasadit do jiného prostoru. Roto-artist je specializovaná pozice ve velkých trikových studiích. Jeho hlavním zaměřením je optimalizovat celý proces a pomocí trackování a různého typu maskování docílit přesného klíče. Na roto se obvykle využívají méně kvalifikovaní lidé z chudších zemí. Několik let zpátky byla za Mekku roto-artistů



Obrázek 14 Rotoscopování pomocí Mocha Pro (2019, Mocha pro)

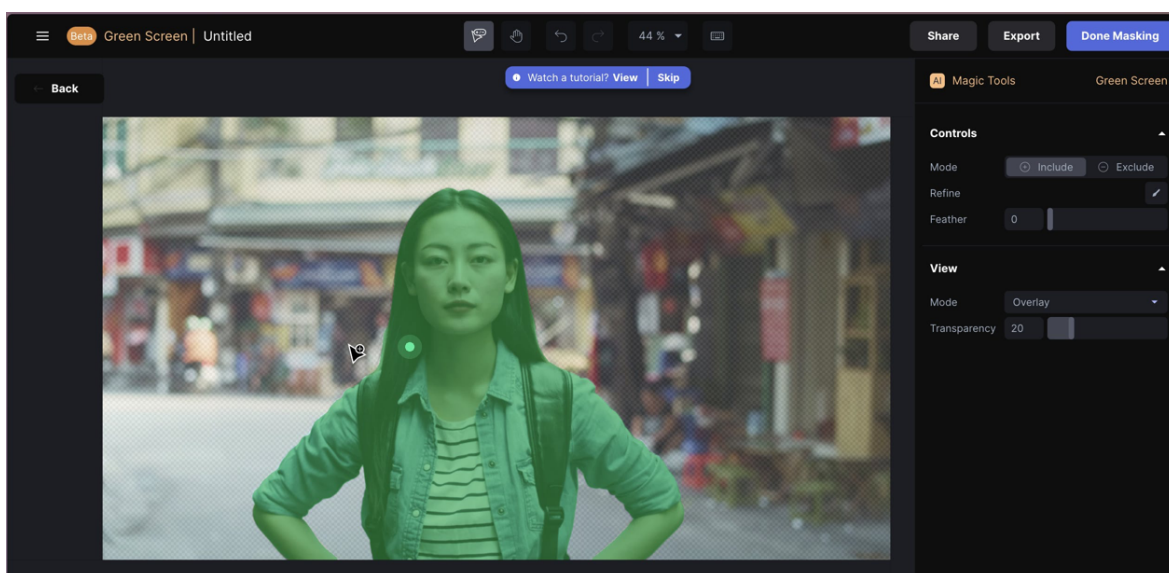
<sup>11</sup> EPSTEIN, David. *Range: Why Generalists Triumph in a Specialized World*. New York: Riverhead Books, 2019.

například Indie. AI může být použita k automatizaci této práce a ušetřit velké množství času a prostředků.

Existuje několik způsobů, jak může AI pomoci s rotoscopováním:

V současné době plně fungujícím programem na rotoscopování lidí je například program Runway<sup>12</sup> běžící skrze webovou aplikaci. Runway má mimo jiné model vytrénovaný na datasetu čítajícím mnoho záběrů lidí. Tento dataset obsahuje velké množství záběrů různě se pohybujících lidí vždy ve dvou verzích. V jedné verzi je původní záběr, a ve druhé je člověk z tohoto záběru vyrotoscopovaný. Pomocí strojového učení z těchto záběrů může program predikovat rotoscoping na nových záběrech. Runway pracuje na vzdálených GPU serverech, takže uživatel nepotřebuje využívat pro rotoscoping výkonný hardware. V době psaní této práce je jasné, že zkušený člověk zvládne doručit lepší výsledek, ale z hlediska časové náročnosti zcela jistě vede AI. Specialisté mají totiž oproti současným AI v oblasti rotoscopie schopnost se kreativně rozhodnout, kterou techniku, v který moment zvolí. Nezřídka se ale vyskytují situace, kdy přesnost výstupu není prioritou a například právě generalisté rádi využijí AI nástroje.

Podobný princip trénování modelu na vytváření masek a hledání siluet je možné rozšířit na větší množství předmětů, než jsou jen osoby, ale v současné době není toto řešení k dispozici.



Obrázek 15 Rotoscopování pomocí Runway (2022, Runway)

<sup>12</sup>RunwayML.(2021).RunwayML:Machine Learning for Creators. [online] Dostupné z: <https://runwayml.com/>

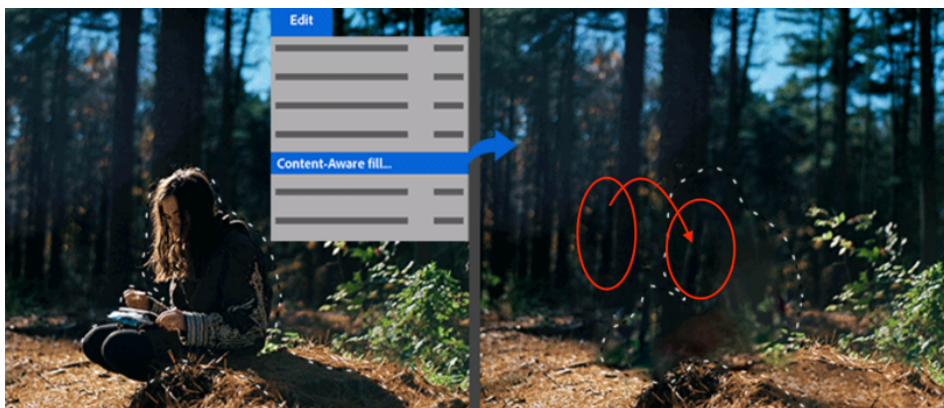
Je pouhou otázkou času, než se toto řešení stane součástí běžných softwarů jako je Nuke, či After Effects. Potenciál pro automatickou úpravu masky například mezi framy, nebo automatické rozpoznávání a upravování specialistou špatně vytvořených masek objektů není technicky příliš vzdáleno.

### 2.7.2 Retuše

V rámci postprodukce se často setkáváme s požadavkem něco nebo někoho odstranit ze záběru. Může se jednat o logo na oblečení, o zvukařské tágo, které se nedopatřením dostalo do záběru, nebo náhodného člověka v pozadí. Retuše podobně jako rotoscoping nejsou většinou prací, vyžadující velkou uměleckou kompetenci.

První možností je, celou retuši provést ručně. To vede ke kontrolovanému výstupu, kdy za pomoci kombinace zkušenosti a množství času získáme funkční výsledek. Snahou softwarových vývojářů je však i tento proces z co největší části zautomatizovat pomocí image processingu.

V kompozičním programu After Effects podobně jako v Photoshop se vyskytuje nástroj content aware fill, který slouží právě k doplnění předtím vystřižené části obrazu. Tento nástroj je používán již několik let, a i když nepracuje pomocí neurovnových sítí, dokáže z jednodušného pozadí obstojně odstranit vybraný objekt. Princip jeho fungování je založen na replikování shluků podobných pixelů v okolí masky. Tyto oblasti pak nástroj kopíruje a různě prolíná do prázdného prostoru, který tak zaplní. Ve scénách, kde se vyskytuje komplexní pozadí (oproti například jednoduché modré obloze) však vznikají viditelné šmouhy a nelogičnosti, kdy nám například část fasády domu překryje částí keře, nebo je zde viditelné duplikování okolního prostředí. *(připojit ukázky).*



Obrázek 16 Ukázka nedokonalosti výstupu Content Aware fill



Image processing se dá zpracovat i pomocí strojového učení. Ideální alternativou je v současné chvíli online aplikace Dalle2, který uživatelé dovoluje vyplnit vymaskovanou část obrazu za pomoci promptu. Na pozadí tohoto procesu už není pouhé kopírování částí



Obrázek 17 původní ulice bez úprav (2023, Veronika Bočková)



Obrázek 18 retuš pomocí AI nástroje Dalle2 (2023, Veronika Bočková)

obrázku z okolí masky, ale generování objektů na základě kombinace psaného promptu a objektů které Dalle2 rozpozná v obrázku. Vznikne tak fotorealistický výstup bez nutnosti časové investice nebo široce specializovaných znalostí a zkušeností dalších retušovacích nástrojů. Samozřejmě v případech, kde je v záběru nějaký pohyb je nutné kombinovat techniku s trackováním, ale AI nástroje zde mohou v mnoha případech pomoci vytvořit kvalitní statický clean plate.

### 2.7.3 Concept art, matte painting

Nástroj Dalle2, podobně jako Stable diffusion nebo Midjourney se dají využít i k ostatním specializovaným disciplínám. Další z nich je tvorba concept artů. Zatímco i zkušenému umělci, trvá naskicovat návrh scény hodiny až dny, strojové učení to zvládne za jednotky sekund. I když je v současné době se gap mezi umělou inteligencí a člověkem v kvalitě výstupu stále poměrně rozpoznatelný. Generátory obrázků umí významně urychlit celý proces. Podobně jako při photobashingu je možné skládat z různých zdrojů například v programu Photoshop. Oproti využití photosbashingu si autor využívající AI generátory může zdrojové materiály tvořit namíru pomocí promptů. V nových verzích programů jako Dalle2 nebo Stable Diffusion je možné přegenerovat a upravovat pomocí dodatečných promptů vybrané části obrázku, takže pro generalistu je process ještě jednodušší.

Tento princip je možné využít i pro matte paingu nebo extenzi setu. Je totiž možné nahrát vybraný frame a přegenerovat pomocí AI softwaru jeho vybranou část tak, jak je možné vidět v příloženém obrázku.



Obrázek 19 matte paint generovaný pomocí Dalle2

Právě zrovna matte painting a concept art je vysoce specializovaná disciplína vyžadující velké množství znalostí a zkušeností, pokud by AI softwaru zvládly pokrýt kompletní část výrobního procesu, byla by to ještě větší změna pro celou industry.

Problém je v současné chvíli s uživatelským rozhraním, které není v AI programech příliš komplexní, a hlavní způsob komunikace uživatele je právě pomocí již mnohokrát zmiňovaného textového promptu.

Například Nvidia se pokouší tuto bariéru omezující kontrolu překlenout pomocí zatím stále demo verze svého softwaru Canva, který využívá barevné štětce jako přídatný ovládací prvek v podobě vizuálního promptu. Každý barva reprezentuje určitou věc/element, takže člověku stačí namalovat zelenou plochu, a Canva místo podle legendy doplní například luční krajinu. Canva v současnosti zvládá generovat pouze krajiny a nemá možnost například nastavit perspektivní údaje, takže výsledky jsou stále poměrně nahodilé. Je to ale další krok

k novému způsobu práce, který bude nejspíše mnohem intuitivnější. V době psaní této práce ale Nvidia zveřejnila náhled dema, které je již schopné generovat HDRI mapy. Tzn 360\*180 kompletní panoramatické textury, což by problém s perspektivou částečně vyřešilo.<sup>13</sup>

#### 2.7.4 Charakterová Animace

Animace je celý jeden obor, kterým se zabývá skupina specialistů, těmto profesionálům se říká animátoři. V kontextu VFX se pod pojmem animace nejčastěji jedná o animaci charakterů.

Profesionál zabývající se rozpohybováním digitálních živých tvorů musí mít nasledovanné chování a kinetiku pohybu těl různých zvířat a lidí v různých situacích a podmínkách včetně fyziky která ovlivňuje způsob jakým se hýbou tvorové různých velikostí. Animátor musí být poté schopen tyto pohyblivé parametry ve své mysli nějak kvantifikovat a promítnout při práci na animovaném charakteru. Charakter může být kůň, pes, drak, nebo člověk, přičemž každý má své specifické pohybové vzorce. U fantazijních charakterů většinou animátor vychází z pohybových vzorců podobných tvorů, kteří existují v realitě.

Pozice animátora je v trikovém průmyslu vysoce ceněna a nenahraditelná, protože vyžaduje schopnosti, a zkušenosti, jež získání vyžaduje poměrně velké množství času.<sup>14</sup>

Existuje nástroj Cascadeur (v době psaní této práce je stále v beta verzi) který obzvláště proces animace značně zjednodušuje. Využívá interní metody AI keyframe asistent k predikování póz animovaného charakteru. Tvůrce tak nemusí editovat při každém klíčovém snímku všechny slidery riggu, ale Cascadeur sám predikuje způsob, jakým by například prudký pohyb paže ovlivnil celkový postoj charakteru. Dále Cascadeur dokáže do animací automaticky přidávat sekundární pohyby zohledňující zákony fyziky a kinetiky<sup>15</sup>.

#### 2.7.5 Motion Capture

V oblasti VFX je funkce neuronových sítí již nyní hojně využívána při méně produkčně nákladných natáčení vyžadujících trikovou postprodukcí v oblasti motion capture. Standartní motion capture systémy se nachází ve specificky vybavených studiích, kde herce omarkerované trackovacími body snímá systém kamer. Pozice jednotlivých bodů se pak

---

<sup>13</sup> Yogesh, Balaji. NvidiaCanva-ImageToImage. [Online] <https://research.nvidia.com/labs/dir/eDiff-I/>.

<sup>14</sup> THALMANN, Daniel a Nadia Magnenat THALMANN. Computer Animation: Theory and Practice. Springer, 1990.

<sup>15</sup>

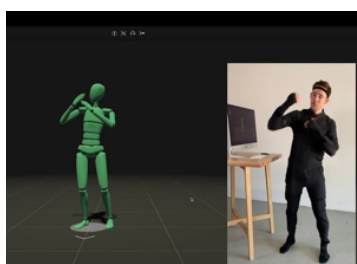
[https://80.lv/.%20Cascadeur%20Research:%20Physics%20In%20Animation.%20\[Online\]%202020.%20http://80.lv/articles/cascadeur-research-physics-in-animation](https://80.lv/.%20Cascadeur%20Research:%20Physics%20In%20Animation.%20[Online]%202020.%20http://80.lv/articles/cascadeur-research-physics-in-animation).



přepočítává v softwaru, který je na to určený (Například Motionbuilder)<sup>16</sup>. Tento postup je velice přesný, ale právě z produkčního hlediska je jeho zajištění poměrně náročné. Indie developeri přišli více než před deseti lety s alternativou, využívající depht senzory ovladače Windows Kinect. Pomocí těchto senzorů bylo možné měřením vzdáleností oddělit postavu od pozadí, a z této siluety pak pomocí algoritmů převádět pohyb na rigg v 3D prostoru. Kinect byl levné řešení, které jak jsem sám zjistil, mělo zásadní problém s trackováním chodidel, které nedokázal přesně oddělit od země. Podobně nepřesný byl i v drobných animacích prstů, které vnímal společně se zbytkem ruky jako jeden objekt. Existuje několik dalších konvenčních alternativ. V současné době se často používají mocap suity, které v sobě mají senzory, měřící svou vzájemnou polohu, a tak je možné používat tento systém i v terénu mimo studio. V posledních letech se začaly objevovat řešení, které více či méně úspěšně extrahovaly siluetu z RGB obrazu. K motion capture tak bylo možné velmi omezeně využívat běžné kamery třeba přímo na setu bez nutnosti externích systémů. Toto řešení podávalo ještě o něco horší výsledky než Kinect. Systémy zaznamenávající pohyby člověka z RGB obrazu (videa) existují sice několik let, ale nyní se je vývojářům podařilo propojit s modely umělé inteligence. S jejich pomocí je možné dosahovat uspokojivě přesných výsledků za použití běžné kamery a celý proces je velice intuitivní. Nejznámější v oboru je v současné době firma Rokoko<sup>17</sup>, které začala mimo své motion capture suity, v oblasti, kterých je světovou špičkou nabízet i vlastní software na trackování pohybu postavy z videa kde jí však rychle dotahuje konkurence jako například Radical motion. Dále existuje například software Deep-motion<sup>18</sup>, který mimo klasický motion capture zvládá z RGB videa



Obrázek 20 Původní technologie motion capture



Obrázek 21 Rokoko suit



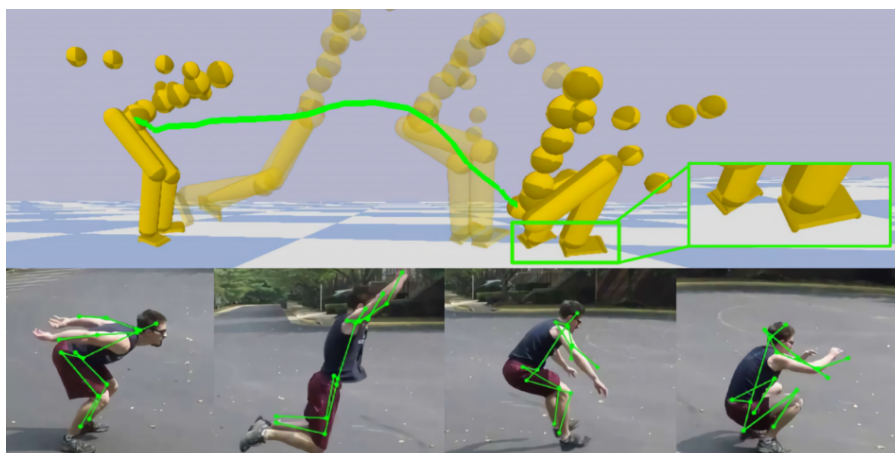
Obrázek 22 Windows Kinect

<sup>16</sup> MENACHE, Alberto. Understanding Motion Capture for Computer Animation and Video Games. San Francisco: Morgan Kaufmann Publishers Inc., 2010. ISBN 978-0123814968.

<sup>17</sup> ROKOKO. Rokoko - Motion capture for all [online]. [cit. 2023-04-07]. Dostupné z: <https://www.rokoko.com/>

<sup>18</sup> DEEPMOTION. DeepMotion - AI-powered motion intelligence [online]. [cit. 2023-04-07]. Dostupné z: <https://www.deepmotion.com/>

trackovat i facial expressions. Nebo move AI<sup>19</sup>, který dokáže kombinovat více kamer a získat tak přesnější výsledek.



Obrázek 23 Vytrénování AI pohybového modelu z RGB obrazového záznamu.

### 2.7.6 Texturování, de-lighting a tvorba materiálů

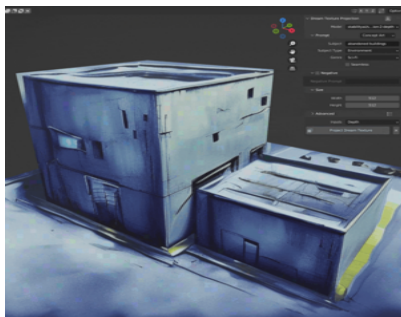
Texturování je podobně jako ostatní expertízy uvedené v této práci oblast, kterou ve velkých studiích zastupují specializovaní umělci. Tito lidé musí zvládat věrohodně napodobit vizuální vlastnosti povrchu objektu tak, aby působil věrohodně. Pracuje se s několika typy obrazových obrázků, které společně tvoří vlastnosti materiálu. Mimo klasickou barevnou texturu existují další typy textur-map, které převádí matematické hodnoty na obrazová data. Existují černobílé mapy, které určují input hodnoty od nuly do jedné. Ale také existují například normálové mapy, které zvládají pomocí obrázku zobrazit tříosá data. Pixely jsou podle svého výsledného náklonu tónované do barev červené, zelené a modré, kdy každá představuje jednu ze tří prostorových os X/Y/Z (*proto jsou normálové mapy zbarveny do modra, protože osa Z je z pohledu textury směrem nahoru*). Specialista zpracovávající materiály musí z modelu vytvořit UV mapu – unwrapovat (rozbalit jednotlivé polygony do plochy s co nejmenší deformací) na kterou poté pomocí specializovaných programů (dnes se nejčastěji využívá Substance Painter/Mari) kreslí jednotlivé textury tak, aby vznikla data, ze kterých se pak skládají v 3d editorech funkční materiály. Celý proces má množství kroků, které (pokud mají být zachována všeobecně uznávaná pravidla) jsou nutné správně provést. Existují však zkratky, které dovolují tyto pravidla v některých případech ohnout, a přesto

---

<sup>19</sup> MOVEAI. MoveAI - Multi-camera motion capture [online]. [cit. 2023-04-07]. Dostupné z: <https://www.moveai.com/>

docílit funkčního materiálu. Existuje například možnost, pracovat pouze s fotografií, a jednotlivé typy texturových map si poté vygenerovat s tím, že dnes je právě generování dalších složek materiálu posíleno o zpřesňující funkce neuronových sítí. Možnost generovat si jednotlivé textury z obrázku například pomocí Photoshopu, nebo jiných programů je v rámci světa 3D technologií totiž poměrně stará. Původní proces extrahoval z obrázku normalizovaná luminance data, která podle kombinace hodnot tmavosti jednotlivých pixelů, manipulace kontrastu a rozostření (opět známý konvulační filtr) převedl na specular, roughness, a displacement mapu, ze které poté spočítal normal mapu. Photoshop nedokázal rozpoznat jakou texturu zpracovává, ale pouze strojově podle jednoho filtru převáděl jeden typ dat na jiný. Tento způsob zpracování nezohledňuje například stíny na fotografii, které poté zkreslují výsledky převodu, protože software nerozlišuje mezi plochou ve stínu a stejnou plochou na světle.

De-lightery jako Agisoft De-lighter<sup>20</sup>, Lightbrush, CrazyBumps, Bitmap2Material apod. pracují na konvenčním principu, kdy uživatel musí vybarvit oblasti stínů, a program je poté porovná s ostatními částmi textury a sjednotí jejich hodnoty. Pro mou práci byla tato metoda z hlediska efektivity a samostatnosti přelomová (zvláště v kombinaci s první metodou, která tak získala o něco více relevance), ale software Substance Alchemist<sup>21</sup> posunul celý de-lighting ještě dále, kdy k němu využil AI technologii a celý proces zautomatizoval a zredukoval na jedno tlačítko. Právě Substance Alchemist umí ke kompletnímu vygenerování materiálových map využít Neuronové sítě, kdy výsledky i z hlediska časové jsou oproti starým metodám působivé. Pomocí klasifikace dokáže generovat jednotlivé mapy s mnohem větší přesností a výsledky jsou například při porovnání s 3D scany při zohlednění časové náročnosti opravdu působivé.



Obrázek 25 Dream texture addon pro Blender



Obrázek 24 Substance Alchemist AI DeLighting

<sup>20</sup> AGISOFT. AgisoftDeligher[online]. [cit.2023-04-07]. Dostupnéz: <https://www.agisoft.com/index.php?id=53>

<sup>21</sup> ADOBE. SubstanceAlchemist[online]. [cit.2023-04-07]. Dostupnéz:

<https://substance3d.adobe.com/magazine/digitize-your-world-with-substance-alchemist/>

Existují ale jiné AI cesty, které dále zefektivňují a zjednodušují informovanému generalistovi právě tvorbu textur. Nejotevřenější je v těchto expanzích program Blender, který je open-source, a tak na něj vzniká velké množství nezávislých addonů. Jedním z těchto usnadňujících texturování pomocí umělé inteligence je Dream texture addon<sup>22</sup> fungující na základech stable diffusion Depth2Image modelu. Ten je schopen za pomoci Z-depth mapy z pozice kamery a textového promptu automaticky promítnout texturu na celou scénu viditelnou scénou. Limitem je fakt, že program generuje pouze obrázek se zapečenými stíny, který unwrapuje pomocí projekce z aktivní kamery, textura tedy nefunguje správně z odvrácených stran modelu. Může to ale skvěle fungovat při přípravě scény nebo konceptů, protože jednotlivé kroky jsou automatizované v rámci addonu. Jinak je možné si tyto kroky provést ručně i mimo blender, ale díky omezeným výsledkům, ceně blenderu (0,-) a pracnosti to nedává příliš smysl. Vzniká ale podobný plugin pro Cinemu 4D

### 2.7.7 Depht mapy

Depht mapy (vyvinuté jako mnoho dalších umělecko/technických postupů E. Catmullem ) jsou další typ obrazových dat, které je možné se ze 3D scény snadno vyrenderovat jako pass, ale získat je ze záběru je mnohem komplikovanější (pokud ovšem kamera výjimečně nedisponuje lidar scannerem). Tyto data jsou v některých případech velmi nápomocna při compositingu elementů do scény, protože díky odstínu pixelů určují vzdálenost od kamery.



Obrázek 26 Z-deph pass 3D render

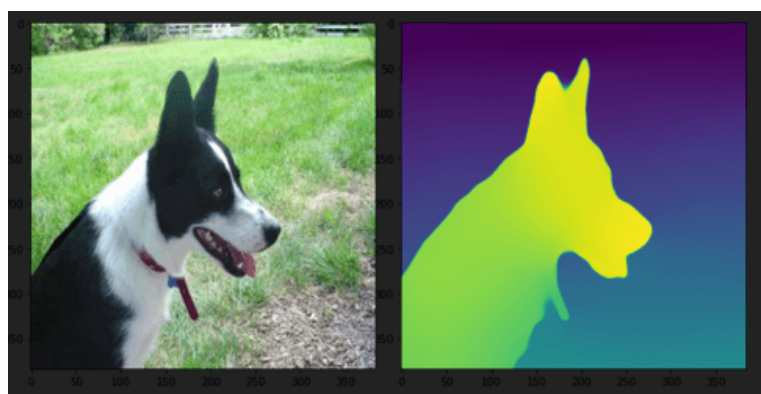
<sup>22</sup> Katri, Carson. GitHub. DreamTextures. [Online] <https://github.com/carson-katri/dream-textures>



Obrázek 27 Vytváření Z-depht mapy pro compositing pomocí masek v Adobe After Effects

Před příchodem technologie AI jsme byli zvyklí je ručně rotoposkovat, nebo kreslit v Photoshopu. To znamenalo využít představivosti, a vybarvit předměty jeden po druhém podle jejich vzdálenosti od kamery do správných odstínů šedi. Vzdálenější objekty jsou tmavší, než ty blíže ke kameře.

AI model midas<sup>23</sup> je v určitých případech řešení, které se vyplatí minimálně zkusit, protože výsledky, které poskytuje jsou poměrně nekonzistentní. Ve většině případů je to ale minimálně dobrý odrazový můstek pro další úpravy. Midas je model vytrénovaný právě k rozpoznávání vzdáleností objektů a generování depht map, který v určitých situacích může ušetřit mnoho času.<sup>24</sup> Existuje i v podobě pluginu do AAE – depht scanner.



Obrázek 28 Depht mapa generovaná pomocí AI Midas

<sup>23</sup> PyTorch.IntelISLMiDaSv2:DepthEstimation[online].Dostupnéz: [https://pytorch.org/hub/intelisl\\_midas\\_v2/](https://pytorch.org/hub/intelisl_midas_v2/)

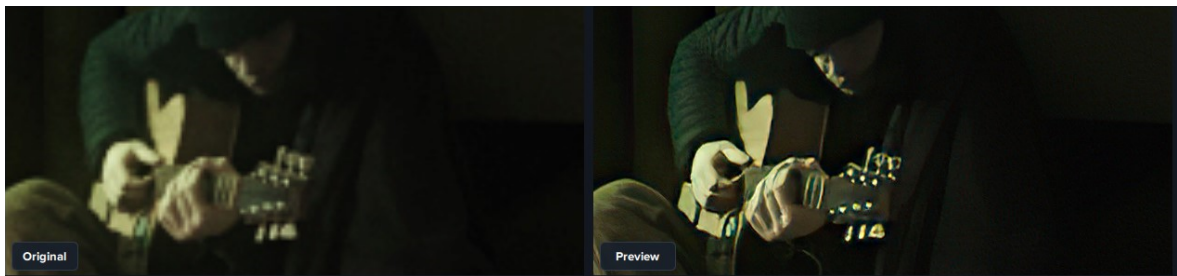
<sup>24</sup> RANFTL, René, KOLTUN, Vladlen a VLADLEN, Kuznetsova. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer. arXiv preprint arXiv:1907.01341, 2019.



### 2.7.8 Optimalizace výkonu

#### **AI up-scaling:**

Většina těchto obrázkových generátorů pracuje s ohledem na rychlost zpracování (ze kterého se odvíjí množství pokusů za stejný čas) lépe s malým rozlišením, (512 – 1024px). Upscaling je spolu s kolorováním jedním z prvních oblastí obrazové postprodukce, kde začalo být AI využíváno. Díky této technologii je možné dopočítávat chybějící pixely obrazu a uměle zvyšovat jeho výslednou kvalitu<sup>25</sup>.



Obrázek 29 Upscalování rozlišení pomocí AI Gigapixel

Oblastí využití je mnoho, ale v praxi nejčastější v mé práci generalisty využívám upscaling pro urychlení renderu scén ze 3D (renderuju v nižším rozlišení, které poté násobím pomocí AI up-scalingu), nebo při nízkém rozlišení zdrojů a elementů oproti materiálu například pro matte-painty či textury. V mé praxi je nejčastěji používaným nástrojem AI gigapixel<sup>26</sup>. V ukázce je možné vidět, jak si gigapixel poradil s jednotlivými složkami snímku. Látka vypadá při upscalu poměrně přesvědčivě, zatímco detaily na obličejích už gigapixel věrohodně dopočítat nedokázal. Vše je ale otázka tréninkových dat, a tak se dá očekávat i na poli gigapixelů velký posun ve šlépějích obrázkových generátorů, které celé odvětví táhnou dopředu.

#### **Interpolace framů:**

V rámci vizuálních efektů i jiných oblastí pohyblivé grafiky je renderování velké téma. S vyvíjejícím se hardwarem se časy vykreslování jednotlivých snímků stále snižují, ale i dnes se nezdá, že se díky renderu může tvůrce dostat do časových problémů. Existuje nespočet metod, jak render optimalizovat, ale navzdory tomu se často vyskytne situace, kdy už není kam optimalizovat, a přesto render jednoho okénka zvláště pokud je

<sup>25</sup> HARIS, Muhammad, SHAKHNAROVICH, Gregory a UKITA, Norimichi. Deep Back-Projection Networks for Super-Resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

<sup>26</sup>Topaz Labs. Topaz Labs: Gigapixel AI [online]. Dostupné z: <https://topazlabs.com/gigapixel-ai/>

potřeba výstupu ve vysokém rozlišení trvá desítky minut. Lidský mozek vnímá 24 snímků za sekundu (při dodržení rozostření pohybu) jako plynulý pohyb, několikasekundový záběr tak může trvat klidně i desítky hodin.

Podobně jako v mnoha dalších oblastech, i pro dopočítávání chybějících mezi snímků existuje vedle nového přístupu i stará konvenční technologie.

Optical flow je softwarová metoda určení pohybu mezi snímky. Základem je porovnávání za sebou následujících snímků, a pomocí analýz vlastností obrazu, jako je srovnání kontrastu a textury oblastí a z nich vyjádření pohybových vektorů. Dílčích přístupů k optical flow je více, ale ve zjednodušené formě se z motion vektorů vypočítává pozice jednotlivých pixelů mezi jednotlivými snímky. Vznikne tak více snímků a tím pádem plynulejší pohyb. Nevýhoda metody je častý vznik různých nahodilých fragmentů, deformací a rozmazání například na neostrých hranách jako jsou odlesky, motion blur nebo jinak nejasně definované plochy či při vnitro záběrové změně nasvícení, kde program mylně interpretuje změny v pixelech jako pohyb<sup>27</sup>. Další nevýhodou je strojová nepřirozenost výsledného pohybu, pohyb pixelů ve snímcích je totiž navyšován čistě matematicky a lineárně. Metodu optical flow posílenou o gpu akceleraci používá ve své době velice oblíbený plugin Twixtor<sup>28</sup>.

Dnes existují modernější a ve všech ohledech lepší možnosti, jak interpolovat framy a zvýšit tak snímkovou frekvenci. Dvě hlavní, obě založené na neuronových sítích je Flow frames a Dain AI. Výsledky obou jsou velice přesvědčivé a rozdíly se hledají jen těžko, přestože technologie za oběma softwary je odlišná, oba jsou založeny na strojovém učení.

Dain AI<sup>29</sup> vylepšuje původní optical flow metodu pomocí dalších dat, které si odvozuje pomocí umělé inteligence (Depth papy, Kontextové vlastnosti, interpolační jádra)<sup>30</sup>, Flow frames tvoří abstraktní reprezentaci pohybu objektů, pomocí které vypočítává chybějící snímky a je schopný pracovat velice rychle díky GPU akceleraci<sup>31</sup>. Je pouhou otázkou času, než se tato řešení podaří integrovat přímo do kompozitorských nástrojů jako kdysi právě Twixtor.

---

<sup>27</sup>BAKER, Simon et al. A database and evaluation methodology for optical flow. International Journal of Computer Vision, 2011, 92(1), 1-31.

<sup>28</sup>RE: VisionEffects.Twixtor: TimeRemapping&OpticalFlowSoftware[online]. Dostupnéz:

<https://revisionfx.com/products/twixtor/>

<sup>29</sup>BAO, Wen. Depth-

AwareVideoFrameInterpolation(DAIN)[online]. Dostupnéz: <https://sites.google.com/view/wenbobao/dain>

<sup>30</sup>DAIN-APP: Depth-AwareVideoFrameInterpolation[online]. GitHubRepository. Dostupnéz:

<https://github.com/GRisk/Dain-App>

<sup>31</sup>Flowframes.Flowframes: AI-PoweredFrameInterpolation[online]. GitHubRepository. Dostupnéz:

<https://github.com/n00mkrad/flowframes>

### Image denoising:

Při renderování 3D grafiky pomocí path tracingu vzniká díky zjednodušené funkčnosti světelných paprsků v 3D prostoru specifický druh šumu. Ke zbavení se tohoto šumu vede tradičně několik cest. Jednou z nich je zvýšit počet světelných samplů, to může být ale náročné na renderovací čas. Druhou možností je využití denoiserů a právě v této oblasti se hojně využívají technologie AI. Nejčastěji používaným 3D denoiserem současnosti je Optix od Nvidia, který se v jednotlivých verzích snaží postupně zajistit fungování v reálném čase. Díky aplikaci AI je rychlý a výsledky jsou oproti starším metodám odstraňování šumu čistší<sup>32</sup>.

AI denoisery je možné použít i mimo 3D rendering, ale umožňují odšumovat i natočený materiál, jako například AI denoiser od TopazLabs.



Obrázek 30 ukázka AI denoiseru (vlevo před, vpravo po)

### 2.7.9 Simulace (Fluid, volume, particles)

Jako generalista se simulacemi pracuji běžně. Ve velkých studiích jsou i na tuto práci zaměřující se specialisté.

Velký potenciál pro využití strojového učení vnímám právě v oblasti 3D simulací. Dnes je totiž k simulacím už kapalin, particlů, nebo volumetrických fluidů zpravidla využíván

---

<sup>32</sup> CHAITANYA, Chakravarty R. Alla et al. Interactive reconstruction of Monte Carlo image sequences using a recurrent denoising autoencoder. ACM Transactions on Graphics (TOG), 2017, 36(4), 98.



hrubý výkon hardwaru, a vypočítání detailních simulací zabere nemalý čas. Prostor je podobně jako u interpolace snímků využívat tyto metody pro predikování pohybů v 3D prostoru. Bohužel v současné situaci existují pouze experimentální demo, které nejsou dostupné pro veřejnost<sup>33</sup>. Po přechodu na gpu akceleraci by byla technologie AI předikovaných simulací další skok, který by podstatně urychlil práci.

Expertíza specialistů totiž do velké míry vychází právě z časové náročnosti každé jednotlivé simulace, a je tedy nutné, aby byly parametry simulace nastaveny co nejpřesněji a jednotlivě opakované pokusy byly omezeny na minimum.

Predikováním pohybu kapalin se na Siggraph konferenci 2020 zabýval paper Wave Curves, a Nvidia pracuje na NeuralVDB, který pomocí neuronových sítí kombinovaných s gpu zvládá zefektivňovat simulace včetně kapalin, fluidů, měkkých i pevných těles.

### 2.7.10 Facial replacement, de-aging

Face replacement je technika využívající počítačovou grafiku, k nahrazení tváře herce ve filmu jinou tváří. Tuto techniku lze využít pro různé účely, například když herec nebyl k dispozici pro některé záběry nebo když se v průběhu natáčení změnil jeho vzhled, případně pokud ve scéně hraje kaskadér, či je herec mrtvý.

V rámci této oblasti si můžeme problematiku face replacementu rozdělit do tří kategorií. Kompletní výměna tváře herce, de-aging, a úprava tváře herce jako je přidání jizev. Poslední kategorii se v této kapitole nebudu věnovat, protože využívá zcela jiné postupy připomínající



Obrázek 31 srovnání skutečného herce, a 3D dubla z filmu Star Wars: Rogue One (2016, Gareth Edwards)

<sup>33</sup>UM, Taesoo, THUREY, Nils, a KIM, Theodore. Data-driven fluids: Wave-curve networks for water simulation. ACM Transactions on Graphics (TOG), 2020, 39(4), 65.

spíše retuše. Z hlediska konvenčního přístupu existují dvě možnosti, jak k problému přistoupit. První je, natočit si hereckou akci s druhým člověkem, kdy je nutné zachovat stejný charakter svícení, poté jeho tvář vymaskovat, vystabilizovat a nakomponovat v kompozičním softwaru. Tato metoda funguje dobře na záběry s malým pohybem, ale ne vždy je nová tvář reálně k dispozici, nezářídka se vyskytují situace, kdy požadovaná tvář patří například herci, který už nežije. Jako příklad poslouží Tarkin ze Star Wars: Rogue One. Jeho tvář byla vymodelována několikačlenným teamem specialistů na základě starých fotografií. Přestože se jedná o fenomenální práci vyžadující velké množství expertíz, je zde přítomný problém, který se při torbě charakterů objevuje poměrně často – uncanny valley.

\*Uncanny valley je teorie v oboru interakce člověk-robot, která popisuje jev, kdy se robot nebo jiná ne-lidská entita stávají pro vnímání člověka stále méně přijatelná, jak se stávají vizuálně více podobnými lidem, až dosáhnou bodu téměř přesvědčivé lidskosti, V tomto bodu, kdy je mozek nedokáže vnímat jako stylizaci, ale ani jako lidi, se stávají pro člověka velmi neklidnými.

Lidský mozek totiž nevnímá odchylky v realitě u různých věcí stejně. Na tváře je člověk obzvláště citlivý díky tomu, že je vídá často, a je nucen číst drobné výrazy, aby ve společnosti přežil. Proto je ve vizuálních efektech velice těžké udělat stoprocentně přesvědčivou fotorealistickou tvář, protože jakákoliv odchylka od reality celý efekt rozbíjí.

Udělat proto přesvědčivou detailní 3D tvář, navíc s nějakou hereckou akcí je výzva pro celý team specialistů. V rámci de-agingu je situace podobná, ale je možné se držet původních rysů tváře a zachovat například oči, což velmi pomáhá udržení věrohodnosti stejně tak je možné si vystačit jen s 2D technologiemi a v některých případech menší produkce pouze dopravit make-up.

Zde ale přichází na scénu technologie Deep fake.

\*Deep fake jsou technologie, které pomocí strojového učení upravují obličej.



Obrázek 32 ukázka deep fake technologie – vložení tváře Nicolase Cage

Pomocí této technologie je možné pomocí syntézy velkého množství fotografií výchozího a cílového obličejové dotrénovat model tak, aby touto tváří nahradil tvář v cílovém videu.

K celému procesu je tedy potřeba co největšího množství dat v co nejvariabilnějších podmínkách jako nasvícení, výrazů tváře, úhlů a vzdáleností od kamery.

K tvorbě deepfaků lze přistoupit dvěma způsoby generativních algoritmů. Tím jednodušším a častějším je deep auto encoder. Vedle něj existuje ještě generative adversarial networks (GANs) ale ten se v praxi v kontextu deepfaků příliš nepoužívá.

Jádrům deep auto encoderu jsou dva neuronové modely. Encoder a decoder. Encoder hledá podobnosti v obličejích a ty převádí na latentní data která jsem již zmínil dříve v této práci. Na těchto datech model trénuje a poté je převádí do decoderu. Zde se data nového obličejové rekonstruují na původní materiál zdrojového obličejové<sup>34</sup>.

Kvalita výsledku je dána množstvím dat, délkou tréninku multiplikovanou HW výkonem a typem cílového záběru.

Limity se objevují například v situacích, kdy je herec snímán z profilu, nebo když se rychle mění jeho nasvícení. Někdy je ve výstupech rozpoznatelná mírná chvilková asynchronicita v pohledu očí, kdy se každé oko dívá jiným směrem, při deep fake je také mírně potlačena intenzita výrazů tváře, které působí oproti reálnému herci více stoickým dojmem. Další limitací je výstupní rozlišení, které je poměrně malé a na scénu tak přichází AI upscaling, který ale přidává obrazu mírně nepřírodně čistý „vyžehlený“ charakter v detailech. Je proto

---

<sup>34</sup>SUWAJANAKORN, Supasorn, SEITZ, Steven M., a KEMELMACHER-SHLIZERMAN, Ira. Synthesizing Obama: Learning lip sync from audio. ACM Transactions on Graphics (TOG), 2017, 36(4), 95.

lepší tento přístup využívat ve větších celcích a vyhnout se detailům. Při deep fake postupu také není možné efektivně nahradit uši, které tím pádem zůstávají původní a je nutné s tím počítat. Drobnou komplikací, se kterou je nutno počítat je občas blending nového obličeje na původní tvář, kdy může mít kůže jiný odstín a je tak nutno tyto odchylky upravit v kompozitingu<sup>35</sup>.

Úpravy deep fake výstupu v další fázi postprodukce je běžný postup, zde je tento výstup využitý spíš jako kompoziční pass. Nejběžnější softwary na tvorbu deep faků jsou DeepFace lab, FakeApp, nebo NeuralFace.

Pro zakrytí chyb deep fake záběrů dobře fungují přehnaná herecká gesta, která odvádí divákovu pozornost od obličeje. Navzdory zmíněným limitům je z hlediska uncanny valley deepfake nástobně efektivnější a přesvědčivější než 3D animace.

### 2.7.11 3D scanning

3D skenování je proces zachycení tvaru a geometrie reálného objektu do digitálního 3D modelu včetně textur. Zahrnuje zachycení datových bodů z povrchu objektu a použití softwaru k vytvoření 3D reprezentace objektu. Tato technologie je v rámci VFX široce využívána.

Jako obvykle existuje více cest, jakými geometrii z objektu zaznamenat. Ke skenování je možné využít LIDAR scannery (Light Detection and Ranging), které ke skenování využívají laserové světlo, kterým měří vzdálenosti. Tyto systémy vysílají a zachytávají laserové pulzy, které se odrážejí z objektu a vracejí se zpět do senzoru, zaznamenaná data se využívají k výpočtu polohy bodů na povrchu objektu. Ze vzniklého mračna bodů se poté ve specializovaném softwaru vypočítá mesh (3D model).

---

<sup>35</sup> CORRIDOR CREW. Professional VFX Artists Explain how to Spot Fake Videos [video]. YouTube, 2022. Dostupné z: [https://www.youtube.com/watch?v=KqIW1TczgsA&ab\\_channel=CorridorCrew](https://www.youtube.com/watch?v=KqIW1TczgsA&ab_channel=CorridorCrew)

Mesh je datová 3D struktura, která představuje objekt pomocí řady vzájemně propojených polygonálních ploch.



Obrázek 33 mesh vygenerovaný za pomoci fotogrammetrie

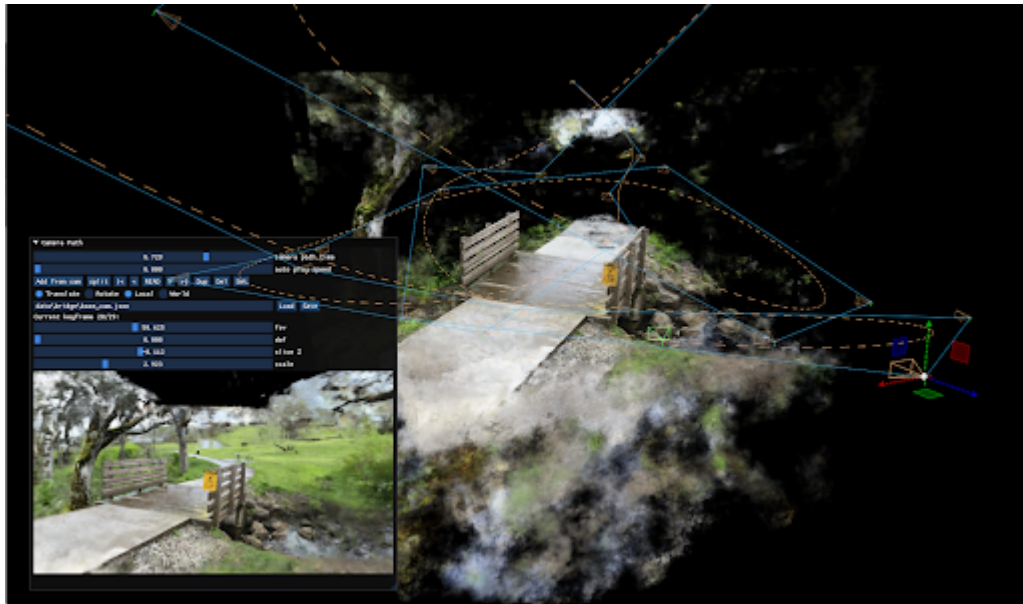
Druhou možností, jak konvenčně zaznamenat mračno bodů je fotogrammetrie.

Princip spočívá v pozorování stejných bodů z různých úhlů. Fotogrammetrie využívá mnoho fotografií objektu pod různými úhly, ve kterých pomocí kontrastních ploch s texturou hledá totožné body. Z těch poté pomocí geometrie a optiky vypočítává referenční body. Výsledkem je opět mračno bodů spolu s texturou připravené na další zpracování.

V praxi je možné použít kombinaci obou metod. LIDAR dodává lepší výsledky při měření drobných detailů, zatímco fotogrammetrie je vhodná na velké rozlehlé objekty jako například architektura.

Výpočty odrazů paprsků světla, na kterých jsou obě metody založeny ale přestávají fungovat v případech, kdy se paprsky odráží od reflexních ploch jako jsou sklo, voda, kovy a jinak světlo deformující materiály. V takových případech dochází ke zkreslení a výsledkem jsou nehezke meshe s různými nesprávně zaznamenanými strukturami.

Tuto situaci postupně řeší AI alternativa NeRF (neural radiance fields)<sup>36</sup>.



Obrázek 34 neural radiance field (NeRF)

Při použití této metody nevzniká mesh v klasickém slova smyslu, ale vzniká volumetrický radiance field. Je to specifická forma mapování 3D prostoru, kterou je možné zobrazit a vyrenderovat pouze uvnitř softwaru na bázi NeRF. Obrovskou výhodou je, že díky tomu, že se nejedná o matematickou metodu, dokáže tento software zachytit a realisticky zobrazit i lesklé a průsvitné plochy, které u předchozích metod deformují výsledky. V rámci NeRF se tak přímo výstup ze skenu bez jakýchkoli úprav limitně blíží hyperrealismu, protože podobně jako AI interpolace se za pomoci sekvence snímků, kombinované s modely AI z omezeného množství snímků generuje prostor mezi nimi<sup>37</sup>.

NeRF v současné době viditelně posouvají firmy LumaLabsAI se svým Luma AI a Nvidia se svým NGP instant NeRF. Další možností, jak se k technologii dostat je internetová platforma NeRF studio.

Současná hlavní limitace NeRF je, že výstupem není 3D model v žádném využitelném a editovatelném formátu kde by se s ním dalo pracovat. Její využití pro 3D je krkolomné a až čas ukáže, jak efektivně se podaří implementovat NeRF do tvůrčího procesu 3D skenování.

<sup>36</sup> Stephens, Jonathan. Getting Started with NVIDIA Instant NeRFs. místo neznámé : Nvidia Developer. Sv. <https://developer.nvidia.com/blog/getting-started-with-nvidia-instant-nerfs/>

<sup>37</sup>Tancik,Matthew.NeRFresearch-matthewtancik.matthewtancik.com.[Online]<https://www.matthewtancik.com/>.

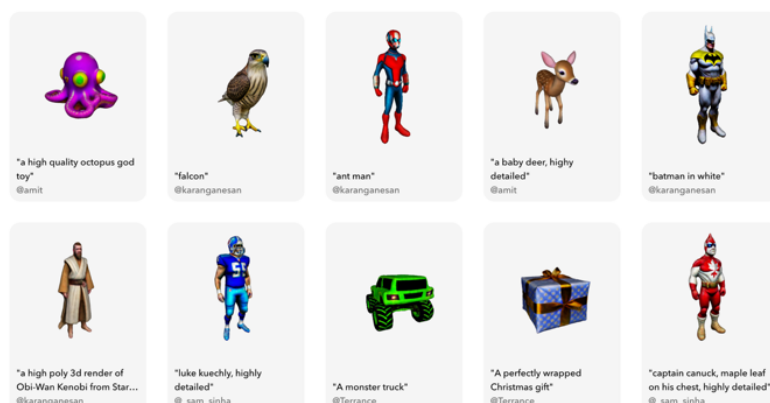


V současné době jediný způsob využití technologie NeRF je importovat do platformy animovanou kameru a poté z ní vyrenderovat video soubor<sup>38</sup>.

### 2.7.12 3D modelování

Způsob využití AI pro generování 3D modelů je podobný, jako při používání obrázkových generátorů. Vývoj v této oblasti postupuje velmi rychle a každých několik měsíců přijde nějaký posun. Dnes je vývojových větví hned několik. Svoje výzkumy dělají velké společnosti jako LumaLabsAI (LumaAI), Nvidia (Magic3D)<sup>39</sup>, nebo Open AI (Point-E).

Principiálně jsou 3D generátory kombinací dvou systémů. V první fázi je využita nějaká forma modelu Text2image, kdy se pomocí promptu generuje obrázek<sup>40</sup>. V následující fázi se pomocí nějaké formy modelu Img2Mesh tvoří z vygenerovaného obrázku pointcloud, jehož zpracování na 3D mesh s texturou je již formalita a léta existující technologie.



Obrázek 35 modely generované pomocí Luma AI

V současné době tyto generátory nemohou ve většině situací konkurovat svými výstupy specialistům na 3D modelování - postrádají detaily, jejich textura je v nízkém rozlišení a disponují pouze diffuse a normal mapou. Dalším nedostatkem je kvalita topologie, kterou mají v aktuální fázi modely žalostnou, a připomíná spíše nepřilíš kvalitní 3D scany. Teoretický smysl dává generování modelů pro nahození assetů do vzdálených koutů 3D

<sup>38</sup>WhyTHISistheFutureofImagery?[Online]CorridorDigital.[https://www.youtube.com/watch?v=YX5AoaWrowY&ab\\_channel=CorridorCrew](https://www.youtube.com/watch?v=YX5AoaWrowY&ab_channel=CorridorCrew).

<sup>39</sup> Chen-Hsuan Lin, Jun Gao, Luming Tang, Towaki Takikawa, Xiaohui Zeng, Xun Huang, Karsten Kreis, Sanja Fidler, Ming-Yu Liu, Tsung-Yi Lin. Magic3D. Magic3D: High-Resolution Text-to-3D Content Creation. místo neznámé : NVIDIA Corporation, 2022. Sv. <https://research.nvidia.com/labs/dir/magic3d/>.

<sup>40</sup> Gao, J., Shen, T., Wang, Z., Chen, W., Yin, K., Li, D. GET3D: A Generative Model of High Quality 3D. [Online] 2022. <https://nv-tlabs.github.io/GET3D/>.

scény nebo layouty. Může být ale otázkou času, kdy se technologie posune do fáze, kdy bude chrlit kvalitně zpracované modely a její využití bude dávat smysl i pro assety v detailnějších záběrech.

### 2.7.13 Psaní scriptů

Nečekaně zajímavou alternativou k ručnímu skriptování je využití textového generátoru opět od firmy OpenAI – ChatGPT<sup>41</sup>. Zmíněný nástroj je chatbot, který pracuje aktuálně na jazykovém modelu GPT-4 (ve vývoji jsou další novější generace tohoto modelu) a je schopen na zadání generovat více či méně přesný text. Generovaný text touto AI produkuje často faktické nepřesnosti a chyby, ale i tak způsobil ve světě velký rozruch, díky schopnosti zapamatovat si předchozí výpovědi a pomocí koherentního textu reagovat na uživatelské vstupy.

ChatGPT zvládá na zadání generovat scénáře, písničky či dokonce diplomové práce. Velkým překvapením pro mnohé bylo, když se ukázalo, že ChatGPT zná programovací jazyky, a umí na povel psát scripty do nástroj, který mu uživatel zadá<sup>42</sup>. Pro generalistu, může být tato možnost obrovskou pomocí. Pomocí skriptů totiž je možné automatizovat části procesu, které se například opakují, a tak pálí generalistovi čas. Skriptování je pro člověka, který se ním aktivně nezabývá (jako jsem například já) poměrně velká technická bariéra, kterou může AI pomoci překlenout. Podobně jako v jiných oblastech i u skriptování dělá ChatGPT chyby, ale může to být dobrý odrazový můstek. Navíc ověřit, jestli script dělá, co má se dá ve většině případů poměrně jednoduše metodou pokus/omyl. Na následujícím funkčním expressionu do adobe after effects ilustruji, jak jednoduše se toto základní skriptování stalo.


---




<sup>41</sup> <https://openai.com/blog/chatgpt-plus>


<sup>42</sup> Kampff, Stephen. Basic scripting with ChatGPT. Turning Ideas Into Animations With ChatGPT and After Effects. místo neznámé : shorturl.at/kBV01, 2022



Model: GPT-4

 Napiš expression do softwaru Adobe after effects, který při nakopírování do position layeru rozanimuje vrstvu na základě sinusoidy, kdy se bude pohybovat ve vteřinových intervalech o 50 pixelů nahoru a poté dolů.

 Můžete použít následující expression pro position atribut vrstvy v Adobe After Effects:  

```
javascript  Copy code
```

```
amplituda = 50;  
frekvence = 1;  
sinusoidni_pohyb = Math.sin(time * frekvence * 2 * Math.PI) * amplituda;  
[value[0], value[1] - sinusoidni_pohyb]
```

Tento expression vytvoří sinusoidní pohyb ve vertikálním směru (nahoru a dolů) s amplitudou 50 pixelů a periodou 1 vteřina.

1. V After Effects, vyberte vrstvu, kterou chcete animovat.
2. Stiskněte "P" na klávesnici, abyste zobrazili "Position" atribut vrstvy.
3. Alt+klikněte na hodnotu "Position" (stopwatch ikona) pro otevření expression editoru.
4. Zkopírujte a vložte výše uvedený kód do expression editoru.
5. Stiskněte Enter/Return na numerické klávesnici (nebo klikněte mimo textové pole) pro aplikaci expression.

Nyní by měla být vrstva animována s sinusoidním pohybem o 50 pixelů nahoru a dolů v jednosekundových intervalech.

Obrázek 36 Expression funkce pro After Effects, vygenerovaná pomocí GPT 4

### 3 OMEZENÍ AI NÁSTROJŮ

Přes svůj obrovský potenciál se prozatím nacházíme ve fázi, kdy nejsou problémy vyplývající z technologie AI v rámci počítačem generované grafiky kompletně vyřešeny. S nedostatky je nutno počítat, a využívat AI v situacích, kdy vyniknou silné, a zaniknou slabé stránky této technologie.

Slabých stránek vnímám několik, ale většina z nich má podobné vyústění. Tím je nedostatek kontroly.

#### 3.1 Kontrola

Kontrola je v tvorbě důležitý parametr, protože rychlost bez kontroly je pro své nepřesné zacílení méně efektivní. V momentě, kdy jsou výstupy kompletně náhodné, je od určitého momentu rychlost jejich zpracování irelevantní.

##### 3.1.1 Omezené UI

Nedostatek kontroly uvnitř procesu vychází z podstaty funkcionality těchto nástrojů. To že je zde jen vstup a výstup omezuje možnosti ovlivnit výsledek. Většina AI nástrojů pracuje s nějakým druhem promptu, což je poměrně omezené UI. Má to samozřejmě i výhody v podobě technické nenáročnosti použití, avšak pro profesionála je omezené UI spíše komplikací.

Například kontrolovat kompozici obrázku je za pomoci AI generátorů minimálně velmi náročné. Obrázkové generátory v posledních verzích přidávají další možnosti, jak kontrolovat výstupy. Hlavním modelem je práce s vizuálním promptem. To umožňuje nahrát nekompletní skicu, která určí, kde se v obraze mají nacházet jaké objekty. Druhou možností je práce s maskami, pomocí které nástroje dovolují pře-generovat pouze vybranou část obrázku a ostatní zachovat.

##### 3.1.2 Flickering

Většinou jako generalista pracuji s video materiálem. Využití obrázkových generátorů na video/sekvenci obrázků produkuje díky náhodnému umístění zrn v šumu tzv flickering. Blikající sekaný efekt, vycházející z toho, že každý generovaný snímek je originál bez vazby na předchozí a následující. Díky tomu vznikají při pokusech o automatickou stylizaci/úpravu videa pohyblivé artefakty, které omezují plynulost obrazu. V současné době se používá vypnutí random noise paternu v kombinaci s Deflicker nástrojem v Davinci, který problém

částečně omezuje<sup>43</sup>. V praxi je zatím pro generalistu většinou nejpraktičtější omezit se pouze na obrázky a ty používat dále v rámci konvenčních postupů.

### 3.1.3 Trénování vlastních modelů

Řešení kontroly výstupů může nabídnout v určitých situacích trénování vlastního modelu, ale to je pro generalistu mého typu technicky i časově náročné. Navíc ani vytrénování vlastního modelu nezaručuje kontrolu výstupu, pouze je to další možnost, jak ovlivnit zpracování vstupu.

Mnoho problémů vychází z nematematické, pro člověka abstraktní podstaty těchto AI procesů.

### 3.1.4 Datové přenosy

Podobným problémem, který se vývojari AI nástrojů v současnosti snaží vyřešit je převod dat na konvenčně využitelná data. Tento problém se konkrétně vyskytuje například v rámci 3D skenovacích modelů NeFR<sup>44</sup>.

Výstupem tohoto modelu je Neural Radiance Field, což je formát, který nedokážou běžné softwary zobrazit. Jedinou možností, jak v současnosti výhod těchto NeFR nástrojů využívat, je renderovat vše jako podklad přímo z nich a pracovat s dílčím výstupem jako se samostatným obrazovým zdrojem<sup>45</sup>.

Existují zde tedy dva oddělené softwarové světy, a je nyní výzvou pro vývojáře a softwarové inženýry tyto světy propojit.

---

<sup>43</sup>CorridorDigital.com.[Online]23.Leden2023.

[https://www.youtube.com/watch?v=\\_9LX9HSQkWo&t=12s&ab\\_channel=CorridorCrew](https://www.youtube.com/watch?v=_9LX9HSQkWo&t=12s&ab_channel=CorridorCrew).

<sup>44</sup>Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. San Diego, Kalifornie, USA : autor neznámý, 19. Březen 2020.[www.matthewtancik.com/nerf](http://www.matthewtancik.com/nerf)

<sup>45</sup>CorridorCrew.Youtube.[Online]Listopad2022.[https://www.youtube.com/watch?v=YX5AoaWrowY&ab\\_channel=CorridorCrew](https://www.youtube.com/watch?v=YX5AoaWrowY&ab_channel=CorridorCrew).

## **II. PRAKTICKÁ ČÁST**

## 4 ZRYCHLENÍ PRÁCE CGI GENERALISTY POMOCÍ AI NÁSTROJŮ

V této kapitole na případové studii demonstruji rozdíly ve způsobu práce za pomoci AI a konvenčních nástrojů v některých modelových situacích.

V rámci obsáhlosti oboru VFX a možností které mohou nastat zobecním problematiku na práci na konkrétním projektu. Zaměřím se na oblasti, ve které jsou AI v současné době pro mou osobní praxi nejvyužitelnější.

### 4.1 Zadání experimentu

Na následujícím experimentu se pokusím ilustrovat a dokázat, že využití AI může být v současnosti velkým multiplikátorem schopností a možností generalisty, který tak podstatně urychluje některé části procesu, a nahrazuje nedostatek zkušeností v konkrétních oblastech. Na druhou stranu předpokládám i nevýhody vyplývající z funkcionality nahodilých iterativních procesů při AI tvorbě.

#### 4.1.1 Cíl experimentu

Cíl experimentu bude v jednotlivých částech procesu tvorby najít ideální přístup, a využít AI k minimalizování času jednotlivých kroků při snaze udržet obrazovou kvalitu. AI nástroje pro vytváření a úpravu podkladů, budu kombinovat s konvenčními postupy.

Vedle snahy najít cestu zkrácení času při udržení kvality si zde kladu za cíl reflektovat způsoby a možnosti AI, jakými dnes generalisté disponují, a které jim dovolují zvyšovat svou konkurenceschopnost vůči teamům specialistů.

### 4.2 Metodologie experimentu

#### 4.2.1 Parametry měření

Kvůli komplexnosti celé práce budu hodnotit pouze jediný parametr. A sice **čas výroby**. Vše ostatní je totiž subjektivní a v podmínkách této práce těžko měřitelné. Přesto ze svého pohledu srovnám výsledky z hlediska subjektivního vnímání kvality.

#### 4.2.2 Slabiny a odchylky experimentu

Samotný výsledek experimentu bude nejspíše do jisté míry zavádějící, protože celá problematika tvorby generalisty zahrnuje velké množství různorodých situací a problémů,

kterým může čelit a u kterých by bylo možné AI využít různorodým způsobem. Ve své práci se zaměřím na ty části konkrétního projektu, kde využití AI nástrojů dává smysl mě osobně. Experiment si neklade za cíl zohlednit lidský faktor, který v praxi představuje různá úroveň výstupů různých generalistů a specialistů vyplývající z jejich zkušeností. Tento fakt, ze kterého vychází omezená replikovatelnost experimentu nelze v mých podmínkách zohlednit.

### 4.3 Případové studie

Srovnat čas konvenčního postupu s časem využití AI při jednotlivých činnostech ve VFX.

#### 4.3.1 Rendering: AI post process motion blur

##### Problém/situace:

V rámci svého modelového automotive projektu jsem se ocitl v situaci, kdy se ve scéně vyžadující motion blur nachází animovaný VDB objekt bez velocity dat.

Díky absenci velocity dat na objekt není možné aplikovat motion blur přímou renderu jako ve zbytku scény. Z důvodu voxelové podstaty VDB objektů nelze použít ani motion vector pass, který pracuje pouze s 3D geometrií. Snímková frekvence je 30fps, což je pro využití post procesového pixel motion blur vzhledem ke scéně nedostatečné.

##### Cíl:

V tomto případě je mým cílem nějakým způsobem získat motion blur volumetrické animace, vedlejším produktem následujícího setupu bude pravděpodobně rychlejší render scény, ale na ten se zaměřím v samostatné studii.

##### Render settings:

Rozhodl jsem se celou scénu vyrenderovat bez VDB objektu, a tento objekt si vyrenderovat samostatně v modifikované verzi scény, kde deaktivuji všechny vlastnosti materiálů, a tak zdecimuji počet rays ve scéně na minimum. Celou tuto scénu přeoptimalizuji pro striktně VDB volumetrický objekt (vypnu reflections, emissions, diffuse...). Abych získal pouze Pass kouře interagujícího s prostředím, nastavím všechny objekty jako matte shadows objekty.

##### Pixel motion blur:

Kvalita vypočítaného motion bluru z pohybu jednotlivých pixelů vychází z množství dat, které nástroj pro pohyb dostane. Pokud získám větší počet snímků, bude z nich vypočítaný

motion blur obsahovat méně (žádné) chyb jako jsou deformace, artefakty nebo rozmazané kontury objektů. Tyto chyby vyplývají z toho, že nástroj hledá ve snímcích pohyby pixelů. Pokud jsou ale snímky příliš odlišné, nedokáže v nich pixel motion blur rozlišit pohyb obrazových struktur správně, a vzniká rozmazání nahodilými směry které tvoří duchy a artefakty.

Cíl je mít pro kvalitní výpočet pixelového motion bluru co největší množství dat (snímků) které tento nástroj bude moci zpracovat. Možností může být nastavení vyšší snímkovací frekvence ze 3D rendereru. Fungovalo by to, ale zároveň by násobně vzrostl časový náklad renderu.

### **AI vstup, zvýšení framerate pomocí interpolace framů**

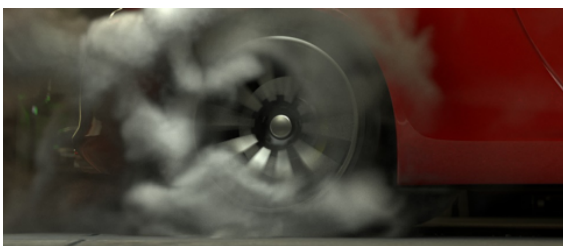
AI software Flowframes, o kterém jsem již psal v teoretické části mi dovolí zvýšit frekvenci 30 snímků za sekundu na frekvenci v libovolné velikosti.



Obrázek 40 původní 3D render pass kouře



Obrázek 39 pass kouře s přidáním motion blur pomocí AI nástrojů



Obrázek 37 původní render celé scény



Obrázek 38 Scéna s přidáním kouře a motion blurem

Pro interpolaci jsem využil AI model RIFE 3.1 v3.

Sekvenci o frekvenci 240 snímků za sekundu jsem rozmazal pomocí pixel motion blur a tuto vrstvu poté zkombinoval s renderem bez VDB kouře v AAE Díky tomuto postupu jsem měl navíc možnost dodatečně upravit intenzitu kouře.

Mimo získání většího množství dat na tvorbu motion bluru se díky tomuto rozdělení scény na dva elementy 10x zkrátil render time (30 min frame s volume objektem ve scéně vs 3min bez volume objektu).

Tabulka 1 tabulka srovnání časů nutných pro realizaci renderu

Render celé scény s VDB	Render scény bez VDB	Render pouze VDB	AI součet
<b>30 min/frame</b>	3 min/frame	1 min/frame	<b>4 min/frame</b>

### 4.3.2 Rotoscoping

Ve svém projektu využívám 2D karty s vyrotoscopovanými postavami pro navození živějšího prostředí a podpoře jeho uvěřitelnosti. „*Mozku, tohle není animák*“ podsouvám tím divákovi. Tyto karty natáčím s nasvícením podobným, jako plánuji v odpovídající 3d scéně. Rotoscoping je ruční práce, která nevyžaduje vysokou úroveň expertízy, ale je velmi náročná na čas. V rámci mé studie nepotřebuji vysokou kvalitu výsledku a přesný klíč jednotlivých vlasů, postačí mi jednoduchá maska s jednotným featherem (rozmazáním okrajů masky).

Využití Mochy pro vymaskování postavy.

Abych zodpovědně ověřil, že je AI maskování opravdu efektivnější cestou, provedu celý proces dvakrát. Poprvé využiji konvenční a osvědčený software Mocha, který je široce používaným řešením na tvorbu masek. Poté na totožný klip využiji automatickou webovou AI aplikaci Runway.

Pro zjednodušení jsem přizpůsobil podmínky natáčení (kostým, pozadí) jeho účelu. Abych se vyhnul složitým tvarům jako jsou vlasy, použil jsem kapuci, a podobně jsem upravil i perspektivu, abych redukoval komplexitu siluety postavy a pracoval s co nejjednodušším tvarem.

První obrázek je ručně vyrotoscopovaný za pomoci programu Mocha, druhý obrázek je vyrotoscopovaný pomocí AI Runway.

Celkový čas Rotoscopování patnácti sekundového záběru v Mocha byl v mém případě dvě a půl hodiny, s tím, že se jednalo o záběr s relativně malými a pomalými pohyby, takže nebyl problém s trackováním masek. Pomocí AI softwaru Runway jsem postup zvládl za necelých deset minut. Rozdíl v kvalitě v tomto případě nelze okem zpozorovat, Free verze, kterou



jsem využil a má sice omezení v podobě pouze HD exportu což se projevuje měkkými okraji masky, ale v tomto případě to je dostačující.

Tabulka 2 srovnání časů nutných pro rotoscoping scény za využití rozdílných nástrojů.

Mocha Rotoscoping	Runway AI Rotoscoping
152 minut	9 minut



Obrázek 41 rotoscoping pomocí nástroje Mocha



Obrázek 42 rotoscoping pomocí nástroje Runway

### 4.3.3 Retušování

Runway má aktuálně poměrně rostoucí škálu možností využití. Dalším z nich je automatické retušování vybraných objektů. Tento nástroj jsem využil v rámci jiného projektu. A sice retušování log z hudebního videoklipu. Retušování obecně podobně jako tvorba masek není příliš tvůrčí proces, navíc bývá v některých případech časově náročný (například retuše objektu za prosvítajícími vlasy herece). Problematické jsou také například situace, kdy v obraze nejsou odpovídající oblasti, které by bylo možné na retušovanou oblast klonovat v těchto případech je potřeba tuto oblast domalovávat a často upravovat snímky zvlášť.

Pro retuš jednotlivých log jsem využil v době psaní této práce nástroj Inpainting, který je součástí testovací verze dříve zmíněného AI generátoru Runway. Postup jsem srovnával s využitím nástrojové sady Adobe v kombinaci s pluginy od Red Giant (spot clone tracker) a Boris FX Mocha. Záběrů bylo celkem sedm, ale pro srovnání času jsem využil jen ten nejkomplicovanější, kde se logo prolíná s vlasy herece, které ho během záběru opakovaně zakrývají a odkrývají.

Jednotlivé vlasy jsem vyrotoscoval pomocí nástroje Mocha. Clean plate oblasti v pozadí jsem vytvořil pomocí klonovacího razítka v Adobe Photoshop. Také jsem musel animovat hodnotu expozice, protože herec vrhá stín, a jeho pohyb ovlivňuje nasvícení retušované oblasti. To vše v kombinaci s dalšíma efektama a maskama v After Effects. Celá retuš záběru trvala při využití této konvenční metody 3 hodiny a 15 minut.

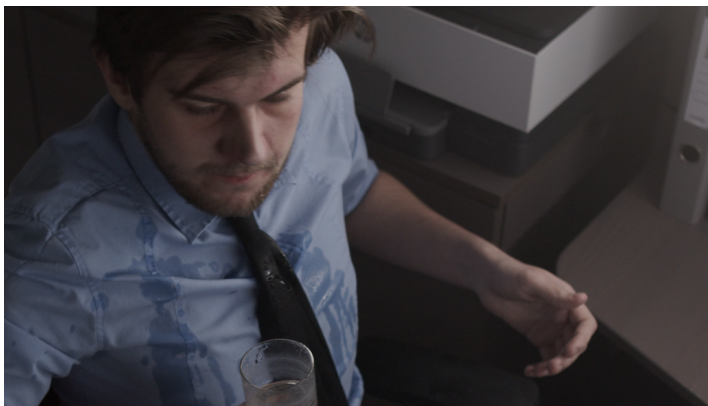
Runaway ve své neplacené verzi umožňuje export pouze 720p, a tak jsem v rámci tohoto nástroje používal pouze výřez obrazu s logem určeným k retuši. V nástroji je možné pomocí štětce označit oblast, kterou je třeba vyretušovat a vše ostatní, včetně trackingu i maskování ostatních objektů přecházejících přes oblast se děje automaticky. U velkých změn probíhajících v obraze je potřeba zkontrolovat, jestli se všechny snímky vyretušovali správně, a v případě artefaktů, vznikajících, když AI nedokáže správně rozpoznat objekt při nějaké podstatné změně, je nutno vytvořit klíčový snímek a znovu pomocí štětce objekt zamalovat.

Vše funguje skvěle v případě, že se v obraze nemění příliš mnoho struktur. Výsledný export je dobrý podklad, ale bylo nakonec nutné udělat úpravy pomocí masek a kopie některých snímků, kterými jsem nahradil ty, na kterých vznikly artefakty. Přes tyto problémy využití dat, které vygenerovala AI podstatně zrychlilo celý proces. Překvapivě dobře si nástroj poradil s oblastmi, kde se objekt překrývá s vlasy, což je aspekt, který sebral při využití

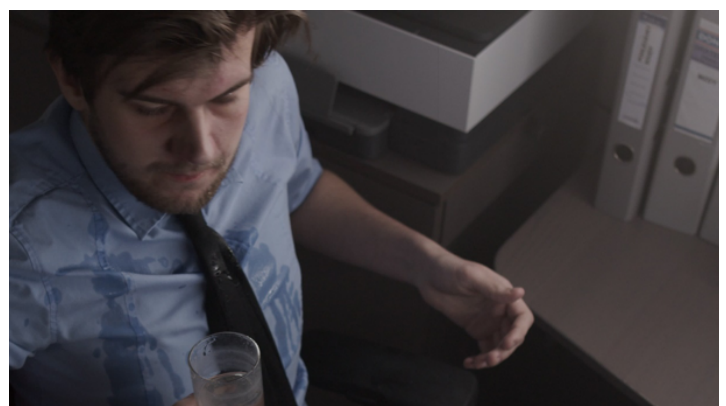
konvenční metody nejvíce času. V rámci automatického AI nástroje je vše hotovo v řádu několika minut, ale při použití hybridní metody, kdy jsem vše ladil v After Effects je tato retuš práce na půl hodiny.

Na přiložených obrázcích je možné vidět výřez záběru před a po retuši, spolu s ukázkou artefaktu.

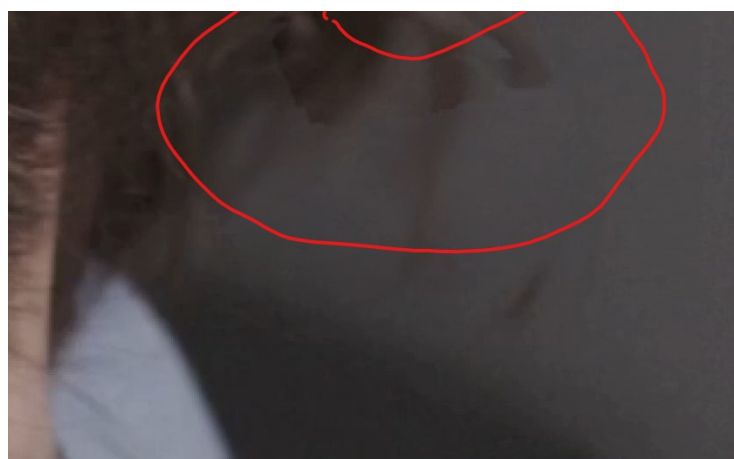
Konvenční retuš	AI retuš	Hybridní přístup
195 minut	3 minuty	30 minut



Obrázek 44 klíčový snímek před retuší v Runway



Obrázek 43 klíčový snímek po retuši v Runway



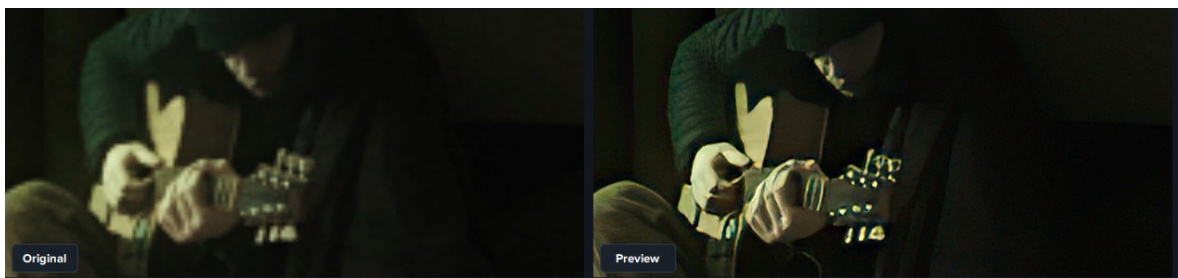
Obrázek 45 detail vzniklého artefaktu při využití Runway

V případě ostatních záběrů byly AI retuše hotové v řádu minut. Ze zkušeností vím, že podobný projekt by byl konvenčníma metodama práce na cca dva dny a AI mi umožnilo ho odbavit za večer.

#### 4.3.4 AI upscale obrazu

Původní render je ve full HD rozlišení, pro větší míru detailů jsem se rozhodl využít AI upscaling sekvence snímků.

Použitím AI Gigapixel jsem úspěšně dosáhl vylepšení obrazové kvality záběru. Tento algoritmus založený na umělé inteligenci umožnil zvětšit obraz v tomto případě z rozlišení Full HD na 4K, aniž by došlo k na první pohled viditelné ztrátě detailů nebo kvality. Díky tomuto procesu jsem značně optimalizoval pracovní postup, protože mi umožnil snížit celkový čas potřebný pro renderování o 50%. Tímto způsobem jsem zefektivnil produkci a dosáhl lepších výsledků bez zbytečného zatížení výpočetních zdrojů.



Obrázek 46 upscale pomocí AI gigapixel (2023, Martin Siuda)



Obrázek 47 upscalovaný snímek pomocí AI gigapixel (2023, Martin Siuda)

#### 4.3.5 Tvorba posteru – snaha kontrolovat výstup generativního AI Midjourney

V následujícím experimentu se budu snažit maximálně přiblížit ručně vytvořenému výsledku. Na tomto experimentu ilustruji nahodilost výstupů z generativní AI Midjourney.

V prvním případě jsem celou scénu vymodeloval, natexturoval, nasvítíl, vyrenderoval a zkomponoval ručně pomocí softwarů ve kterých běžně pracuji (Cinema4d, Redshift, Photoshop, After Effects) ve druhém případě jsem se pokusil zadání podobné scény splnit za pomoci generativní AI Midjourney 5. Nutno dodat, že pokud by byl cílovým dílem pouze obrázek, jakožto člověk bych AI nemohl z hlediska času v rámci této situace absolutně konkurovat. Midjourney trvalo vygenerovat výstup okolo deseti sekund. Mě scéna zabrala tři týdny práce. Vzhledem ale, že se jedná o snímek z filmu, zůstává mi stále výhoda v podobě přidání pohybů což AI zatím příliš neumí. Další věcí je, že i když jsem se pokusil Midjourney přesně popsat, jak má výstup vypadat, kontrola je omezená, a i když obrázky mají správnou atmosféru a jsou esteticky vizuálně vyvážené, nevypadají přesně tak, jak jsem chtěl. Ve třetím případě jsem použil svůj render jako image prompt, zde byl výsledek mnohem přesnější, ale ani tak se nepodařilo kontrolovat jednotlivosti jako pozici sloupů. Pro určení kompozice a barevné palety ale image prompt funguje v současnosti velmi dobře.



Obrázek 48 originální 3D render (2023, Martin Siuda)





Obrázek 50 poster vygenerovaný pouze na základě textového promptu pomocí Midjourney (2023, Midjourney)



Obrázek 49 poster vygenerovaný pomocí kombinace textového a image promptu za pomoci Midjourney (2023, Midjourney)

## ZÁVĚR

Jak může generalista využít AI nástroje pro svou práci? Jaké komplexní oblasti a nástroje může AI suplovat? A kde se odhalují limity této nové, explodující technologie?

Existuje sada AI nástrojů, které pomáhají překlenout mezeru, mezi myšlenkou a cílem. Sada nástrojů, které stírají technologické bariéry mezi tvůrcem a dílem.

Při svém bádání a praktických případových studiích, jsem došel k závěru, že v době psaní této práce je umělá inteligence pro generalistu využitelná hlavně jako nástroj poskytující dodatečná data pro další zpracování. V profesionálním užití AI je hlavním problémem zatím nedostatečná modularita. Tato modularita vyplývá z principu, kterým AI pracuje, a v tuto chvíli se pro běžného uživatele jedná o blackbox, do kterého není možné zasahovat a kontrolovat výchozí output.

Aktuální výzva pro vývojáře je vyřešit tuto kontrolu a integrovat nástroje do softwarů, využívaných ve VFX průmyslu a vyřešit workflow, která bude využívat kombinaci silných stránek obou přístupů. Kombinace konvenčních postupů s nahrazováním určitých kroků pomocí AI.

V současnosti, i s přihlédnutím na limitace AI velice silně rozšiřuje možnosti generalisty, a v mnoha ohledech nahrazuje technologické schopnosti a znalosti, kterými generalista oproti specialistům kvůli svému širokému záběru nemůže do hloubky disponovat.

Za dobu, kdy jsem psal tuto práci se odvětví dynamicky rozvíjelo a každý měsíc přinese nějakou novinku z oblasti umělé inteligence. Problémy, které její využití způsobuje jsou velice rychle řešeny. Vývojáři postupně integrují AI do konvenčních nástrojů. Jen měsíc po dopsání vydala společnost Runway modely Gen1 a Gen2, které dokážou generovat videa z holého textu, nebo s použitím referenčního videa. Adobe začalo publikovat ukázky svého vlastního modelu AI Firefly zakomponovaného do balíčku svých programů. Cinema 4D pracuje na pluginu umožňující upravovat rendery pomocí Stable diffusion...

S ohledem na překotný vývoj, kterého jsme svědky se dá očekávat, že aktuálnost této práce bude nejspíše postupně klesat, avšak nutno dodat, že i v době psaní práce mi byly nástroje založené na umělé inteligenci velkým přínosem.

Doufám, že si tyto nástroje osvojí co nejvíce lidí, což může přispět k obecné zvýšení úrovně a další demokratizaci celého odvětví.

## SEZNAM POUŽITÉ LITERATURY

- Chen-Hsuan Lin, J. G.-Y.-Y. (2022). Magic3D. *Magic3D: High-Resolution Text-to-3D Content Creation*, <https://research.nvidia.com/labs/dir/magic3d/>. NVIDIA Corporation.
- CorridorCrew. (Listopad 2022). *Youtube*. Načteno z [https://www.youtube.com/watch?v=YX5AoaWrowY&ab\\_channel=CorridorCrew](https://www.youtube.com/watch?v=YX5AoaWrowY&ab_channel=CorridorCrew)
- Corridordigital.com*. (23.. Leden 2023). Načteno z [https://www.youtube.com/watch?v=\\_9LX9HSQkWo&t=12s&ab\\_channel=CorridorCrew](https://www.youtube.com/watch?v=_9LX9HSQkWo&t=12s&ab_channel=CorridorCrew)
- Baker, S. S. (2011). A database and evaluation methodology for optical flow.
- Ben Mildenhall, P. P. (19. Březen 2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. San Diego, Kalifornie, USA.
- Brinkmann, R. (2008). *The Art and Science of Digital Compositing: Techniques for Visual Effects, Animation and Motion Graphics*. Morgan Kaufmann.
- Deacon, T. W. (1998). *The Symbolic Species: The Co-evolution of Language and the Brain*. W. W. Norton & Company.
- Gao, J. S. (2022). *GET3D: A Generative Model of High Quality 3D*. Načteno z <https://nvtlabs.github.io/GET3D/>
- https://80.lv/*. (2020). Načteno z Cascadeur Research: Physics In Animation: <https://80.lv/articles/cascadeur-research-physics-in-animation>
- Ian Goodfellow, Y. B. (2017). *Deep Learning*. MIT Press.
- Julianna Delua, S. I. (12. Březen 2021). *IBM.com*. Načteno z <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>
- Kampf, S. (2022). Basic scripting with ChatGPT. *Turning Ideas Into Animations With ChatGPT and After Effects*. [shorturl.at/kBV01](https://shorturl.at/kBV01).
- Katri, C. (nedatováno). *GitHub*. Načteno z DreamTextures: <https://github.com/carson-katri/dream-textures>
- Kozin, V. (23. Prosinec 2020). *Záznam online konference*. Načteno z [https://www.youtube.com/watch?v=miLIwQ7yPkA&ab\\_channel=EscapeStudios](https://www.youtube.com/watch?v=miLIwQ7yPkA&ab_channel=EscapeStudios)
- Leon A. Gatys, A. S. (2016). *Image Style Transfer Using Convolutional Neural Networks*. Načteno z [https://openaccess.thecvf.com/content\\_cvpr\\_2016/papers/Gatys\\_Image\\_Style\\_Transfer\\_CVPR\\_2016\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2016/papers/Gatys_Image_Style_Transfer_CVPR_2016_paper.pdf)
- McCorduck, P. (2004). *Machines who think: a personal inquiry into the history and prospects of artificial intelligence*. MA: A.K. Peters.
- Massobrio, A. (nedatováno). *Iterative Design Process: A Guide & The Role of Deep Learning*. Načteno z <https://www.neuralconcept.com/>: <https://www.neuralconcept.com/post/the-iterative-design-process-a-step-by-step-guide-the-role-of-deep-learning>
- Menache, A. (2010). *Understanding Motion Capture for Computer Animation and Video Games*. Morgan Kaufmann Publishers Inc.
- Stephens, J. (nedatováno). Getting Started with NVIDIA Instant NeRFs. <https://developer.nvidia.com/blog/getting-started-with-nvidia-instant-nerfs/>. Nvidia Developer.
- Tancik, M. (nedatováno). *NeRF research - matthewtancik*. Načteno z [matthewtancik.com](https://www.matthewtancik.com/): <https://www.matthewtancik.com/>
- Thalmann, D. &. (1990). *Computer Animation: Theory and Practice*. Springer.
- Um, T. T. (2020). Data-driven fluids: Wave-curve networks for water simulation. *ACM Transactions on Graphics (TOG)*, 39(4), 65.



*Why THIS is the Future of Imagery?* (nedatováno). (CorridorDigital) Načteno z [https://www.youtube.com/watch?v=YX5AoaWrowY&ab\\_channel=CorridorCrew](https://www.youtube.com/watch?v=YX5AoaWrowY&ab_channel=CorridorCrew)  
Yogesh, B. (nedatováno). *NvidiaCanva-ImageToImage*. Načteno z <https://research.nvidia.com/labs/dir/eDiff-I/>



## SEZNAM POUŽITÝCH SYMBOLŮ A ZKRATEK

AI Umělá inteligence

VFX Vizuální efekty

AAE Adobe After Effects

CGI Počítačem generovaná grafika

**SEZNAM OBRÁZKŮ**

Obrázek 1 Talós (2023, Midjourney).....	13
Obrázek 2 Symbolová komunikace .....	15
Obrázek 3 Efekt rozpixelování z filmu Westworld (1973, Michael Crichton).....	16
Obrázek 4 Schéma deep learning procesu .....	17
Obrázek 5 Srovnání schématu nodového a layered systému .....	17
Obrázek 6 Malování dokreslovačky na sklo ve filmu Star Wars VI (1983, George Lucas).....	18
Obrázek 7 Latentní prostor .....	20
Obrázek 8 Hierarchie AI pojmů.....	21
Obrázek 9 Fáze generování obrázku ze šumu. (2021, OpenAI).....	23
Obrázek 10 Příklad funkce Sharpen založené na konvolučním filtru.....	24
Obrázek 11 červená karkulka vygenerovaná pomocí Midjourney (2023, Midjourney).....	25
Obrázek 12 Dataset LegoBricks (2020, LegoBricks).....	26
Obrázek 13 Srovnání času postupů retuší fotografií – vlevo původní fotografie, uprostřed retuš v Photoshopu 2 hodiny, vpravo Dalle 2 retuš 10 minut. (2023, Veronika Bočková).....	29
Obrázek 14 Rotoscopování pomocí Mocha Pro (2019, Mocha pro) .....	31
Obrázek 15 Rotoscopování pomocí Runway (2022, Runway).....	32
Obrázek 16 Ukázka nedokonalosti výstupu Content Aware fill.....	33
Obrázek 17 původní ulice bez úprav (2023, Veronika Bočková).....	34
Obrázek 18 retuš pomocí AI nástroje Dalle2 (2023, Veronika Bočková).....	34
Obrázek 19 matte paint generovaný pomocí Dalle2.....	35
Obrázek 20 Původní technologie motion capture.....	37
Obrázek 21 Rokoko suit.....	37
Obrázek 22 Windows Kinect.....	37
Obrázek 23 Vytrénování AI pohybového modelu z RGB obrazového záznamu. ....	38
Obrázek 25 Substance Alchemist AI DeLighting.....	39
Obrázek 24 Dream texture addon pro Blender .....	39
Obrázek 26 Z-depth pass 3D render .....	40
Obrázek 27 Vytváření Z-depht mapy pro compositting pomocí masek v Adobe After Effects .....	41
Obrázek 28 Depht mapa generovaná pomocí AI Midas.....	41
Obrázek 29 Upscalování rozlišení pomocí AI Gigapixel .....	42
Obrázek 30 ukázka AI denoiseru (vlevo před, vpravo po) .....	44
Obrázek 31 srovnání skutečného herce, a 3D dubla z filmu Star Wars: Rogue One (2016, Gareth Edwards) .....	45
Obrázek 32 ukázka deep fake technologie – vložení tváře Nicolase Cage.....	47

---

Obrázek 33 mesh vygenerovaný za pomoci fotogrametrie.....	49
Obrázek 34 neural raidance field (NeFR).....	50
Obrázek 35 modely generované pomocí Luma AI.....	51
Obrázek 36 Expression funkce pro After Effects, vygenerovaná pomocí GPT 4 .....	53
Obrázek 37 původní render celé scény .....	59
Obrázek 38 Scéna s přidaným kouřem a motion blur.....	59
Obrázek 39 pass kouře s přidaným motion blur pomocí AI nástrojů .....	59
Obrázek 40 původní 3D render pass kouře.....	59
Obrázek 41 rotoscoping pomocí nástroje Mocha .....	61
Obrázek 42 rotoscoping pomocí nástroje Runway .....	61
Obrázek 43 klíčový snímek po retuši v Runway .....	63
Obrázek 44 klíčový snímek před retuší v Runway .....	63
Obrázek 45 detail vzniklého artefaktu při využití Runway .....	63
Obrázek 46 upscale pomocí AI gigapixel (2023, Martin Siuda) .....	64
Obrázek 47 upscalovaný snímek pomocí AI gigapixel (2023, Martin Siuda).....	64
Obrázek 48 originální 3D render (2023, Martin Siuda).....	65
Obrázek 49 poster vygenerovaný pomocí kombinace textového a image promptu za pomoci Midjourney.....	66
Obrázek 50 poster vygenerovaný pouze na základě textového promptu pomocí Midjourney .....	66



**SEZNAM TABULEK**

Tabulka 1 tabulka srovnání časů nutných pro realizaci renderu.....	60
Tabulka 2 srovnání časů nutných pro rotoscoping scény za využití rozdílných nástrojů....	61

