Umělá inteligence jako asistent režie: Koncepční rámec pro generování videoklipů na základě textů písní

Zpracoval: Tým oborových specialistů

Datum: 25. května 2025

Verze dokumentu: 1.0

Shrnutí

Tato zpráva předkládá podrobnou analýzu a koncepční návrh inovativní softwarové aplikace, která funguje jako inteligentní „asistent režie“ pro tvorbu videoklipů. Jádrem aplikace je její schopnost transformovat text písně na strukturovaný vizuální scénář. Systém přijímá jako vstup text písně, volitelná uživatelská média (obrázky, videoklipy) a požadovanou celkovou délku skladby. Následně provádí hloubkovou sémantickou a strukturální analýzu textu, na jejímž základě generuje očíslovanou sekvenci promtů (pokynů) pro tvorbu vizuálních materiálů v externích generativních platformách. Klíčovou inovací a jedinečnou hodnotou aplikace je její překladový engine, který převádí abstraktní lyrické koncepty – jako jsou emoce, témata a narativní struktura – do konkrétního filmového jazyka, včetně pokynů pro kameru, osvětlení a vizuální styl.
Aplikace je navržena tak, aby zapadla do pracovního postupu technicky zdatných tvůrců, jako jsou nezávislí hudebníci a tvůrci obsahu pro sociální média, kteří hledají efektivní způsob, jak vizualizovat svou hudbu. Místo toho, aby se snažila nahradit lidskou kreativitu, funguje jako kolaborativní partner, který automatizuje časově náročnou fázi pre-produkce a tvorby storyboardu. Tím, že poskytuje strukturovaný plán a sadu kreativních návrhů, umožňuje uživatelům soustředit se na finální umělecké doladění. Zpráva podrobně rozebírá technologický základ, včetně zpracování přirozeného jazyka (NLP), generativní umělé inteligence a počítačového vidění, a navrhuje modulární, na API založenou architekturu. Závěrem konstatuje, že tato aplikace má potenciál zaplnit významnou mezeru na trhu s kreativními nástroji AI tím, že nabízí jedinečný most mezi analýzou textu a vizuální tvorbou.

Sekce 1: Koncepční rámec: Asistent režie pro videoklipy s podporou AI

1.1 Úvod: Nové paradigma v kreativní spolupráci

Vstup umělé inteligence do kreativních odvětví ohlašuje posun od nástrojů pro pouhou automatizaci k sofistikovaným kolaborativním partnerům. Navrhovaná aplikace není koncipována jako náhrada lidské kreativity, ale jako inteligentní spolutvůrce, který rozšiřuje umělecké možnosti. Moderní AI nástroje se vyvíjejí z pouhých vykonavatelů úkolů na

spolutvůrce a asistenty režie, kteří pomáhají překonávat tvůrčí bloky, navrhují nové nápady a přebírají pracné, opakující se činnosti. Hlavním účelem této aplikace je překonat „vizuální tvůrčí blok“, který často postihuje hudebníky a tvůrce, a přeložit jejich zvukovou a lyrickou vizi do hmatatelného vizuálního scénáře, tedy storyboardu. Aplikace tak demokratizuje proces tvorby videoklipů a umožňuje umělcům realizovat vizuální složku jejich díla s dříve nedosažitelnou efektivitou a kreativní podporou.

1.2 Konvergence technologií

Navrhovaná aplikace se nachází na průsečíku tří klíčových domén umělé inteligence, jejichž nedávná vyspělost a synergie činí takový integrovaný nástroj poprvé technicky realizovatelným. Těmito doménami jsou:

1. **Zpracování přirozeného jazyka (NLP):** Tato technologie je základem pro dekonstrukci a pochopení textu písně. Moderní NLP modely, zejména architektury založené na transformerech, jako jsou BERT nebo GPT, umožňují hloubkovou sémantickou analýzu, rozpoznávání emocí a identifikaci klíčových témat v textu.
2. **Generativní umělá inteligence (Text-to-Image/Video):** Tato oblast se zabývá vytvářením vizuálního obsahu na základě textových pokynů (promtů). Vzestup platforem jako Midjourney, Sora od OpenAI nebo Veo od Googlu demonstruje schopnost generovat vysoce kvalitní a komplexní vizuální scény, což je klíčové pro realizaci navržených vizuálních konceptů.
3. **Počítačové vidění:** Tato disciplína umožňuje strojům „vidět“ a interpretovat vizuální data. V kontextu této aplikace se využívá k analýze a automatickému označování (tagování) médií nahraných uživatelem, což umožňuje jejich inteligentní zařazení do časové osy videoklipu na základě jejich obsahu a tematické relevance.

Právě souběžná zralost těchto tří pilířů, zejména nástup výkonných multimodálních modelů, jako je Gemini, které dokáží zpracovávat text i obraz současně, otevírá dveře pro vytvoření takto komplexního a integrovaného tvůrčího nástroje.

1.3 Definice uživatele: Technicky zdatný tvůrce

Cílovou skupinou pro tuto aplikaci je technicky zdatný kreativní profesionál. Může se jednat o nezávislého hudebníka, tvůrce obsahu pro sociální média (YouTube, TikTok, Instagram) nebo malé produkční studio, které potřebuje efektivně vytvářet vysoce kvalitní vizuální obsah. Tito uživatelé nehledají jednoduchou „černou skříňku“, která za ně udělá veškerou práci, ale spíše výkonný a přizpůsobitelný nástroj, který se integruje do jejich stávajícího pracovního postupu. Požadavek na snadno kopírovatelné prompty je toho jasným důkazem. Uživatel chce mít kontrolu a flexibilitu použít vygenerované nápady ve svých oblíbených generativních platformách. Aplikace tedy neslouží jako uzavřený ekosystém, ale jako centrální mozek pro vizuální ideaci, který spolupracuje s ostatními specializovanými nástroji v arzenálu moderního tvůrce.

Zásadní je, že primární hodnota aplikace nespočívá v samotném generování finálního videa, ale ve strukturované ideaci. Zatímco schopnost vygenerovat jeden videoklip se stává komoditou , skutečnou výzvou pro tvůrce je vytvořit koherentní sekvenci vizuálů, která vypráví příběh a odpovídá emocionálnímu oblouku písně. Uživatelův dotaz klade důraz na „očíslované prompty“ a „časovou osu“, což signalizuje, že hlavním problémem je koncepční práce na storyboardu. Nejhodnotnějším výstupem aplikace je tedy strukturovaná a načasovaná sekvence nápadů. Tím se nástroj redefinuje z pouhého „video editoru“ na „generátor vizuálního zpracování“, který

automatizuje náročnou pre-produkční fázi.

Sekce 2: Dekonstrukce narativu: Engine pro analýzu textu

2.1 Základní vrstva: Strukturální segmentace

Prvním a nejdůležitějším krokem je rozdělení textu písně na jeho základní stavební kameny, jako jsou sloky (Verse), refrény (Chorus), mosty (Bridge) a další. Tato struktura tvoří kostru, na které je postavena celá časová osa videoklipu.
Metody:

- **Analýza založená na opakování:** Základním principem je, že refrény se z definice opakují. Systém využije algoritmy k vytvoření tzv. samo-podobnostní matice (Self-Similarity Matrix, SSM). V této matici je každá řádka textu porovnána s každou další řádkou. Pokud se dvě řádky shodují, příslušná buňka v matici získá vysokou hodnotu (např. 1).
- **Vizualizace opakování:** Tuto matici lze vizualizovat jako tepelnou mapu (heatmap), kde diagonální linie jasně ukazují na opakující se sekvence (typicky refrény), zatímco souvislé bloky podobných barev značí homogenní sekce, jako jsou sloky.
- **Pokročilé techniky:** Pro dosažení vyšší přesnosti, zejména u textů, kde se refrény mírně liší, systém nasadí modely neuronových sítí. Modely pro sekvenční značkování (sequence labeling) nebo hierarchické sítě s mechanismem pozornosti (Hierarchical Attention Networks) mohou být natrénovány na anotovaných textech, aby se naučily rozpoznávat strukturu písní s větší robustností. Tímto způsobem se systém posouvá od prostého porovnávání řetězců k naučenému, kontextuálnímu chápání struktury písně. Pro tuto úlohu by mohly být adaptovány i knihovny původně určené pro analýzu audia, jako je msaf.

2.2 Analýza tematického a emocionálního jádra

Jakmile je struktura identifikována, každá sekce (sloka, refrén) je analyzována z hlediska svého významu a nálady.

- **Sentiment a emoce:** Systém využije specializované API nebo modely k přiřazení emocionálních značek. To může sahat od jednoduchého rozlišení na pozitivní/negativní sentiment až po granulárnější emoční stavy založené na zavedených psychologických modelech, jako je Ekmanův atlas emocí, který využívá služba LyricIQ ve spojení s IBM Watson. Služby jako Nyckel nabízejí předtrénované klasifikátory pro specifické lyrické sentimenty, jako jsou ‚melancholický', ‚nadějný' nebo ‚rozzlobený'.
- **Tematické značkování:** Kromě emocí systém identifikuje i zastřešující témata. Služby jako Cyanite.ai a Sonoteller.ai explicitně nabízejí funkci značkování „lyrických témat". Tímto způsobem lze identifikovat koncepty jako ‚láska', ‚sociální komentář', ‚oslava' nebo ‚příroda'. Toho je dosaženo pomocí pokročilých NLP modelů, jako jsou transformery (např. BERT, DistilBERT), které rozumí kontextu a sémantice slov.

2.3 Extrakce vizualizovatelných prvků: Rozpoznávání pojmenovaných

entit (NER)

Tento krok představuje nejpřímější most mezi textem a vizuální stránkou. Systém využije nejmodernější model pro rozpoznávání pojmenovaných entit (Named Entity Recognition, NER) k identifikaci a klasifikaci konkrétních, vizualizovatelných podstatných jmen.

- **Standardní NER:** Identifikuje osoby, organizace a místa (např. „Johnny“, „bar“, „Paříž“).
- **Jemnozrnné a vlastní NER:** Aplikace musí jít nad rámec standardních kategorií. Je nutné ji natrénovat nebo doladit na datasetech, jako je Few-NERD , aby dokázala rozpoznat širší škálu objektů, konceptů a dokonce i abstraktních vizuálních prvků (např. „rozbité zrcadlo“, „jedoucí vlak“, „blednoucí fotografie“). Analýza kreativních textů, jako jsou písně, je obzvláště náročná, protože standardní NER modely trénované na zpravodajských článcích mohou selhávat při interpretaci poetického a metaforického jazyka.
- **Multimodální NER:** Pokročilé implementace by mohly využívat i multimodální NER, který k lepšímu pochopení nejednoznačného textu využívá přidružené obrázky. Tento koncept, zkoumaný pro příspěvky na sociálních sítích, je pro tuto aplikaci vysoce relevantní.

Pro efektivní fungování je nezbytná hierarchická analýza. Nestačí provést plochou analýzu celé písně, protože její emocionální cesta se mění. Strukturální segmentace (2.1) poskytuje primární hierarchii (Sloka 1, Refrén 1, Sloka 2...). Tematická, emocionální a NER analýza (2.2 a 2.3) musí být aplikována na úrovni jednotlivých segmentů. To umožňuje systému pochopit, že sloka může být melancholická a zmiňovat „deštivé ulice“, zatímco refrén je nadějný a zmiňuje „slunce“. Tato hierarchická datová struktura (Píseň -> Segment -> Řádky -> Entity/Emoce) je základním datovým modelem, který musí být předán enginu pro generování promtů.

Zvláštní výzvou je tzv. „problém refrénu“. Text refrénu se opakuje, ale videoklip zřídka ukazuje pro každý refrén naprosto stejné záběry. Naivní systém by generoval identické prompty pro každý refrén, což by vedlo k monotónnímu videu. Inteligentní „asistent režie“ musí tento problém řešit. Měl by rozpoznat, že se jedná o refrén, a aplikovat specifickou logiku: jádro tématu a subjektu zůstane konzistentní, ale vizuální prvky jako úhel kamery, typ záběru nebo detaily scény se budou měnit. Například prompt pro první refrén může znít: „Široký záběr na ženu stojící na útesu, dívá se na oceán, osvětlení zlaté hodiny.“ Prompt pro druhý refrén by mohl být: „Střední detail téže ženy na útesu, vítr jí fouká do vlasů, vypadá odhodlaně, osvětlení zlaté hodiny.“ To demonstruje sofistikované chápání filmového jazyka, nikoli jen lyrického opakování.

Služba / API	Základní funkce	Podkladová technologie	Granularita analýzy	Silné stránky pro tuto aplikaci	Omezení pro tuto aplikaci
Sonoteller.ai	Analýza nálady, žánru, nástrojů, tempa, lyrických témat, struktury (refrén)	Proprietární AI	Úroveň písně a sekcí	Komplexní hudební a lyrická analýza v jednom, identifikace „zlaté minuty“.	Méně granulární než specializované NLP nástroje pro analýzu na úrovni řádků.
LyricIQ (LyricFind)	Analýza emocí (Ekmanův model), sentimentu, tematické filtry (31 kategorií)	IBM Watson	Úroveň písně	Velmi detailní emoční analýza, široká škála obsahových filtrů (násilí,	Zaměřeno primárně na obsahovou klasifikaci, méně na vizuální

Služba / API	Základní funkce	Podkladová technologie	Granularita analýzy	Silné stránky pro tuto aplikaci	Omezení pro tuto aplikaci
				náboženství atd.).	sémantiku.
Cyanite.ai	Značkování žánru, nálady, nástrojů, lyrických témat, vyhledávání podobnosti	Proprietární AI	Úroveň písně	Silné v identifikaci nálady a lyrických témat, což je klíčové pro vizuální překlad.	Analýza je spíše holistická, nemusí poskytovat entity na úrovni slov.
Nyckel API	Klasifikace sentimentu textu s předtrénovaný mi modely pro texty písní	Vlastní modely	Úroveň textového vstupu	Nabízí specifické a relevantní emoční kategorie (např. ‚melancholický', ‚nadějný').	Vyžaduje integraci jako samostatná komponenta; neposkytuje komplexní analýzu.
Google Cloud NLP API	Rozpoznávání entit (NER), analýza syntaxe, klasifikace obsahu	Google Transformer modely	Úroveň věty/textu	Velmi výkonné a přesné NER pro extrakci konkrétních vizualizovatelný ch objektů a míst.	Standardní modely mohou vyžadovat doladění pro kreativní a poetický jazyk.

Tabulka 1: Srovnávací analýza API pro analýzu textů písní. Tato tabulka poskytuje přehled pro výběr optimální kombinace nástrojů třetích stran pro sestavení enginu pro lyrickou analýzu, přičemž zvažuje rovnováhu mezi náklady, schopnostmi a snadností integrace.

Sekce 3: Od slov ke světům: Sémanticko-kinematografický překladový engine

3.1 Anatomie filmového promptu

Tato podsekce definuje cílový formát výstupu. Jednoduchý prompt jako „smutný muž" je pro tvorbu kvalitního vizuálu nedostatečný. Systém musí generovat bohaté a strukturované prompty, které obsahují filmový jazyk. Na základě osvědčených postupů pro pokročilé text-to-video modely, jako je Google Veo , a obecných příruček pro prompt engineering , bude každý prompt složen z následujících klíčových prvků:

- **Subjekt:** Kdo nebo co je v záběru. Odvozeno z NER (sekce 2.3) nebo tematické interpretace (např. „osamělý vlk").
- **Akce:** Co subjekt dělá. Odvozeno ze sloves a kontextu v textu.
- **Scéna/Prostředí:** Kde a kdy se děj odehrává. Odvozeno z NER nebo tematické asociace (např. téma ‚ztráta' -> ‚prázdný pokoj').
- **Jazyk kamery:** Typ záběru (detail, široký záběr), úhel (podhled, nadhled), pohyb (nájezd, pomalý švenk).

- **Osvětlení:** Nasvícení scény, které udává náladu (např. „drsné neonové světlo“, „měkké ranní světlo“).
- **Styl/Estetika:** Celkové vizuální pojetí (např. „fotorealistické, 4K“, „styl akvarelové malby“, „drsný styl grafického románu“).

3.2 Jádro logiky: Překladová matice

Toto je „tajná ingredience“ celé aplikace. Jedná se o systém založený na pravidlech a strojovém učení, který mapuje analytické výstupy ze sekce 2 na komponenty promptu z podsekce 3.1.
Příklad logického řetězce:

- **Vstup:** Segment je ‚Sloka‘, emoce je ‚Melancholická‘, téma je ‚Ztráta‘, NER našel entity ‚fotografie‘ a ‚déšť‘.
- **Překlad -> Styl:** ‚Melancholická‘ + ‚Ztráta‘ se mapuje na styl ‚Film Noir‘ nebo ‚Desaturovaná barevná paleta‘.
- **Překlad -> Osvětlení:** ‚Déšť‘ se mapuje na ‚low-key osvětlení‘, ‚odrazy na mokré dlažbě‘.
- **Překlad -> Kamera:** ‚Melancholická‘ se mapuje na ‚pomalý pohyb kamery‘, ‚statické záběry‘, ‚pomalý zoom‘.
- **Překlad -> Subjekt/Akce:** Kombinací NER se vytvoří scéna: „Detail na ruku držící vybledlou fotografii, v pozadí stékají kapky deště po okenní tabuli.“

Tato logika bude kodifikována v koncepční matici, která slouží jako základ pro vývoj (viz Tabulka 2).

3.3 Sekvencování promtů a narativní tok

Aplikace nebude generovat prompty izolovaně. Bude brát v úvahu předchozí a následující prompty, aby zajistila logický a plynulý vizuální tok. Využije k tomu základní filmové principy, jako je zahájení scény úvodním širokým záběrem (establishing shot) a následné přechody na střední záběry a detaily pro zdůraznění emocí a detailů. Přechody mezi lyrickými sekcemi (např. ze sloky do refrénu) spustí specifické typy vizuálních přechodů, jako je rychlejší střih nebo změna barevného tónování.
Aby systém nevytvářel klišé (např. ‚smutek‘ se vždy rovná ‚déšť‘), musí disponovat rozsáhlým a hlubokým „vizuálním slovníkem“. Tento slovník není jen prostým mapováním slova na obrázek, ale spíše nálady -> filmového jazyka. Tato znalostní báze, obsahující tisíce filmových technik, uměleckých stylů a vizuálních metafor, je klíčovým aktivem aplikace. Může být kurátorována filmovými experty nebo získána trénováním modelu na datasetech filmových scénářů a jejich filmových zpracování. Kvalita a nuance tohoto interního slovníku přímo určují kreativní rozsah a kvalitu výstupů.
Zatímco automatizovaný překlad je hlavní funkcí, pokročilí uživatelé budou chtít tento proces ovlivnit. Aplikace by měla umožnit zadání globálních „stylových pokynů“ na začátku, například „Vytvoř to v estetice synthwave 80. let“ nebo „Režijní styl: Wes Anderson“. Tyto vysokoúrovňové instrukce by omezily a usměrnily volby, které překladový engine provádí. Pokyn „Wes Anderson“ by například upřednostnil symetrické kompozice, specifické barevné palety a statické výrazy postav. Tím se vytváří víceúrovňová uživatelská zkušenost: jednoduchý, automatizovaný režim pro začátečníky a výkonný, „režírovatelný“ režim pro profesionály, což dramaticky zvyšuje tržní atraktivitu aplikace.

Lyrické téma / Emoce	Asociované vizuální metafory	Doporučený umělecký styl	Dominantní barevná paleta	Typický jazyk kamery	Příklad fragmentu promptu
Nadějný vzdor	Fénix z popela, květina prorážející beton, východ slunce po bouři	Vysoce kontrastní realismus, epická fantasy ilustrace	Teplé zlaté tóny, jasná bílá, hluboké kontrastní stíny	Podhled (low-angle shot) směřující vzhůru, pomalý nájezd (push-in), stoupající jeřábový záběr	„...jediná zářivá květina proráží popraskaný beton, podhled, filmové, zalité teplým světlem úsvitu...“
Melancholická nostalgie	Staré fotografie, prachové částice tančící ve světle, opuštěné hřiště	Sépiový tón, měkké zaostření (soft focus), styl starého filmu (vintage film grain)	Tlumené, teplé barvy (sépie, jantarová), nízký kontrast	Statický záběr, pomalý zoom na detail, záběr přes rameno na prázdnou scénu	„...stará sépiová fotografie leží na zaprášeném stole, paprsky světla odhalují tančící prach, pomalý nájezd na fotografii, melancholické...“
Euforická svoboda	Let ptáka, tanec v dešti, jízda po otevřené silnici s větrem ve vlasech	Jasné, živé barvy, styl GoPro/FPV dronu, dynamická animace	Syté primární barvy, vysoký jas, sluneční odlesky (lens flare)	Rychlý sled záběrů (fast cuts), ruční kamera, 360stupňový rotační záběr	„...z pohledu první osoby (FPV drone shot) letící nad pobřežní silnicí při západu slunce, pocit svobody, teplé barvy, dynamický pohyb...“
Paranoidní úzkost	Dlouhé stíny, sledování z dálky, zkreslené odrazy, blikající světla	Film Noir, německý expresionismus, found footage (nalezený záznam)	Monochromatická s vysokým kontrastem, studené modré a zelené tóny, ostré světlo	Nakloněná kamera (dutch angle), rychlé zoomy, trhané pohyby, záběry zpoza rohu	„...muž rychle kráčí prázdnou ulicí v noci, dlouhé stíny ho pronásledují, nakloněná kamera, drsné osvětlení z pouliční lampy, styl film noir...“

Tabulka 2: Koncepční příklad překladové matice z lyrických témat na vizuální prompty. Tato tabulka ilustruje, jak aplikace převádí abstraktní koncepty na konkrétní filmové instrukce a slouží jako základní blueprint pro vývoj jejího klíčového inteligentního jádra.

Sekce 4: Montážní linka: Interaktivní časová osa a integrace médií

4.1 Uživatelské rozhraní: Video editor založený na storyboardu

Centrální uživatelské rozhraní nebude připomínat tradiční, komplexní video editor (jako Adobe Premiere). Místo toho bude založeno na konceptu storyboardu, který je pro tuto úlohu mnohem intuitivnější a přístupnější. Tento přístup je inspirován nástroji jako Pictory a Wisecut , které používají scénář nebo storyboard jako hlavní editační rozhraní.
Rozložení: Obrazovka bude zobrazovat časovou osu s jasně vyznačenou strukturou písně (Sloka 1, Refrén 1 atd.). Pod každou sekcí bude umístěna sekvence karet. Každá karta reprezentuje jeden záběr a obsahuje:

1. Odpovídající řádek (řádky) textu písně.
2. Očíslovaný, AI-generovaný prompt.
3. Tlačítko „Kopírovat prompt“ (klíčový požadavek uživatele).
4. Zástupný prostor (placeholder), kam může uživatel nahrát vygenerovaný obrázek nebo video.

Tento design musí plynule propojovat proces generování založeného na promptech s editací na časové ose. Kreativní proces je iterativní: uživatel vygeneruje klip, prohlédne si ho a bude chtít upravit prompt nebo klip nahradit. Karta na storyboardu je pro tento cyklus ideálním kontejnerem, protože obsahuje prompt, výsledek a ovládací prvky pro opětovné generování nebo nahrazení. Tím se liší od nástrojů, které jsou buď pouze pro generování (jako Discord bot), nebo pouze pro editaci. Síla navrhované aplikace spočívá v jejím integrovaném pracovním postupu „vygeneruj a umísti“, což představuje významnou inovaci v uživatelském zážitku.

4.2 Inteligentní integrace médií nahraných uživatelem

Tato funkce řeší část uživatelského dotazu „mohu vložit média“. Když uživatel nahraje vlastní videoklipy nebo obrázky, systém je musí inteligentně zařadit.

- **Analýza pomocí počítačového vidění:** Aplikace využije výkonné API pro počítačové vidění, jako je Google Cloud Vision nebo Azure AI Vision , k analýze nahraných souborů. Toto API vygeneruje popisné značky (tagy), například „pláž“, „západ slunce“, „dav lidí“, „usmívající se tvář“. Pokročilejší modely jako Gemini Pro Vision mohou poskytnout i komplexnější popisy scén.
- **Tematické párování:** Systém následně provede sémantické vyhledávání. Porovná značky z uživatelských médií s lyrickými tématy, emocemi a entitami (NER) z každého segmentu písně. Například uživatelem nahraný klip označený jako „pláž, západ slunce“ bude automaticky navržen pro lyrickou sekci s tématem ‚romantika‘ nebo se slovy ‚oceán‘ a ‚večer‘. Stejný princip, jaký používá Cyanite.ai pro vyhledávání podobnosti mezi audio soubory, se zde aplikuje na párování vizuálního obsahu s textem.

4.3 Sestavení finálního střihu a export

Jakmile uživatel vygeneruje nebo nahraje vizuály pro každou kartu na storyboardu, aplikace nabídne nástroje pro jejich sestavení.

- **Automatizovaná montáž:** Jedním kliknutím může systém vyrenderovat pracovní verzi

videa, přičemž načasuje každý vizuál na dobu trvání odpovídající lyrické sekce. Může také automaticky přidat jednoduché přechody, podobně jako funkce „smart transitions“ v nástroji Descript.
- **Načasování a tempo:** Uživatel může na storyboardu ručně upravit délku trvání každého záběru. Aplikace bude průběžně počítat celkovou délku a poskytovat zpětnou vazbu, aby pomohla uživateli dosáhnout požadované délky písně.
- **Export:** Finální výstup není omezen pouze na video soubor. Systém by měl nabízet export v několika formátech:
    1. Finální video ve formátu MP4.
    2. Seznam střihových rozhodnutí (Edit Decision List, EDL) nebo soubor XML pro import do profesionálních střihových programů jako DaVinci Resolve nebo Final Cut Pro. To umožňuje další profesionální úpravy a je klíčovou funkcí pro pokročilé uživatele.

Funkce „Kopírovat prompt“ je více než jen tlačítko; je to filozofie otevřenosti. Uživatel explicitně požádal o tuto funkci, nikoli o to, aby aplikace měla vlastní, proprietární generátor videa. To naznačuje touhu po flexibilitě. Uživatel chce využívat nejlepší dostupné modely pro generování obrazu a videa (Midjourney, Sora, Veo atd.), které se rychle vyvíjejí. Tím, že aplikace externalizuje krok generování, se vyhýbá závislosti na jediném, potenciálně zastaralém generativním modelu. Její hodnota tak zůstává v inteligenci tvorby promtů, nikoli v renderování médií. To činí aplikaci odolnější vůči budoucím změnám a přátelštější k tvůrcům, kteří mají své preferované nástroje. Stává se tak centrálním „mozkem“, který orchestruje ostatní specializované nástroje.

Sekce 5: Technická architektura a strategické postavení

5.1 Modulární, na API založená systémová architektura

Tato podsekce navrhuje vysokoúrovňový technický blueprint. Systém by měl být postaven jako soubor mikroslužeb nebo modulů propojených přes API, nikoli jako monolitická aplikace. Tento přístup zajišťuje flexibilitu, škálovatelnost a snadnější údržbu.

- **Modul 1: Vstup a lyrická analýza:** Přijímá text, audio soubor písně. Využívá kombinaci open-source knihoven (např. pro výpočet SSM) a komerčních API (např. Google NLP pro NER, Nyckel pro sentiment).
- **Modul 2: Sémanticko-kinematografický engine:** Jádro proprietární logiky. Může být implementován jako velký jazykový model (např. GPT-4) doladěný na datech z „Překladové matice“ (Tabulka 2) a na principech filmové tvorby. Přijímá strukturovaná data z Modulu 1 a na výstupu poskytuje sekvenci strukturovaných promtů.
- **Modul 3: Analýza uživatelských médií:** Rozhraní napojené na API pro počítačové vidění (např. Azure AI Vision, Google Vision API) pro značkování nahraných souborů.
- **Modul 4: Frontend / Uživatelské rozhraní:** Webová aplikace postavená na moderním frameworku (např. React, Vue), která implementuje editor založený na storyboardu.
- **Modul 5: Montáž a export:** Využívá backendovou knihovnu pro zpracování videa (jako FFMPEG) nebo cloudovou službu pro sestavení finálního videa a generování exportních souborů (MP4, EDL).

5.2 Konkurenční prostředí a jedinečná hodnotová propozice (UVP)

Analýza trhu ukazuje, že navrhovaná aplikace zaujímá unikátní pozici.

- **Analýza konkurence:**
    - **Hudební nástroje:** Suno, Udio , Staccato , Lyrics Into Song – Tyto nástroje se zaměřují na *generování hudby*, nikoli videa. Jsou doplňkové, nikoli přímými konkurenty.
    - **Video nástroje:** Pictory , Vizard , Descript , Captions.ai – Jedná se o nástroje pro převod scénáře na video nebo o editory. Chybí jim hloubková lyrická analýza a tematické porozumění potřebné pro tvorbu videoklipu. Jejich vstupem je typicky próza, nikoli poezie/texty písní.
    - **Nástroje pro tvorbu lyric videí:** HeyGen – Jedná se o jednodušší nástroje zaměřené na synchronizaci textových animací se zvukem, nikoli na generování tematických vizuálů.
- **Jedinečná hodnotová propozice (UVP):** Unikátní pozice navrhované aplikace spočívá v její roli **inteligentního mostu** mezi těmito dvěma světy. Nevytváří hudbu a (nutně) nerenderuje video. Vykonává vysoce hodnotnou kognitivní úlohu **vizuálního překladu a tvorby storyboardu**, což je nika, která je v současné době na trhu neobsazená.

Obchodní model by měl být založen na předplatném (SaaS) s několika úrovněmi přístupu. Například: **Free/Hobby úroveň** s omezeným počtem projektů a vodoznakem; **Creator úroveň** s více projekty, pokročilou analýzou a bez vodoznaku; a **Pro/Studio úroveň** s neomezenými projekty, nejpokročilejšími možnostmi ovládání stylu, funkcemi pro týmovou spolupráci a profesionálními exportními formáty (EDL/XML).

Obranný příkop aplikace, tedy její klíčová konkurenční výhoda, nespočívá v použitých základních AI modelech, které jsou do značné míry komoditou. Skutečnou, těžko replikovatelnou hodnotou je proprietární databáze kreativních mapování – „Vizuální slovník“ a logika „Překladové matice“, které pohánějí sémanticko-kinematografický engine. Tato znalostní báze, budovaná kombinací expertní lidské kurace a strojového učení, by byla pro konkurenci velmi obtížné a časově náročné napodobit.

5.3 Budoucí směřování a pokročilé funkce

Aplikace má značný potenciál pro další rozvoj a rozšíření svých schopností:

- **Přímá integrace API:** Místo tlačítka „Kopírovat prompt“ nabídnout přímou integraci s hlavními generativními platformami (např. tlačítko „Generovat s Midjourney“, které automaticky odešle prompt).
- **Automatická synchronizace rtů (Lip-Sync):** Pro videa, kde vystupuje zpěvák, integrovat technologii pro synchronizaci pohybu rtů , která by oživila vygenerovaného avatara nebo dokonce statickou fotografii uživatele.
- **Dynamické generování doplňkových záběrů (B-Roll):** Systém by mohl identifikovat mezery ve vizuálním vyprávění a automaticky generovat relevantní doplňkové záběry na základě okolních témat.
- **Audio-reaktivní vizuály:** Analyzovat nejen text, ale i samotnou audio stopu na BPM, energii a instrumentaci (pomocí nástrojů jako Sonoteller nebo Tunebat). Tato data by mohla ovlivnit tempo střihu a pohyb ve vygenerovaných videích. Například sekce s výraznými bicími by mohla spustit rychlejší střihy.
- **Kompletní tvůrčí cyklus:** Integrovat se s AI generátory hudby. Uživatel by mohl zadat

pouze „smutná pop-punková píseň o rozchodu“ a systém by nejprve vygeneroval text a hudbu (pomocí nástroje jako Suno nebo Udio) a poté by je přímo vložil do vlastního procesu generování videa, čímž by vytvořil kompletní audiovizuální dílo z jediné věty.

Závěr

Navrhovaná aplikace představuje významný krok vpřed v oblasti kreativních nástrojů poháněných umělou inteligencí. Její hodnota nespočívá v nahrazení lidského umělce, ale v poskytnutí výkonného nástroje pro spolupráci, který zásadně zefektivňuje a demokratizuje proces tvorby videoklipů. Tím, že se zaměřuje na kognitivně náročnou úlohu překladu lyrického obsahu do vizuálního jazyka, zaplňuje zjevnou mezeru na trhu mezi AI generátory hudby a obecnými video editory.
Její modulární architektura a filozofie otevřenosti, symbolizovaná funkcí kopírování promtů, zajišťují flexibilitu a relevanci v rychle se měnícím technologickém prostředí. Klíčem k dlouhodobému úspěchu a obranyschopnosti na trhu je budování a neustálé zdokonalování proprietárního sémanticko-kinematografického enginu – znalostní báze, která je srdcem její kreativní inteligence. S jasně definovanou cílovou skupinou, silnou jedinečnou hodnotovou propozicí a bohatými možnostmi budoucího rozvoje má tento koncept potenciál stát se nepostradatelným nástrojem v arzenálu moderního digitálního tvůrce.

Citované zdroje

1. (PDF) LANGUAGE, LEARNING, AND LYRICS: NLP APPLICATIONS ..., https://www.researchgate.net/publication/391060127_LANGUAGE_LEARNING_AND_LYRICS_NLP_APPLICATIONS_IN_SONGWRITING_AND_LYRICISM 2. Staccato | AI Tools for Music Makers | MIDI Music, https://staccato.ai/ 3. Implementation of Classification of Poetry Text into the Emotional States Using Deep Learning Technique - ijmrset, https://www.ijmrset.com/upload/42_Implementation.pdf 4. Exploring Genre and Success Classification through Song Lyrics using DistilBERT: A Fun NLP Venture - arXiv, https://arxiv.org/html/2407.21068v1 5. text-to-video - The Prompt Engineering Institute, https://promptengineering.org/tag/text-to-video/ 6. The Easiest (One-prompt!) AI Music Videos with Pro lip-sync - YouTube, https://www.youtube.com/watch?v=mgVBR5oCBAU 7. Cloud Vision API documentation, https://cloud.google.com/vision/docs 8. Azure AI Vision with OCR and AI | Microsoft Azure, https://azure.microsoft.com/en-us/products/ai-services/ai-vision 9. Image understanding | Gemini API | Google AI for Developers, https://ai.google.dev/gemini-api/docs/image-understanding 10. Script To Video Creation In Minutes with Pictory, https://pictory.ai/pictory-features/script-to-video 11. Making AI-Enhanced Videos: Analyzing Generative AI Use Cases in YouTube Content Creation - arXiv, https://arxiv.org/html/2503.03134v1 12. Descript: Edit Videos & Podcasts Like a Doc | AI Video Editor, https://www.descript.com/ 13. I tried 100 AI Music Tools… These are the ONLY ones worth using - YouTube, https://www.youtube.com/watch?v=1oj0Usyy_ds 14. (PDF) NLP based Poetry Analysis and Generation - ResearchGate, https://www.researchgate.net/publication/313874773_NLP_based_Poetry_Analysis_and_Generation 15. Lyrics Segmentation: Textual Macrostructure Detection using Convolutions - ACL Anthology, https://aclanthology.org/C18-1174.pdf 16. A CHORUS-SECTION DETECTION METHOD FOR LYRICS TEXT - ISMIR, https://archives.ismir.net/ismir2020/paper/000088.pdf 17. Visualize repetition in song lyrics - The DO Loop - SAS Blogs,

https://blogs.sas.com/content/iml/2018/03/14/visualize-repetition-lyrics.html 18. A Method to Detect Chorus Sections in Lyrics Text - ResearchGate, https://www.researchgate.net/publication/373596878_A_Method_to_Detect_Chorus_Sections_in_Lyrics_Text 19. TuringTrain/lyrics_segmentation - GitHub, https://github.com/TuringTrain/lyrics_segmentation 20. rupakvignesh/Lyrics-to-Audio-Alignment - GitHub, https://github.com/rupakvignesh/Lyrics-to-Audio-Alignment 21. AmishaMurarka/Sentiment-Analysis-of-Song-Lyrics - GitHub, https://github.com/AmishaMurarka/Sentiment-Analysis-of-Song-Lyrics 22. LyricIQ - Home, https://lyriciq.lyricfind.com/ 23. Identify song lyrics sentiment using AI - Nyckel, https://www.nyckel.com/pretrained-classifiers/song-lyrics-sentiment/ 24. Cyanite.ai | AI-powered Music Tagging and Similarity Search, https://cyanite.ai/ 25. SONOTELLER.AI - AI song analyzer including lyrics and music, https://sonoteller.ai/ 26. Named Entity Recognition (NER) - Papers With Code, https://paperswithcode.com/task/named-entity-recognition-ner 27. A New State of the Art for Named Entity Recognition - PrimerAI, https://primer.ai/developer/a-new-state-of-the-art-for-named-entity-recognition/ 28. Named entity recognition - NLP-progress, http://nlpprogress.com/english/named_entity_recognition.html 29. What do we really know about State of the Art NER? - ACL Anthology, https://aclanthology.org/2022.lrec-1.643/ 30. Multimodal Named Entity Recognition for Short Social Media Posts - ACL Anthology, https://aclanthology.org/N18-1078/ 31. Veo video generation overview | Generative AI on Vertex AI - Google Cloud, https://cloud.google.com/vertex-ai/generative-ai/docs/video/overview 32. Veo: A Comprehensive Guide to Usage, Features, and Effective ..., https://daily.promptperfect.xyz/p/veo-prompt-guide 33. Prompt Engineering for AI Guide | Google Cloud, https://cloud.google.com/discover/what-is-prompt-engineering 34. Generative AI: Prompt Engineering Basics by IBM - Coursera, https://www.coursera.org/learn/generative-ai-prompt-engineering-for-everyone 35. How to write effective text prompts to generate AI videos? - FlexClip Help Center, https://help.flexclip.com/en/articles/10326783-how-to-write-effective-text-prompts-to-generate-ai-videos 36. Write your Veo 2 prompts here and i will generate them (First 30-40 comments with prompts i will generate) : r/singularity - Reddit, https://www.reddit.com/r/singularity/comments/1k7d7xp/write_your_veo_2_prompts_here_and_i_will_generate/ 37. Text generation and prompting - OpenAI API, https://platform.openai.com/docs/guides/text 38. Wisecut | Transform Your Videos with AI Editing, https://www.wisecut.ai/ 39. Vision AI: Image and visual AI tools | Google Cloud, https://cloud.google.com/vision 40. Image Tagging - Azure AI services - Learn Microsoft, https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/concept-tagging-images 41. Lyrics Into Song AI - Turn Lyrics To Song Free Online, https://lyricsintosong.ai/ 42. AI Video Editor: Edit Your Clips for Free - Vizard.ai, https://vizard.ai/tools/ai-video-editor 43. AI Video Editor - Easily Create & Edit Videos with AI - Captions, https://www.captions.ai/tools/ai-video-editor 44. Make a Lyric Video Fast with AI Tools - HeyGen, https://www.heygen.com/blog/make-a-lyric-video 45. " The Ultimate Guide to Creating Viral AI Music Videos (Step-by-Step Tutorial)" - Full Transcript Inside! | YTScribe, https://ytscribe.com/v/XJ5WbMOm7uc 46. 4 Best AI Tools for Music Analysis (+1 Bonus) in 2025 - - Beatoven.ai, https://www.beatoven.ai/blog/best-ai-tools-for-music-analysis/