



UMEÅ UNIVERSITET

"What a journey"

*Ett autoetnografiskt
utforskande av att
använda AI-stödda
verktyg under en
gestaltande process*

Henrik Bengtsson

Sammanfattning

I dagens samhälle talas det dagligen i media om AI, både som hot och möjlighet, och det kommer nya verktyg med mer och mer AI-stöd hela tiden. Men hur är det att faktiskt använda dessa i en gestaltande process? Vilka frågor och problem möter jag som kreatör när jag vill använda mig av AI-stödda verktyg? Kan jag lita på resultatet som jag får från verktygen? Kan jag lita på leverantörerna och de resultat som produceras. Är resultaten representativa, inkluderande och faktabaserade eller behöver jag kritiskt dem? Genom att ha använt mig av AI-stödda verktyg under en gestaltande process, där jag skapat en berättelse som har visualiserats och sedan tonsatts, så har jag undersökt hur dessa verktyg har påverkat mitt gestaltande och samtidigt undersökt vilka risker och problem som kan vara associerade med användandet av dessa AI-stödda verktyg. Vi kommer att se att även om det snabbt går att nå resultat, så är det flera frågor vi behöver ta ställning till under vår gestaltande process.

Abstract

In today's society, there is daily talk in the media about AI, both as a threat and an opportunity, and new tools with more and more AI support are being released all the time. But what is it like to actually use these in a creative process? What questions and problems do I face as a creator when I want to use AI-supported tools? Can I trust the results I get from the tools? Can I trust the suppliers and the results produced. Are the results representative, inclusive and fact-based or do I critically need them? By using AI-supported tools during a design process, where I created a story that was visualized and then set to music, I have explored how these tools have affected my design and at the same time explored what risks and problems may be associated with the use of these AI-supported tools. We will see that even if it is possible to achieve results quickly, there are several questions we need to decide on during our creative process.

Innehåll

Inledning	1
Syfte och frågeställningar	2
Gestaltande arbete	2
Bakgrund.....	3
Designprocesser	3
Gemensamma delar för designprocesser	3
Filmproduktion	4
En kort historia om AI	6
AI-modeller.....	6
Large Language Models (LLM)	7
GPT Modellen.....	7
Generativa bildmodeller.....	8
AI-verktyg inom musik.....	9
Ljudande AI-hjälpmedel – instrument	9
Ljudande AI-hjälpmedel – sång.....	9
Generativa musikaliska modeller.....	10
Problematisks beteenden hos AI modeller	10
Hallucinationer.....	10
Prompt-manipulation/injicering	11
Transparens	11
Datainterferens och kvalité	11
Representation inom AI modeller och deras resultat.....	12
Teori.....	12
Arts-based research	12
Metod	13
Autoetnografisk metod.....	13

Planerat genomförande	14
Designprocessen i detta arbete.....	14
Traditionella metoder och verktyg.....	15
Val av AI-stödda verktyg.....	15
Generativt textverktyg: ChatGPT och Bing CoPilot	15
Generativt bildverktyg: Midjourney	15
Musikaliska verktyg: Noteperformer.....	16
Val av språk	17
Genomförande.....	17
Fas 1: Berättelse och Manusförfattande.....	17
Första försök – Draken och flickan.....	17
Ändring av stil.....	21
Andra försöket – science fiction	21
Lägga till en twist i berättelsen	24
Revidering och redigering av manus	27
Fas 2: Visualisering och Produktions-design.....	27
Visualisering av viktiga händelser	28
Att skriva detaljerade prompter	29
Skapande av ögonblicksbilder och första tidslinjen.....	34
Den iterativa visualiseringen.....	34
Nya versioner under arbetets gång (ChatGPT 4 & Midjourney v 5).....	35
Slutgiltig klippning till klickspår	37
Fas 3: Tonsättande	37
Viktiga händelser i berättelsen.....	37
Musikaliska element	37
Att hitta information om Maya-musik	38
Orkestrering för Mayamusik.....	41

Skalor och tonal harmoni	43
Minimalism	43
Komponerande	43
Resultat	46
Diskussion.....	46
Trovärdighet hos verktygen	46
Juridiska frågeställningar	47
Användarvänlighet	48
Representation och transparens.....	48
Tonsättningen.....	49
Slutsatser	49
Framtida forskning.....	50
Referenser	52
Figurer.....	54
Appendix 1 – Manus	56
Appendix 2 – Storyboard	58
Appendix 3 – Notbild.....	59

Inledning

Hur jag använder verktyg för att stödja mitt gestaltande, och vilken påverkan dessa verktyg kan ha på själva processen, har alltid fascinerat mig. Med verktyg menar jag här hjälpmedel som exempelvis programvaror för att skapa noter och musik, virtuella instrument som simulerar en orkester, programvaror för att generera eller underlätta skapandet av bildmaterial, text och ordbehandlare och liknande. Att verktygen blir en viktig del i processen och att nya typer av verktyg kan förändra skapandet, samt leda till nya former och stilar av skapande är något som jag är mycket intresserad av.

Ända sedan människan gick från att doppa fingrarna i färgpulver till att binda ihop en bunt gräs och sen djurhår till en pinne, så har vårt användande av verktyg påverkat vårt kreativa gestaltande skapande (Robertson, et al., 2021, p. 36). I detta arbete skall jag studera och analysera hur användandet av verktyg med inbyggt stöd utav artificiell intelligens (AI), härnå kallat AI-stödda verktyg, kan förändra en gestaltande process.

Trots att AI har pratats om ända sedan Alan Turing publicerade sin artikel; "Computing Machinery and Intelligence" (Turing, 1950) så är det först under det senaste decenniet som området har utvecklats med en nästan dramatisk hastighet. Nya verktyg och modeller för AI kommer ut nästan dagligen och har potential att drastiskt förändra den gestaltande processen hos kreatörer (Exploring the Impact of Artificial Intelligence on the Creative Industry: Opportunities and Challenges, 2023).

I slutet av år 2022 och i början av år 2023 så skedde det ett antal stora utvecklingar inom just AI-stödda verktyg med mer avancerade modeller och implementeringar av dessa.

Allmänheten fick möjlighet att börja testa AI-stödda verktyg för bland annat språkanalys/generering och bildgenerering, som tidigare varit tillgängliga endast för forskare inom området. Verktygen blev initialt väldigt populära och fick en stor användarbas, men snart började etiska, sociala, kulturella, inkluderande och juridiska frågeställningar att uppstå. Kan jag lita på svaren jag får på frågorna jag ställer? Svaren låter trovärdiga och de ser trovärdiga ut, men stämmer de? De bilder som den AI styrda bildgenereringen tar fram ser väldigt välgjorda ut, men hur har mina instruktioner tolkats? Vad finns det för viktning och vinkling på det jag har skrivit och hur kan jag som kreatör styra dessa?

Genom att genomföra en gestaltande process, från idé, till visualisering och till slutgiltig artefakt, så kommer jag att undersöka och reflektera över hur just AI-stödda verktyg kan

förändra min process. Fokus kommer att vara på hur verktygen kan stödja mitt arbete, samt vilka etiska, sociala, kulturella, inkluderande och juridiska aspekter som jag behöver ta hänsyn till. I min egenskap av tonsättare har jag valt att göra den slutgiltiga artefakten i form av en tonsättning av en berättelse. Under framtagandet av denna berättelse kommer jag dock att beröra flera olika discipliner inom gestaltning såsom prosa, illustrationer, kinematografi och för att assistera mig i dessa delar så skall jag använda mig av AI-stödda verktyg.

Jag kommer inte att gå in djupt på den matematiska eller tekniska aspekten av de olika AI-modeller och tekniker som verktygen använder sig av, utan endast ge en översikt så att läsaren kan förstå lekmannamässigt hur de fungerar. Då kan läsaren också bedöma verktygets lämplighet och om det kan finnas några aspekter av användandet som läsaren behöver ta särskild hänsyn till.

Syfte och frågeställningar

Uppsatsens syfte är att utforska hur användandet av AI-stödda verktyg påverkar och förändrar min gestaltande process vid framställandet av en konstnärlig artefakt. För att besvara syftet formulerade jag följande forskningsfrågor:

1. Hur kan jag använda mig av AI-stödda verktyg i en gestaltande designprocess för att antingen ersätta eller komplettera traditionella verktyg och metoder?
2. Vilka möjligheter och utmaningar medför användandet av AI-stödda verktyg i en gestaltande designprocess?

Gestaltande arbete

För att besvara syftet och frågeställningarna så skall jag genomföra en gestaltande process. Denna process kommer att börja med framtagandet av en berättelse, som sedan visualiseras. Från denna visualisering skall jag skapa ett animerat bildspel (storyboard) och slutligen tonsätta detta bildspel.

Slutartefakten blir ett musikaliskt verk som är min gestaltning inspirerad av det visuella och textbaserade materialet. Artefakten kommer att bestå av ett partitur och en ljudande version av partituret samt ett bildspel baserad på det visuella materialet ihop med den ljudande versionen. Längden på musikstycket kommer att vara beroende på omfattningen av referensmaterialet.

Bakgrund

I det här kapitlet ges först en genomgång av designprocesser och en beskrivning av en variant för filmproduktion som jag använder mig av för hela det gestaltande arbetet. Sedan följer det en kort översikt över Artificiell Intelligens och dess utveckling fram tills då denna rapport skrevs under början av år 2023. Jag kommer även att belysa vissa parametrar som kan generera problem för användare av AI-stödda verktyg och varför användaren behöver vara medveten om dessa vid användning av verktygen i sitt gestaltande arbete.

Designprocesser

Designprocesser har funnits i någon form sedan människan först började fundera på hur hon kunde tillverka eller designa produkter, men började definieras först under 50-talet och etablerades mer spritt under 70- och 80-talet (Jones, Mitchell, & Jones, 1992). Med sin grund inom produktdesign och industri så har designmetodologi spridit sig till många olika discipliner. Fält som arkitektur, formgivning, reklam, mediaproduktion och till och med informations- och systemarkitektur tillämpar någon form av designprocess för att lättare kunna förmedla idéer och koncept mellan människor och grupper (Dubberly, 2004).

Gemensamma delar för designprocesser

De flesta designprocesser slutar i någon gestaltande artefakt (Jones, Mitchell, & Jones, 1992), och det kan argumenteras att även processer inom informationsarkitektur och systemarkitektur kan anses vara kreativa artefakter, även om det inte är ett fysiskt föremål som skapas utan mer abstrakt informationsarkitektur (Dubberly, 2004). Det som dessa processer har gemensamt är vissa moment som återkommer oavsett vad målet är nämligen divergens, transformation och konvergens. Med detta menas att först definiera vad som behöver lösas, vad problemet är, sedan transformera problemet till en specifikation (designspecifikation/brief) och sedan utförs designen/lösningen utefter denna specifikation (Jones, Mitchell, & Jones, 1992). Dessa tre moment delas ofta upp i ytterligare fyra delar, Discover, Define, Develop och Deliver, som kort kan beskrivas på följande vis (Choudhary, 2019) (Elmansy, 2021):

Discover

Vad skall skapas/designas? Vilket problem skall lösas? Detta är den initiala idén som startar allt. Den kan komma från en kund, ett behov internt, eller en kreatör som vill skapa något.

Define

Här kartläggs vad projektet skall omfatta, samt vilka faktorer som påverkar det. Idén börjar ta form och detaljeras baserat på både fakta- och informationsinsamling. Ofta hålls så kallade brainstorming-möten där idéer får flöda fritt och lösningar hittas ibland som kanske inte kommit med vid mer styrda möten. Denna delprocess resulterar i att det tas fram en designspecifikation, ofta kallad design-brief, som i detalj beskriver vad som skall designas eller skapas.

Develop

När nu designbriefen är klar så påbörjas den produktiva fasen, där artefakten börjar ta form. Detta är ofta en linjär fas där det byggs vad som önskas utefter designspecifikationen. Skulle det här uppstå frågor som inte har behandlats av designspecifikationen så kan dessa läggas tillbaka på discoverfasen och köras igenom processen en gång till för att få en mer komplett designspecifikation.

Deliver

Efter att kraven har uppfyllts som designspecifikationen har definierat så är det dags att leverera. Nu finns förhoppningsvis en färdig artefakt att leverera till kund eller ta till nästa projekt.

För att återkoppla till Jones definition av divergens, transformation och konvergens så är "Discover" början på divergensfasen, där en idé bryts ner denna i många delar och definitioner (Choudhary, 2019). "Define"-fasen fokuserar sedan dessa delar till en detaljerad designspecifikation och processen övergår från definitionsfas till produktionsfas (transformation). Produktionsfasen "Develop" fokuserar sedan hela designspecifikationen ner till en lösning vilket då motsvarar konvergensfasen och i "Deliver"-fasen så har en färdig artefakt skapats utefter specifikationen (Choudhary, 2019).

Filmproduktion

Då min grundtanke i detta arbete har varit att undersöka en designprocess genom att utföra den så valde jag tidigt att skapa något inom filmproduktion. Detta för att jag har mest erfarenhet av just filmproduktion och mitt gestaltande arbete i form av tonsättning är en vanlig del i en filmproduktion. Projektet leder då över från idé till visualisering till tonsättning på ett naturligt sätt och i enlighet med processen.

Designprocesser inom filmproduktion

Man kom tidigt fram till att någon form av visualisering av berättelser gjorde arbetet mycket lättare, då hela produktionsteamet blev mer fokuserat på vad som skulle göras och hur slutresultatet skulle se ut.

Det finns många olika processer inom filmproduktion och den processen som jag har valt att använda mig av används främst inom första fasen av en filmproduktion, ofta kallad förproduktionsfas (pre-production).

Förproduktionsprocessen delas in i följande moment (Katz, 1991):

Manusförfattande

Manuset som berättar berättelsen är oftast skrivet i en specifik manusform med scener och dialog tydligt separerade och utmärkta. Det tidigaste exemplet av denna manusform anses förekom redan år 1902 i George Méliès film "*A trip to the moon*" och har sedan dess utvecklats till att bli en standard inom film-, tv- och teaterproduktioner (Azlant, 1997).

Produktions-design

Denna fas handlar om visualisering av och att skapa ett visuellt språk för berättelsen. Här finner vi olika tekniker som används för att kunna visualisera de olika delarna i berättelsen, såsom moodboards och storyboards, men även miniatyrer, modeller och andra fysiska artefakter. Även storyboards användes av Méliès runt sekelskiftet för att planera visuella effekter och kameravinklar för sina filmer (Gress, 2014). Den form som vi oftast ser idag utvecklades senare av Disney på 1930-talet när de började ta vissa viktiga ögonblick i berättelsen, illustrera dessa och sätta upp dem på en tidslinje för att visa hur berättelsen fortskred (Katz, 1991).

Under 70-talet blev storyboarding väldigt viktig när visuella effekter skulle planeras som var kostsamma och svåra att genomföra fysiskt i samband med filminspelningen. Dessa kunde då enklare visualiseras och planeras genom att göra storyboards med illustrationer av scenerna.

Manusanalys

Nu när vi har en visualisering av berättelsen så börjar vi planera själva filminspelningen och genomförandet. Efter en gjord manusanalys skapas så kallade shotlists som kartlägger vilka scener som behöver filmas, vilka vinklar som behövs, vilka aktörer som är med, klädesdetaljer et cetera. Även här kan vi använda sig av storyboarding för att i ännu större detalj planera scenerna.

Filmfotografi (cinematography)

I de sista faserna innan själva filminspelningen påbörjas så kopplar vi in filmfotografen, som står för kunskapen om ljussättning, kameror, objektiv, filmmedia/digitalt media, et cetera, för att skapa en visuell identitet för filmen. Det kan vara en känsla som regissören, i samråd med filmfotografen, vill förmedla genom exempelvis ljussättning och färgval.

Generalrepetition

Sist men inte minst så är det dags för att öva scener. Detta innebär inte bara att aktörerna och skådespelarna skall öva utan det kan också innefatta hela teamet, särskilt när det gäller fysiska effekter och koreografi-element i scener såsom dans eller slagsmål/actionsekvenser.

En kort historia om AI

Artificiell intelligens (AI) har varit ett forskningsområde inom först matematiken (Turing, 1950) och sedan inom ingenjörskonst och även design (Simon, 1996). Allt eftersom datorer har blivit alltmer kraftfulla har även datamängder ökat och behovet av hjälpmedel för att behandla stora mängder data och även kunna analysera eller till och med ta beslut efter vissa regler. Ända sedan Turings tid så har det funnits viktiga milstolpar som försökt uppnås men datorkraften som fanns tillgänglig då var helt enkelt inte tillräcklig. Allt eftersom datorkraften har ökat så har dessa milstolpar sakta men säkert uppnåtts. Exempelvis så slogs den regerande världsmästaren i schack, Gary Kasparov av IBMs Deep Blue 1997 (Anoyoha, 2017) och bara några månader senare var det den kinesiske Go-mästaren Ke Jies tur att bli besegrad av Googles Alpha Go. Både schack men framför allt Go är spel som kräver enorma mängder data att behandla och ett stort antal permutationer för att kunna förutspå möjliga speldrag och utgångar och rangordna dessa, vilket har varit svårt att kunna behandla trots att det varit relativt enkla datatyper (Silver, Huang, Maddison, & Guez, 2016). En AI som skall lära sig spela schack gör inget annat och behöver inte ta hänsyn till något annat än att spela massor av omgångar mot sig själv eller mot människor.

AI-modeller

Nedan presenterar jag några olika AI-modeller, LLM/NLM, GPT, och generativ bildmodell, för att ge en bättre insikt i vilka problem som kan vara inbyggt i modellerna och hur dessa kan hanteras.

Large Language Models (LLM)

En av de största målstolparna inom AI har varit möjligheten att tyda, analysera och koda information som kommer i naturlig språkform. Det är lätt att förstå svårigheten i att tyda mänskligt språk som beroende på vilket språk som används har enorma variationer i syftningar, betydelser och där kultur och sociala faktorer spelar en stor roll i hur det skall tydas (Canestrino, Magliocca, & Li, 2022). Turing skapade redan på sin tid ett test som blev döpt efter honom "The Turing Test" (Turing, 1950), där en AI skulle lura en människa i ett blint test att tro att hen kommunicerade med en annan människa.

Detta test anses av en del ha passerats redan 2014 då en ukrainsk AI-bot skall ha övertygat 30% av domarna i ett blint test utfört av University of Reading, England, och trots kontrovers kring resultatet så började redan då det lyftas röster om att Turing testet behöver uppdateras (Lewis, 2014).

År 2018 släppte Google sin första språkmodell BERT, som kunde koda av texter för vidare behandling vid till exempel datainsamling eller analys (Devlin, Chang, Lee, & Toutanova, 2018). Redan året efter släppte OpenAI den första GPT modellen (GPT-2) som sedan låg till grunden för GPT-3 och GPT 3.5 som i 2022 fick ett webgränssnitt som kallas för ChatGPT.

Det är viktigt att förstå vad en LLM är, och vad som är tanken bakom den, vad den kan göra och vad den inte kan göra för människor har en tendens att tillskriva mänskliga attribut till dessa objekt. Särskilt om interaktionen med objektet görs med metoder som vi är vana att interagera med andra människor genom (Epley, Waytz, & Cacioppo, 2007).

Det finns många olika LLM modeller, men jag skall kortfattat gå igenom GPT-modellerna (GPT 3, GPT 3.5 och GPT 4) då ChatGPT och även den efterföljande Bing assistenten från Microsoft användes för text-generering och analys i detta arbete.

GPT Modellen

Språkmodellen GPT 3.5 var den första som gjordes publikt tillgänglig genom ett webbaserat interface som heter ChatGPT och fick stor spridning och användning.

Denna modell använder sig av en så kallad "Next word prediction" där modellen försöker hela tiden förutsäga vad nästa ord borde vara, baserat på tidigare ord i meningen/sammanhanget (Radford, o.a., 2019).

Man har sedan tränat modellen på enorma mängder text-data för att lära den att bli bättre och bättre på att analysera och kodifiera data men även kunna generera nya data baserat på vad svaret skulle kunna vara. Och det är det här ”skulle kunna vara” som kan vara problematiken med modeller som GPT när de implementeras och släpps fria.

GPT-modellen har endast ett mål och det är att generera ett för människor läsbart och förståeligt svar. Det innebär att den i olika generationer av implementationen av GPT modellerna har varit mer eller mindre ”kreativ” i sina svar. Den har tränats på existerande data som finns tillgänglig för modellen att skapa sina svar från, den går alltså inte ut och söker upp informationen som en senare implementation som Bing CoPilot gör. Detta har gjort att den skapar svar som låter korrekta, men som inte alltid är det. Dessa påhittade svar kallas för hallucinationer och jag går igenom lite mer om dessa i avsnittet ”Problematiska beteenden hos AI modeller”

En av intentionerna bakom GPT modellerna (Radford, o.a., 2019) har varit att skapa ett gränssnitt för att kunna interagera med en funktion med hjälp av vårt mänskliga språk.

Detta språkbaserade gränssnitt kan sedan kopplas ihop med andra funktioner, till exempel Bings sökmotor som ger ett effektivare sätt att kunna interagera mer naturligt med en sökmotor. Detta är exakt hur det är implementerat i Bings nuvarande ”CoPilot”, där GPT 4 används för att kodifiera inputen oavsett språk, sedan görs uppslag mot Bings sökmotor, och svaren som återkommer analyseras och jämförs mot frågan och en ny sammanställning med referenser till svarpunkterna presenteras. Så för användaren ser det ut inte bara att hen kan ställa en mer eller mindre generell fråga, utan svaren kommer också att sammanfattas och presenteras även om det är information från multipla källor. Bing CoPilot är sedan februari 2023 integrerad i Microsoft Edge webbläsare.

Generativa bildmodeller

En annan implementation av en naturlig språkmodell (NLM) är som gränssnitt för en generativ AI, som skapar bilder eller till och med video baserat på en prompt i mänskligt språk. Denna prompt kodifieras av en språkmodell och parametrarna skickas sedan in i en bildgenerativ modell. Dessa modeller har tränats på omfattande datamängder med bilder som redan kodifierats med beskrivande nyckelord eller texter på vad bilden föreställer (Xiaojin, Goldberg, Eldawy, Dyer, & Strock, 2007). Vem som gjort kodifieringen och även i vissa fall vilka bilder som använts redovisas inte alltid och är ett stort problem när transparens behövs

för att kunna bedöma materialet som modellen tränats på med avseende på representation, inkluderande och andra faktorer.

Dessa modeller är väldigt beroende på vilket bildmaterial som använts som datamängd, och hur dessa har kodifierats eftersom detta styr ju vilka resultat som sedan genereras, och här kan direkt ses lite potentiellt problematiska beteenden. Om Midjourney skall avbilda ”a beautiful person” så kommer den att generera fyra bilder på kvinnor, trots att person ej är könskodat. Och lika så den skall avbilda ”A criminal” så genereras idag bild på fyra män. Eftersom det är språkligt kodade variabler så kan det spekuleras i att det statistiskt sett troligen är fler kvinnor som har kodningen vackra och fler män som har kodningen brottsling i datamängden, men utan transparens eller uppföljning så blir det svårt att göra en bedömning hur representativ eller inkluderande träningsmaterialet är.

AI-verktyg inom musik

Jag har nu beskrivit verktyg som hjälper med språkanalys och bild-generering, men hur är det med musikaliska verktyg? Här skiljer utvecklingen sig åt på generativa modeller och ljudande hjälpmedel. När det gäller de generativa modellerna, de som faktiskt skapar noter och musik så är det lite mindre utvecklat än när det gäller ljudande hjälpmedel.

Ljudande AI-hjälpmedel – instrument

AI-tekniker har funnits ett tag för just instrument. Ett av de mer kända är Noteperformer som används inom mjukvara för att skriva/skapa notskrift, till exempel Dorico, Sibelius och Finale. Tyvärr finns det inte mycket publicerat om hur dessa funkar i grunden men enligt hemsidan för Noteperformer så är detta ett instrument som har en träningsmodell (Wallander, u.d.) i grunden och som tränats på att förutsäga hur en musiker skulle spela en passage. Genom att läsa musiken i förväg så kan den anpassa teknikerna som används och spelas upp så den gör en mer naturlig och verklighetstrogen övergång mellan tekniker, anslag, och så få ett mer naturligt uppspelande. Besluten som tas är då tränade på en datamängd över hur en mänsklig musiker skulle ha spelat samma passage och detta översätts till hur dynamik, anslag och vilka tekniker som skall spelas upp.

Ljudande AI-hjälpmedel – sång

En av de senare utvecklingarna inom ljudande instrument som använder AI-teknik är nära besläktat med språkmodeller och även spelande modeller och det är instrument som genererar sång baserat på noter och text. I detta arbete kommer jag dock inte att gå in speciellt djupare

på användandet av dessa då det tonsatta stycket inte kommer innehålla någon sång, men för den intresserade så finns det mycket gjort inom språksyntes och talsyntes där sång kan ses som en variant av genererat syntetiskt tal/språk.

Generativa musikaliska modeller

Tidigare under 2023 kom Google Research med information om en generativ motor (MusicLM) för att skapa musik baserat på ett text-baserat prompt som beskriver musiken (Agostinelli, o.a., 2023). Denna motor fanns inte tillgänglig för publikt användande vid skrivande stund men det finns ett antal exempel på musik som genererats och vilka text-prompt som används.

Dessa ”text till musik”-modeller är idag designade för att generera färdig musik, inte assistera vid komponerande vilket snarare skulle ersätta mig som tonsättare än stödja mig.

När jag tittade på andra verktyg på marknaden så som AIVA, Soundful och Soundraw så genererade dom dock väldigt generisk och repetitiv musik som i många fall kändes mer slumpmässiga än planerade.

Problematiska beteenden hos AI modeller

Det sätt som verktyg som använder sig av LLM-modeller för kodifiering av indata och även produktion av utdata har vissa inbyggda problem. Här skall jag beröra fyra av de mest tydliga och uppmärksammade problemen med att använda sig av dessa modeller.

Hallucinationer

Begreppet ”Hallucinationer” inom AI refererar till när en språkmodell är mer intresserad av att generera ett svar, än att svaret skall vara baserat i fakta (Ji, et al., 2023). När ChatGPT lanserades så blev det en enorm tillströmning av användare, och det upptäcktes tidigt att modellen var mer inriktad på att generera ett svar som lät bra, än som var helt korrekt. Om ChatGPT tillfrågades om en person som antingen var uppdiiktad eller som inte fanns i det material som modellen var tränad på, så skapade modellen ett svar som lät korrekt. I senare generationer (GPT-4) så informerade modellen om den inte hade information nog att svara på frågan. Och numera finns även texten ”ChatGPT may produce inaccurate information about people, places, or facts” inlagd under inmatningsfältet för att ytterligare förtydliga att svaren inte är helt tillförlitliga.

Även Bing CoPilot fick stora problem tidigt med svar som den inte kunde styrka med fakta, och svar som kunde låta bra men som var tydligt fabricerade (Vincent, 2023).

Jag som kreatör måste vara medveten om detta när jag använder mig av den här typen av generativa språkmodeller, alltid verifiera den information jag får.

Prompt-manipulation/injicering

Ett annat problem som upptäcktes med dessa språkmodeller är att de kan manipuleras för att presentera resultat som kanske inte alltid är önskvärda. Genom att skriva frågorna och manipulera prompten till modellen så kan eventuellt de filter och hinder som lagts in av modellens skapare för att hindra att den används för att till exempel sprida information som kan vara skadlig eller förnedrande kringås (Selvi, 2002).

Transparens

En mer och mer viktig faktor inom alla AI modeller och verktyg är att leverantören kan presentera vilka datamängder och källor modellerna är tränade på, vilka regelverk och filter som är inlagda och öppet och tydligt visa hur och varför resultatet blev vad det blev.

Här varierar just nu hur transparanta de olika företagen är med storlek på och ibland vilka datamängder som de tränats på. GPT modellerna av OpenAI fram till och med version 3.5 redovisade hur stor datamängd och hur många datapunkter som den tränats på men från och med version 4 så har inga sådana siffror redovisats.

Sedan behöver också hänsyn tas till innehållet på de data set som använts, och vad för information dessa innehåller med hänsyn till värderingar, kulturella och sociala faktorer, representation och fakta. Det kan ju vara svårt nog att få bort problematiska faktorer från facklitteratur och datakällor över lag och datamängderna som omfattas här är rent utsagt enorma. Som exempel kan nämnas att GPT-2 tränades på c:a 40 GB av text från internet-källor och det motsvarade ungefär 10 miljarder datapunkter (tokens) medan GPT-3 tränades på 499 miljarder datapunkter (Li, 2020).

Datainterferens och kvalité

Ytterligare ett problemområde är om språkmodell kontinuerligt tränar på material från internet kommer detta att innebära att den tränas även på material som genererats av annan AI (eller i vissa fall den själv). Detta är också relaterat till problematiken med varierande kvalité på källor på internet. Om informationen hämtas från osäkra källor så behöver den

analyseras och viktas annorlunda än om informationen kommer från säkrare källor (Xu, Cuesta-Infante, Berti-Equille, & Veeramachaneni, 2022).

Representation inom AI modeller och deras resultat

Vi har redan berört lite problematiken med att det kan vara svårt och veta vad för datamängder som dessa modeller är tränade på, och således kunna bedöma hur representativa de är för olika grupperingar. Vidare kan datamängder användas med intentioner som de inte var tänkta för från början vilket kan skapa och förstora eventuella fördomar eller stereotyper.

Idag är verktygen väldigt hårt bundna till just vad de är tränade på och vilken kodifiering som gjorts (i de fall där detta är relevant), vilket också innebär att vi som användare måste vara kritiska till de resultat vi får och värdera om de verkligen är inkluderande. Att blint lita på de AI-stödda verktyg som används utan att veta vilka eventuella bias, värderingar, fördomar, stereotyper som kan finnas i materialet som det tränats på, kan resultera i väldigt skeva och annorlunda resultat.

Det behövs även en översikt över hur träningsmaterial används inom modeller och bedömningar görs om det kommit in värderingar i kodifieringen. Och för att få detta krävs mer transparens på hur och på vilka datamängder som modellerna har tränats (Liliequist, Tubella, Danielsson, & Cocq, 2023).

Teori

Mitt huvudsakliga teoretiska grundområde kommer att vara arts-based research där jag genom att utföra ett gestaltande arbete också tittar på hur detta påverkas av AI-stödda verktyg och tekniker som jag använder under arbetets gång. Jag kommer att beröra resultat och forskning från några andra vetenskapliga områden i mitt arbete i diskussion om de olika aspekterna som berörs, men dessa områden kommer att vara av en stödjande natur snarare än huvudfokus för min forskning.

Arts-based research

Under mitt utforskande arbete så använder jag mig av ABR, eller Arts-Based Research då denna tittar extra mycket på hur utforskande konstnärligt baserad forskning kan utöka vår kunskap om sociala och kulturella fenomen (Leavy, 2020). Den konstnärliga skapandeprocessen i sig blir forskningsobjektet genom att jag tittar på hur jag skapar, vilka

metoder och tekniker som används och i detta specifika fall, hur användandet av specifika verktyg kan ändra min gestaltande process.

Hannula et.al skriver i sin bok *Artistic Research – Theories, Methods and Practices* ” (Hannula, Suoranta, & Vadén, 2005) om konstnärlig forskning som ”...it is an activity which challenges and exposes, opens up and activates in order to consider who we are, where we are, and how we are. Detta för mig är en bra beskrivning över den utforskande och explorative aspekten som finnes inom konstnärlig forskning. Genom att genomföra den gestaltande processen och titta på vilka faktorer som jag upplever som inkluderande, etiska, sociala eller kulturella när jag jobbar med de AI-stödda verktygen så får jag en djupare insikt i hur dessa faktorer påverkar mig själv och min gestaltande process.

Leavy skriver ”Arts-based researchers are not discovering new research tools, they are carving them” (Leavy, 2020) och att undersöka den gestaltande processen i sig och hur den påverkas genom teknologi ger en bra insyn i hur vår gestaltning påverkas av tekniken och verktygen som används. I detta arbete tittar jag ju specifikt på hur AI-stödda verktyg påverkar min gestaltande process och att kunna utforska och analysera hur jag måste förhålla mig till de resultat och produkter som kommer från att använda dessa verktyg.

Metod

I denna del beskriver jag vilken huvudmetod som jag använder mig utav samt hur genomförandet går till i mer detalj. Jag kommer att beskriva hur arbetet är uppdelat i tre faser, berättelse, visualisering och tonsättning samt vilka AI-stödda verktyg jag valt att använda mig av.

Autoetnografisk metod

Jag har valt att använda en autoetnografisk metod där jag håller en löpande diskussion och reflektion under arbetets gång och tittar på hur det jag gestaltar och de tekniker och/eller verktyg som jag använder kan påverkas av kulturella, sociala faktorer och värderingar. Denna metod kombinerar kultur-analys och tolkning i en berättande form och använder sig av antropologiska och socialvetenskapliga frågeställningar snarare än en beskrivande eller performativa (Chang, 2008).

Planerat genomförande

Arbetet kommer att delas upp i tre faser; Berättelse, visualisering och tonsättning. De första två syftade till att skapa ett visuellt material som skall ligga till grund som inspiration till tonsättningen i fas tre. Både text-baserat material och visuellt grafiskt material kommer tas fram med hjälp av AI-stödda verktyg i en traditionell designprocess, och resultera i en visuell artefakt som kan användas för nästa tonsättande fas.

Designprocessen i detta arbete

Då jag inte har tid, budget eller mankraft att skapa en hel film i denna rapport, hur trevligt detta hade varit, så har jag fokuserat mig på just delar av förproduktionsprocessen och dess designprocess.

Fas 1 – Berättelse och Manusförfattande

Jag skall med hjälp av språkgenerativ AI skapa en berättelse utifrån en enkel start-idé, och få fram ett grovt manus för en berättelse. I denna fas så kommer jag att starta med en enkel idé, och sedan utveckla denna vidare med egna idéer som jag kan få utifrån inspiration av det som AI-verktyget genererar. AI-verktyget får inte fria händer här utan jag kommer att guida den för att nå något som jag känner funkar i berättelsen som helhet.

Fas 2 – Visualisering och Produktions-design

Denna fas kommer att bli en av de större och viktigare. Här utgår jag från manuset och skapar visualisering med hjälp av bildgenerativa AI-verktyg. Här kommer AI-verktyget stå för grovarbetet i att faktiskt skapa bilden, och jag kommer att mer eller mindre detaljerat styra den för att få fram en illustration som jag tycker passar in i berättelsen. Denna fas blir väldigt mycket repeterande försök för att få fram en bild som jag bedömer passar, och det blir alltid baserat på min bedömning att bilden passar in i helheten som avgör.

Här skulle jag kunnat titta på att bara vara helt inspirerad av vad AI-verktyget presenterar och låta fantasin flöda fritt, men jag har valt att se det som ett rent verktyg, som styrs och hanteras av mig för att uppnå ett mål. Om jag hade låtit verktyget styra berättelsen helt fritt så blir det ett väldigt annorlunda projekt och det får bli ett annat arbete.

Fas 3 Tonsättning

De sista tre delarna i förproduktionsprocessen som beskrivits tidigare behövs endast om jag hade tänkt att genomföra hela filmproduktionen från start till slut. I detta arbete har jag dock

valt att endast göra en tonsättning trots att den delen normalt sett kommer efter förproduktionsprocessen.

Traditionella metoder och verktyg

Normalt sett i en förproduktionsprocess för film hade jag haft ett helt team av manusförfattare, illustratörer och creative-directors och samlat in kollage bilder för att skapa en visualisering och ett designspråk för berättelsen. Detta kommer jag att ersätta i mitt arbete med AI-styrda generativa verktyg, både för text och för bild.

Det finns även verktyg för att underlätta att skapa storyboard och bildspel för visualisering, mer eller mindre enkla att använda. Jag väljer här att visualisera med hjälp av bildgenerativa AI-verktyg och sedan klippa ihop resulterande storyboardet till ett bildspel och använda mig av ett så kallat klick-spår för att hålla en rytm. Jag beskriver mer om detta i genomförande delen.

Val av AI-stödda verktyg

Här beskriver jag nu kort vilka faktiska AI-stödda verktyg jag valt att använda i min gestaltande process.

Generativt textverktyg: ChatGPT och Bing CoPilot

För textgeneration så har jag främst att använda ChatGPT version 4 för generering och utveckling av berättelsen och Bing CoPilot för informationssökning. ChatGPT kan i skrivandets stund finnas genom att gå till <https://chat.openai.com/>. Modellerna som används i både ChatGPT och Bing CoPilot är GPT-3.5 och GPT-4 för ChatGPT och GPT-4 för Bing CoPilot.

Generativt bildverktyg: Midjourney

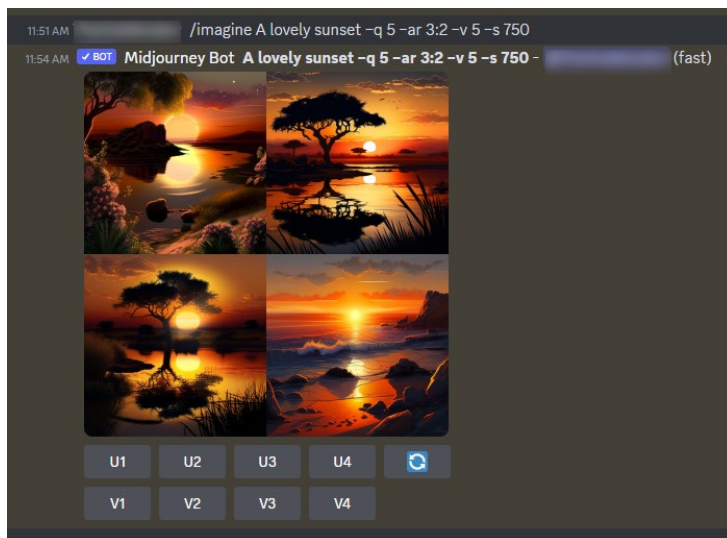
Verktyget som jag valt för att använda för bild generativa problem är Midjourney version 4 och 5. Arbetet började med version 4 men under arbetets gång så släpptes version 5 med en mer fotorealistiskt tränad modell så jag växlade till denna under visualiseringsprocessen. För att använda sig av Midjourney så styrs verktyget med prompts och olika parametrar. Generellt går det till att ett kommando (/imagine) skrivs följt av en beskrivande textprompt och sedan eventuella parametrar för att styra viktning av resultatet.

Ett exempel på en sådant prompt kan vara till exempel

/imagine A lovely sunset -q 5 -ar 3:2 -v 5 -s 750

Där "A lovely sunset" kommer kodifieras med hjälp av dess LLM modell, och sedan genereras bilden utifrån dessa. Parametrarna som följer styr då aspekter av arbetet, där till exempel *q* står för kvalitet (*quality*) vilket är en grov beskrivning av hur mycket processorkraft som kommer dediceras till arbetet och hur länge den kommer att jobba på tolkningen, *ar* är bildförhållande (*aspect ratio*), *v* styr vilken version av modell som används och *s* står för style vilket är en viktning på hur "kreativ" den skall vara i sin tolkning av prompten.

När prompten sedan skickats till modellen så genereras fyra resultat. Dessa resultat kan sedan skalas upp (upscale) till större upplösning eller användas för nya varianter av samma prompt baserade på den varianten som valdes. Jag har även en möjlighet att helt begära en ny omarbetning av prompten från början ifall ingen av de fyra förslagen är vad användaren söker.



Figur 1 */imagine a lovely sunset -q 5 -ar 3:2 -v 5 -s 750*

Musikaliska verktyg: Noteperformer

Efter att ha analyserat vad som finns tillgängligt för AI-stödd musik och komponerande i skrivande stund så har jag valt att gå vidare med att använda Noteperformer för generering/uppspelning av den ljudande artefakten. Detta är dock det enda AI-stödda verktyget som kommer att användas för just tonsättandet då övriga verktyg inte ger någon fördel när det gäller att skapa och komponera själva musiken.

Val av språk

När jag fick tillgång till de språkgenerativa verktygen så var de i huvudsak tränade på Engelska, och därför har jag valt att använda detta som språket i all textgenerering samt promptskapande.

Genomförande

I detta avsnitt går jag igenom hur arbetet har genomförts och de reflektioner jag haft under arbetets gång.

Fas 1: Berättelse och Manusförfattande

Första delen är att definiera en idé och utveckla denna till en berättelse. För att göra detta så använder jag mig av ChatGPT för att snabbt få ut en berättelse. Inspirerad av denna så kan jag sedan med hjälp av samma verktyg vidareutveckla denna till en färdig berättelse att visualisera.

Första försök – Draken och flickan

En av idéerna jag hade när jag påbörjade denna del i processen var att göra en liten saga om en liten flicka och hennes drak-vän som skulle ut och leta efter en bortsprungen liten kattunge. Verktöget som jag behöver främst i detta skede är ett språkgenererande verktyg, mer specifikt ChatGPT (GPT-3.5 och senare GPT-4).

Den kinesiske filosofen Lao Tse skrev i sitt verk Tao Te Ching; En resa på tusen mil börjar med ett steg, och så även denna resa. Fröet till storyn som skall skapas, den första prompten till AI-hjälpmidlet.

Jag börjar med att mata in följande prompt till ChatGPT för att få ut en första version av berättelsen:

Write a story about a girl going on an adventure with her pet dragon to find a lost kitten named fluffels.

Detta resulterade i följande resultat:

Once upon a time, in a world filled with magic and mythical creatures, there lived a girl named Ava. Ava was an adventurous young girl who loved exploring the vast forests and mountains that surrounded her village. She lived with her pet dragon, Blaze, who was a loyal and faithful companion.

One day, Ava heard about a lost kitten named Fluffels that had wandered away from its owner. Fluffels was a well-loved pet in the village and everyone was worried about its safety. Ava, being the kind-hearted person she was, offered to help find the kitten. She approached Blaze and explained the situation.

Blaze was more than eager to help. With a roar, he spread his magnificent wings and lifted Ava onto his back. Together, they set off on their journey to find Fluffels.

They flew over towering mountains and swooped down into deep valleys. They searched through dense forests and meadows filled with wildflowers. Along the way, they met many creatures who offered to help, but none had seen Fluffels.

After days of searching, Ava and Blaze were about to give up hope when they heard a faint meowing coming from a nearby cave. They cautiously approached the cave, and there they found Fluffels, scared and shivering.

Ava gently picked up the kitten and cradled it in her arms. Blaze let out a soft roar, and Fluffels, sensing that Blaze was a friend, relaxed and started purring.

Together, Ava, Blaze, and Fluffels made their way back to the village where Fluffels was reunited with its owner. The entire village celebrated their return, and Ava was hailed as a hero for her bravery and kindness.

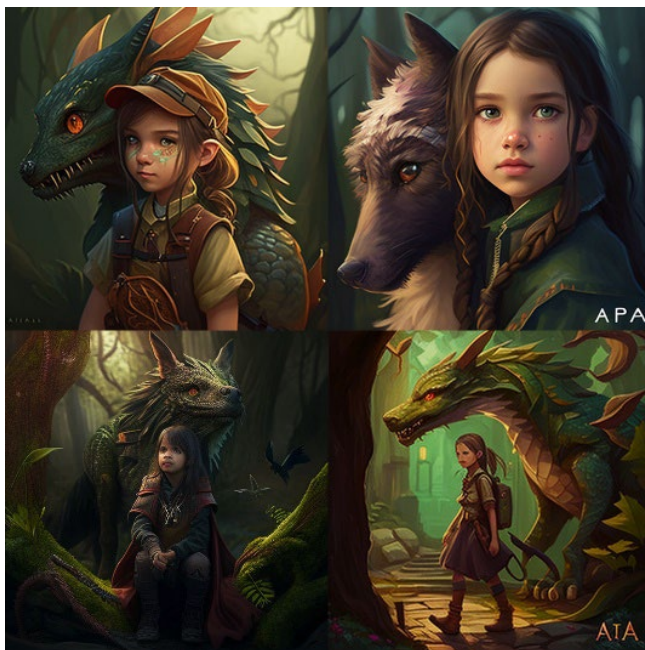
From that day forward, Ava and Blaze went on many more adventures together, and Fluffels always tagged along, always ready for their next quest. They traveled to distant lands, met new friends, and discovered wonders beyond their wildest imaginations. And every time, Ava and Blaze stood by each other, determined to face any challenges that came their way.

The end.

ChatGPT har nu alltså skapat ett litet frö till en historia som handlar om vad jag frågade efter. För att testa verktygen, eller mer specifikt visualiseringen så beslutar jag mig för att pröva att låta AI generera en bild baserat på text från vad ChatGPT har genererat. Så jag matar in den första stycket in i det bild-generativa verktyget (Midjourney v4) för att skapa en första illustration:

/imagine Once upon a time, in a world filled with magic and mythical creatures, there lived a girl named Ava. Ava was an adventurous young girl who loved exploring the vast forests and mountains that surrounded her village. She lived with her pet dragon, Blaze, who was a loyal and faithful companion –v 4

Detta resulterar i fyra små bilder genererade av Midjourney som sedan kan väljas ut att gå vidare med och antingen förstora (Upscale) eller se nya varianter av (Variants)¹.



Figur 2 /imagine once upon a time, in a world filled with magic and mythical creatures, there lived a girl named ava. ava was an adventurous young girl who loved exploring the vast forests and mountains that surrounded her village. she lived with her pet dragon

¹ En liten notering är att tillägget –v 4 är för att tvinga den till att använda den äldre motorn av version 4. I skrivandets stund (Maj 2023) är Midjourney uppe i version 5 och utseendet är markant annorlunda mellan versionerna.

Värt att notera här är att alla fyra bilder visar en vit flicka, trots att etnicitet inte var specificerat. Vidare är två av bilderna signerade, vilket kan vara en hallucinations-artefakt där en hel del av bildmaterialet som modellen är tränad på kan antas vara signerade originalverk och AI kan inte avgöra att det inte är en del av bildkompositionen². Då uppstår omedelbart två viktiga frågor om träningsmaterialet nämligen materialets representativitet och vems bilder är det tränat på.

”Hur representativt är materialet?”

Om jag inte specificerar parametrar så försöker den skapa något som så väl som möjligt stämmer med frågan. Och om materialet kan undersökas som modellen är tränad på så kan det vara som så att statistiskt sett är det fler illustrationer av vita flickor i sago-/fantasygenren. Och har modellen inga specifika parametrar att den skall skapa en illustration av andra etniciteter, kön, kulturella eller sociala aspekter så kommer den att generera något som enligt hur modellen har tränats har högst sannolikhet att träffa rätt. Dock kan det vara svårt att få tillgång till exakt vilken datamängd som modellen är tränad på, särskilt om det är kommersiella aktörer som kontrollerar modellen.

”Har vi rättigheter att använda de bilder som modellen tränats på?”

De flesta av dessa bildgenerativa modeller är tränade på väldigt stora mängder bilder, och det har varit begränsad transparens av vilka datamängder som de använt. Det pågår i skrivandets tid multipla juridiska processer där konstnärer har kunnat identifiera delar av sina egna konstverk och till och med sin egen tydliga stil i genererade bilder från olika modeller. Dessa konstnärer har dock inte blivit tillfrågade om de vill ha sin upphovsrättsskyddade konst behandlad på detta vis. En annan fråga som också blir aktuell är när verktyget uppmanas kopiera en stil efter en konstnär. Detta är inte ett lika tydligt kopieringsbrott men jag personligen kan tycka det är inte god sed att så tydligt kopiera en annan konstnärs stil.

² Samma situation kan uppstå när AI modellen är tränad på bilder som innehåller så kallad vattenmärkning. Den kan inte avgöra att vattenmärkningen inte är en normal del av bilden och således kan bilder genereras som tydligen baserats på bilder från vattenmärkta samlingar. Detta gissar jag är en indikation på att modellen tränats på bilder som leverantören egentligen inte har rättigheter till, eftersom då borde bilderna levererats utan vattenmärkning.



Figur 3 /imagine an oil painting of darth vader in the style of leonardo da vinci. composition like the mona lisa. --ar 3:2 --s 750 --q 2 --v 5.1 --style raw

Ändring av stil

Även om berättelsen skulle kunna gå vidare med den lilla flickan och draken på sina äventyr, så väljer jag här efter ett par dagars experimenterande att ändra inriktning på berättelsen helt. Valet för detta är främst baserat på det faktum att den berättelsen hade blivit för omfattande för arbetet och kräva multipla tonsättningar.

Andra försöket – science fiction

Jag beslutar mig nu i stället att skapa en lite mindre berättelse, som har lite fler parametrar definierade i början. Jag tänker här att jag skall begränsa historien till ett kortare scenario, och vara tydligare med vissa punkter som jag vill ha med. Jag provar följande prompt:

Write me a scene where a deep space exploration ship arrives at a deep green jungle planet after a long hyperspace jump. I want scenes where you see the captain observe from the command deck how the planet appears before him, and he marvels at its beauty.

Vilket genererade följande respons från ChatGPT:

The deep space exploration ship emerged from the hyperspace jump with a jolt, its engines roaring as it came to a stop in the void of space. The ship's

captain, Commander Jackson, stood on the command deck, his eyes glued to the viewscreen as the ship's scanners slowly mapped out the new planet they had arrived at.

As the ship's scanners began to reveal the planet's surface, Commander Jackson watched in awe as a vast expanse of deep green jungle spread out before him. The thick foliage was punctuated by towering trees, their branches reaching out like fingers towards the sky. Streams and rivers snaked their way through the dense vegetation, their crystal-clear waters reflecting the light of the nearby star.

Commander Jackson couldn't help but feel a sense of excitement well up within him as he gazed upon this stunning new world. He had spent years traveling through the cold emptiness of space, searching for something new and exciting, and finally, he had found it.

"Open a channel to the science team," he said, his voice crackling over the ship's intercom. "Tell them to prepare for a landing on the planet's surface."

As the ship descended towards the planet, Commander Jackson watched as the dense jungle came into sharper focus. He could see flashes of color as birds and insects flitted through the canopy, and he marveled at the planet's sheer biodiversity.

The ship touched down on a wide open plain, and Commander Jackson led the way as the science team disembarked, eager to explore this new world. They walked through the jungle, taking in the sights and sounds of the alien ecosystem around them.

As they made their way deeper into the jungle, Commander Jackson couldn't help but feel a sense of wonder and awe. This was a world unlike anything he had ever seen before, and he knew that they had only scratched the surface of what this planet had to offer.

For the first time in a long time, Commander Jackson felt a sense of purpose, a sense of belonging. He knew that this was where he was meant to be, exploring the unknown, and discovering the wonders of the universe.

Detta gav ganska mycket berättelse att jobba vidare, och jag vill nu direkt börja titta på berättelsen i manusform. Jag ber därför ChatGPT att skriva om samma berättelse, fast nu i manusform. Jag skriver in följande prompt:

Rewrite this in script form

Eftersom motorn är kontext-känslig så vet den att "this" hänvisar till historien som den skapade i föregående moment, så får jag nu följande manus från ChatGPT:

FADE IN:

EXT. DEEP SPACE - DAY

A deep space exploration ship bursts out of hyperspace with a jolt.

INT. COMMAND DECK - DAY

COMMANDER JACKSON, the ship's captain, stands on the command deck, eyes glued to the viewscreen as the ship's scanners map out the new planet.

The scanners reveal a vast expanse of deep green jungle, punctuated by towering trees and crystal-clear streams.

COMMANDER JACKSON

(whispering)

My God... it's beautiful.

He takes a deep breath, feeling a sense of excitement well up within him.

COMMANDER JACKSON

(opening a channel)

Open a channel to the science team. Tell them to prepare for a landing on the planet's surface.

The ship descends towards the planet, the dense jungle coming into sharper focus.

EXT. PLANET'S SURFACE - DAY

The ship touches down on a wide open plain, and the science team disembarks, eager to explore the new world.

They walk through the jungle, taking in the sights and sounds of the alien ecosystem around them.

As they make their way deeper into the jungle, Commander Jackson can't help but feel a sense of wonder and awe.

COMMANDER JACKSON

(to himself)

This is where I'm meant to be.

He leads the team further into the jungle, determined to discover the wonders of this new world.

FADE OUT.

Här identifierar jag ett fyra viktiga händelser i berättelsen:

- När skeppet anländer till planeten
- Commander Jackson som står och ser utöver planeten på kommandobryggan
- Landningsskeppet med utforskningsteamet som landar
- Utforskningsteamet som söker genom djungeln

Själva berättelsen är dock lite intetsägende, så jag vill lägga till en händelse där något viktigt händer som ändrar hela den underliggande känslan i berättelse. Min tanke här är att få in något element som kan knyta till vår egen planet, och då skapa en fråga hos åskådaren om de två planeterna är ihopkopplade på något vis.

Lägga till en twist i berättelsen

Nu går jag tillbaka till ChatGPT och ställer ett nytt prompt med mer information om vad jag vill se:

Write me a scene where a deep space exploration ship arrives at a deep green jungle planet after a long hyperspace jump. I want scenes where you see the captain observe from the command deck how the planet appears before him, and he marvels at its beauty. I want the scene to include that we see a research vessel undock and travel down to the surface, land in the jungle and a small team of explorers disembark and start exploring the jungle. They will be observed and followed by a tribe of blue skinned aliens. After a while exploring they will arrive at a large Mayan temple structure in the middle of the jungle to the surprise of the explorers.

Språket i detta prompt har några punkter som är viktiga; det är mer direkt och specifikt på vad jag vill ha i stället för att vara mer ”artigt”. Inga ”skulle du kunna...”, ”om du vill...” utan rakare tydligare instruktioner.

Jag lägger in en referens till Maya-tempel som är kopplingen till vår planet. Att jag just väljer Mayakulturen är dels för att det inte finns väldigt mycket information om den, och för att det finns en hel del pseudovetenskapliga konspirationsteorier om att den civilisationen var i kontakt med utomjordiska varelser (Wikipedia, Erich von Däniken, 2023). Jag tänker här att det kan vara liten rolig populärkulturell referens just på grund av detta.

Eftersom jag fortfarande är i manus-kontext så producerar nu ChatGPT följande scen:

FADE IN:

EXT. DEEP SPACE - DAY

A deep space exploration ship emerges from hyperspace with a jolt.

INT. COMMAND DECK - DAY

Captain PARKER stands on the command deck, his eyes fixed on the viewscreen as the ship's scanners reveal the new planet below.

As the image sharpens, a vast expanse of deep green jungle comes into view, stretching out as far as the eye can see.

CAPTAIN PARKER

(whispering)

Wow. It's beautiful.

He takes a deep breath, feeling a sense of excitement well up within him.

CAPTAIN PARKER

(opening a channel)

"Attention all crew, this is the Captain speaking. Prepare for landing on the planet's surface."

EXT. PLANET'S SURFACE - DAY

The ship lands in a clearing in the dense jungle, and a research vessel undocks and descends towards the planet's surface.

As the research vessel lands, a small team of explorers disembark and begin to explore the jungle.

As they make their way deeper into the jungle, they start to notice that they are being observed and followed by a tribe of blue-skinned aliens.

One of the explorers, Dr. LEE, steps forward and holds out her hand in a gesture of peace.

DR. LEE

(in a friendly tone)

"Hello, we come in peace. We are explorers from another planet."

The blue-skinned aliens hesitate for a moment, then cautiously approach the explorers.

As they travel deeper into the jungle, the explorers are amazed by the diversity of plant and animal life they encounter.

Eventually, they arrive at a clearing where a massive Mayan temple structure looms over them.

DR. LEE

(in disbelief)

"Is that what I think it is?"

The team approaches the temple, and as they do, they realize that it is far older than anything they have ever seen before.

As they begin to explore the temple, they discover that it is filled with intricate carvings and artifacts, evidence of a long-lost civilization that once thrived on this planet.

The team marvels at the discoveries they have made, and as they prepare to return to their ship, they can't help but feel a sense of wonder at the mysteries that still lie undiscovered on this alien world.

FADE OUT.

Revidering och redigering av manus

När ett antal grundblock i storyn skapats som jag kan jobba med, och jag redigerar manuset så jag får en tydligare röd linje och lite mindre dialog och moment som inte kommer att visualiseras. Jag lägger till vissa moment för att förhöja den visuella delen av filmen, till exempel färden genom hyperrymden i början och avslutande bilderna i rymden där planeten tonar ut, och jag tar bort mötet mellan människorna och invånarna på planeten för att mer direkt låta människorna bli överraskade av mayatemplet i djungeln.

Det färdiga manuset kan ses i appendix 1 och återfinns även i den animerade storyboarden som text vilket återfinns i appendix 2.

Fas 2: Visualisering och Produktions-design

När jag har en grundstory så börjar jag att jobba med att visualisera denna. Traditionellt sett när det skall göras rörlig film så skapas det ett bildspel, så kallat storyboard för att beskriva berättelsen över tid (Katz, 1991). Dessa metoder finns beskrivna i tidigare kapitel men det vanliga är att visualiseringen utförs av illustratörer eller skapas med miniatyrer eller 3D-visualiseringsverktyg på dator.

Eventuellt skapas också collage som definierar visuella element i filmen, färgtoner, känsla och detta kallas ofta för moodboard. Traditionella metoder för att tillverka dessa går från att göra färgpaletter, klippa och klistra till att använda sig av programvaror för bildbehandling och producera det digitalt.

För detta arbete så valde jag att bygga endast en storyboard då de illustrationer som genererats för storyboardet är utvalda att vara stilmässigt liknande. Ett moodboard skulle reflektera ungefär samma sak så jag tyckte inte det tillförde något extra.

Visualisering av viktiga händelser

För att visualisera en berättelse delas den upp i ett antal händelser, ögonblicksbilder ut berättelsen som beskriver eller visar något viktigt som händer eller syns. Dessa blir sedan beståndsdelarna i storyboardet för att visa vad som händer i berättelsen.

Jag testar att lägga in text direkt från manuset till bild-generativa verktyget och då stöter jag snabbt på ett av problemen med språkbaserade modeller.

/imagine The deep space exploration ship emerged from the hyperspace jump with a jolt, its engines roaring as it came to a stop in the void of space. The green jungle planet expands beneath it as the ship enters orbit.

Denna prompt resulterar i följande bilder:



Figur 4 /imagine the deep space exploration ship emerged from the hyperspace jump with a jolt, its engines roaring as it came to a stop in the void of space. the green jungle planet expands beneath it as the ship enters orbit.

Som jag ser här så tolkas ”The green jungle planet” lite väl bokstavligt och är inte exakt vad jag hade tänkt mig. Just detta beteende är ett bra exempel på att vi kan tolka språk på olika sätt, orden är desamma men jag hade tänkt mig en annan innebörd. Jag försöker med en variant med lite mer förtydligande beskrivning:

/imagine The deep space exploration ship emerged from the hyperspace jump with a jolt, its engines roaring as it came to a stop in the void of space. The planet expands beneath it as the ship enters orbit. --q 5 --ar 3:2 --s 750

Vilket får modellen att tolka det på ett annorlunda sätt och där en planet faktiskt syns i bilden.



Figur 5 /imagine the deep space exploration ship emerged from the hyperspace jump with a jolt, its engines roaring as it came to a stop in the void of space. the planet expands beneath it as the ship enters orbit. --q 5 --ar 3:2 --s 750

Att skriva detaljerade prompter

Läsaren har säkert hört ”En bild är värd tusen ord”, som tillskrevs Arthur Brisbane redan 1911 och ett problem kan vara att det omvända resultatet kan mycket väl vara korrekt. Ju mer specifik beskrivningarna är vad som önskas, desto bättre kan modellen generera något som

är närmare det som sökes. Detta gör att vissa prompts kan bli enormt stora, och ett exempel på detta är det prompt som jag använt för att generera porträtt av de blå alien-humanoiderna.

/imagine blue alien woman, dancing in the jungle, full body shot, wide field of view, centered, showing entire body dancing, award-winning cinematography, award-winning photographyfashion, photoshooting, detail on a beautiful old cosmic shaman, vogue pose, a real life like old beautiful shaman in colorful stars theme clothes, interesting pose, dynamic pose, spiritual dance, dancing, hiperrealistic photography, Premium Portrait Photography::4 well-lit, sharp-focus, high-quality, artistic, unique, award-winning photograph, Canon EOS 5D Mark IV DSLR, f/8, ISO 100, 1/250 second, close-up, natural light, professional, flattering, headshot, glibatree style::3 black and white, grainy, High-speed photography, Color Grading, Ultra-Wide Angle, Depth of Field, hyper-detailed, insane details, intricate details, Unreal Engine, Gorgeous__, Cinematic , Color Grading, Editorial Photography , Photography, Photoshoot, Shot on 70mm lens, Depth of Field, DOF, Tilt Blur, Shutter Speed 1/5000, F/6.3, White Balance, 32k, Super-Resolution, Megapixel, Pro Photo RGB , VR , Lonely, Good, Massive, Half rear Lighting, Backlight, Natural Lighting, Incandescent, Optical Fiber, Moody Lighting, Cinematic Lighting, Studio Lighting, Soft Lighting, Contour, Beautiful Lighting, Accent Lighting, Screen Space Global Illumination, Ray Tracing Global Illumination, Optics, Scattering, Glowing, Shadows, Rough, Shimmering, Ray Tracing Reflections, Lumen Reflections, Screen Space Reflections, Diffraction Grading, Chromatic Aberration, GB Displacement, Scan Lines, Ray Traced, Ambient Occlusion, Anti-Aliasing, Fxaa, TAA, RTX, SSAO, hdr, deformed, --ar 3:2 --v 4 --q 2

Denna prompt genererade följande bild:



Figur 6 bild genererad utifrån väldigt detaljerad prompt

Behövs all denna text? Svaret är lite komplicerat. All den text kodifieras av Midjourneys egna språkanalyserande LLM modell, och det blir då vikter i när verktyget skall sätta ihop olika aspekter i bilden. Så ett ord i listan kan göra att en viktning blir större, och den använder mer element från en källa än en annan. Så även om vi skulle tolka texten på ett sätt när vi läser den, och det kan till och med upplevas som motsägande fraser (Natural Lighting samtidigt som Studio Lighting och Cinematic Lighting?). Vad som exakt används i den slutgiltiga sammanslagningen av bilderna är mer dolt och endast de som byggt modellen vet hur mycket som det viktas.

I vissa modeller, exempelvis DALL-E 2 och Stable Diffusion, kan viktningen styras på enskilda fraser och ord för att påverka resultatet lite mer vilket kan snabbt resultera i väldigt komplexa prompts.

En annan sak att notera här är att dessa kodifieringar ger ett resultat i version 4 av Midjourneys modell, medan i version 5 ger det ett väldigt annorlunda resultat:



Figur 7 samma prompt i midjourney version 5

Just modellen som finns i verktyget Midjourney är lite mer låst och de tränar den mellan modellversionerna på olika fokusområden. För version 5 var fokus mer på fotorealism medan för version 4 var det mer illustrativt fokus.

Som jag nämnde tidigare kan det bli ett stort sökande på att hitta rätt specifika ord eller termer att använda för att Midjourney skall kunna skapa det som önskas. Ett ytterligare exempel på detta som jag stötte ihop med var när jag skulle visualisera scenen med landingsfarkosten. Väljer jag använda ordet "Shuttle" så tolkade modellen det direkt som en amerikansk SpaceShuttle som inte var vad jag ville ha.

*/imagine a small shuttle exiting the hangar bays of a large spaceship --q 5 -
-ar 3:2 --s 750*



Figur 8 /*imagine a small shuttle exiting the hangar bays of a large spaceship --q 5 --ar 3:2 --s 750*

När jag byter ut shuttle mot det mer generella landing craft så får jag lite mer generella skepp som genereras, dock ser jag här att den inte kodat av alla delar korrekt:



Figur 9 /*imagine a small landing craft exiting the hangar bays of a large spaceship into space. extreme long shot. 28mm. view from outside the ship. --q 5 --ar 3:2 --s 750*

Och den sista iterationen som resulterar i en bild jag väljer att använda (bild nummer 2) för visualisering:



Figur 10 /imagine a small landing craft exiting the hangar bays of a large spaceship into space. extreme long shot. 28mm. view from outside the ship. --q 5 --ar 3:2 --s 750

Detta letande efter rätt ord att beskriva bilden blir en väldigt stor del av arbetandet med det bildgenerativa verktyget och det är nästan omöjligt att förutsäga vad för resultat som fås genom att ord läggs till, dras ifrån eller förändras.

Skapande av ögonblicksbilder och första tidslinjen

Jag genererar först med hjälp av AI-verktyg en visualisering av ett antal händelser i berättelsen och börjar lägga in dessa i en tidslinje. Detta gör jag för att få en uppfattning om vad för mer detalj jag kommer att behöva allt eftersom arbetet fortskrider. För att enkelt kunna koppla varje visualisering till en scen, så lägger jag också in hela manuset i form av textning ovanpå bildmaterialet. Nu kan jag ta scen för scen och bryta ner denna i mindre visuella delar, kameravinklar, klippningar och effekter för att skapa mer rörelse och detalj på tidsaxeln.

Den iterativa visualiseringen

Nu börjar ett konstant hoppande fram och tillbaka mellan att klippa upp scenen i mindre beståndsdelar, och till den bildgenerativa modellen för att visualisera nya ögonblick, kameravinklar och delar av scenen. Här hade det varit önskvärt att kunna göra animerade sekvenser i stället för bara stillbilder, men även här behövs i skrivande stund mer förfining och utveckling av verktyg för att generera video baserat på textinput.

Efter att ha bearbetat manuset, nyckelbilderna och delat upp allt i mindre delar så innehåller slutversionen 37 nyckelbilder totalt och dessa återfinns i appendix 2.

För att komma fram till dessa 37 bilder, så gjordes 219 olika prompter som i sin tur genererade 1-4 bilder vardera. Och från alla dessa skapade jag sedan bildspelet som sedan ligger till grund för tonsättningen genom att lägga in det i ett video-redigeringsprogram på en tidslinje med ett klick-spår för att hålla takten.

Nya versioner under arbetets gång (ChatGPT 4 & Midjourney v 5)

AI utvecklingen under skrivandets tid har verkligen varit extremt intensiv och två av mina valda verktyg fick stora uppgraderingar. ChatGPT bytte sin språkgenerativa LLM modell från GPT 3.5 till GPT 4 och OpenAI/Microsoft släppte Bing CoPilot som hade GPT 4 modellen i grunden men hade möjlighet att söka information via Bings sökmotor.

Även Midjourney uppgraderade sin modell från v 4 till v 5 med ett större fokus på att generera fotorealistiskt material. Detta skapade en tydlig visuell skiljelinje mellan v 4 och v 5 och gjorde att jag skapade om några av frågorna i version 5 för att se om det blev ett bättre resultat.

Prompt som använts i följande två exempel:

*/imagine a blue alien woman in Mayan outfit looking out from behind bushes.
Extreme closeup on face, cinematic lighting. --q 2 --ar 3:2 --s 50*



Figur 11 /imagine a blue alien woman in Mayan outfit looking out from behind bushes. Extreme closeup on face, cinematic lighting. --q 2 --ar 3:2 --s 50--v 4



Figur 12 /imagine a blue alien woman in Mayan outfit looking out from behind bushes. Extreme closeup on face, cinematic lighting. --q 2 --ar 3:2 --s 50 --v 5

Som syns ganska tydligt ger version 5 ett mer fotorealistiskt uttryck än version 4, och i vissa situationer så funkade det bättre medan i andra fall så föredrog jag estetiken från version 4 då den mer stiliserade kändes mer konsekvent med övriga bilder.

Slutgiltig klippning till klickspår

När jag nu hade samlat ihop alla visualiseringar, och de scenerna och kameravinklar/effekter/rörelser som skulle vara med i berättelsen, så klipptes detta ihop mot ett så kallat klick-spår vilket är ett ljudspår med en metronom som går i ett specifikt tempo. Jag ville inte att mitt eget tonsättande skulle färgas av vilken temporär musik som användes vid klippningen så därför valde jag att klippa mot ett klick-spår i stället.

Fas 3: Tonsättande

Efter visualiseringen har jag en visuell artefakt i en tidslinje som berättar en berättelse över tid och den har ett satt tempo som jag kan komponera mot. Men innan jag kan gå in på tonsättandet måste jag titta igenom vad det är för berättelse jag har och vad för element som kan bygga på.

Viktiga händelser i berättelsen

När jag spelar igenom det animerade bildspelet så identifierar jag några moment och hållpunkter i tiden som jag kan använda för att bygga upp min tonsättning.

Först är tempot och det faktum att betraktaren kommer in i berättelsen relativt sakta med långa visuella sekvenser och övergångar mellan scener. Detta pågår under första minuten i relativt sakta tempo på ungefär 60 slag per minut (bpm) fram till scen 4 (Takt 1-18, 0:00-1:05) där vi ser kaptenen på bryggan som ser ut över planeten.

Nästa stora händelse är sedan när landningsskeppet landar på planeten i scen 8 (Takt 32-34 2:00 – 2:13) där en större musikalisk förändring skulle ge en auditiv förändring av miljö genom att jag ändrar tempo, dynamik, och orkestrering.

Efter att utforskarna tagit sig genom djungeln så kommer de upp till Maya-templet i scen 12 (Takt 52 3:39) och sedan scen 13 (Takt 56 4:00). Här vill jag ha ytterligare en förändring av musiken som skall leda över till slutscenen i scen 14 (Takt 60 4:19) där kameran åker sakta bakåt och sedan tonas bilden ut till svart.

Musikaliska element

Nu när jag har en grov struktur så tittar jag på andra faktorer som jag kan vilja ha med när jag komponerar musiken såsom orkestrering och vilken harmonik eller musikalisk stil jag vill använda mig av.

Jag väljer att börja stycket i en mer traditionell västerländsk form med en standard orkesteruppsättning för att sedan lägga in andra mer etniska element i orkestreringen när vi kommer till djungelmiljön i berättelsen.

I slutet av stycket tänkte jag också återgå till den mer traditionella västerländska orkestreringsformen så jag kommer att behöva ha en fungerande övergång mellan dessa två orkestreringar. Min tanke här är att jag inte byter orkestrering helt, utan komponerar med element från den västerländska orkestreringen med i den mer etniska delen även om dessa inte hade använts traditionellt i den etniska formen.

Musik till science-fiction filmer har traditionellt sett ofta innehållit elektronisk experimentell musik och det var faktiskt inte förrän under 60 och 70-talet som orkestral musik började bli populär, främsta exemplet är Stanley Kubricks film från 1968, *2001: A space odyssey* som använde sig av musik utav Johan Strauss II (Also sprach Zarathustra) och avant-gardé kompositören György Ligeti (Atmosphères) men även George Lucas *Star Wars VI: A new hope* från 1977 som använde sig av orkestral musik komponerad av John Williams (May, 2020).

Att hitta information om Maya-musik

Jag valde ju tidigt att ta med element från Mayakulturen som fanns i Mayariket (250 VT – 900 VT³) som sträckte sig från södra México genom Centralamerika ner till Nicaragua (Britannica, 2023). Min tanke är nu att ta med element från Maya-musik i min tonsättning men det finns ett stort problem: jag kan inget om Mayakultur eller musik.

Min lösning här är att använda mig av en annan funktion som språk-generativa och analyserande modeller har, och det är att sammanställa information. För att få en startpunkt så valde jag att fråga först ChatGPT:

I am writing a score for a short movie about deep space exploration and the first encounter with a civilization that has direct similarities with the Mayan culture on Earth. I want you to write me a theme melody with Mayan influences combined with western tonal harmony.

³ VT refererar här till vår tideräkning (v.t. eller VT) som är ett sätt att skriva årtal som inte refererar till någon religiös tideräkning (Wikipedia, Vår tideräkning, 2022).

Och fick följande svar tillbaka:

As an AI language model, I am unable to directly create or provide musical notation or audio files. However, I can offer some suggestions for creating a theme melody that combines Mayan influences with Western tonal harmony.

- 1. Start with a pentatonic scale: Mayan music often utilizes a pentatonic scale, which consists of five notes per octave. A common pentatonic scale is the minor pentatonic scale, which includes the root, minor third, perfect fourth, perfect fifth, and minor seventh notes of the natural minor scale. You can use this as a starting point for your melody.*
- 2. Incorporate Mayan instruments: The Mayan culture features a variety of unique instruments, such as the tunkul (a type of slit drum), marimba, and various flutes. Consider using the sounds of these instruments to add authenticity to your composition*
- 3. Use syncopated rhythms: Mayan music often employs syncopated rhythms, which can create an exciting and energetic feel. Syncopation involves emphasizing off-beat notes, creating an unexpected and engaging rhythmic pattern.*
- 4. Combine with Western harmony: To merge Mayan influences with Western tonal harmony, consider using common Western chord progressions and harmonies alongside your pentatonic melody. For example, you could use a I-IV-V progression (in a major key) or a i-iv-VII progression (in a minor key).*
- 5. Add a drone: A drone is a sustained note that is held throughout a piece of music, and it can be used to provide a foundation for your melody. This technique can be found in both Mayan and Western music, and it can help to create a cohesive sound between the two styles.*
- 6. Develop a memorable theme: To create a memorable theme melody, try to develop a simple, singable melody that can be easily recognized and repeated throughout your short movie. You can experiment with different*

motifs and phrases, combining Mayan and Western elements, until you find the perfect balance.

Remember to experiment and iterate on these ideas to create a unique theme that captures the essence of deep space exploration and the first encounter with a civilization resembling the Mayan culture.

Nu vill jag ha lite mer specifik information så jag ställer följande prompt till ChatGPT:

Can you give me some references to information about Mayan tonal harmony and scales

ChatGPT svarade med:

Although there is limited information available about the specific tonal harmony and scales used in ancient Mayan music, it is generally understood that their music was based on pentatonic scales and utilized various unique instruments. Here are some aspects of Mayan music that could be helpful in understanding their tonal harmony and scales:

- 1. Pentatonic scales: As mentioned earlier, Mayan music is believed to have utilized pentatonic scales, which consist of five notes per octave. The most common types of pentatonic scales are the major pentatonic (C, D, E, G, A in the key of C) and the minor pentatonic (C, Eb, F, G, Bb in the key of C). These scales have a distinct sound that can give your music an authentic Mayan feel.*
- 2. Modes: Mayan music could have also used modes, which are scales derived from the diatonic major scale by starting on a different note. While it is unclear which specific modes were employed by the Mayans, incorporating modes into your composition can provide an ancient and exotic sound.*
- 3. Instruments: Mayan music featured a variety of unique instruments, such as the tinkul (a type of slit drum), marimba, and different kinds of flutes, including the clay ocarina and the wooden tzab (end-blown flute). These*

instruments played an essential role in creating the characteristic sound of Mayan music.

- 4. Rhythms and polyrhythms: Mayan music often employed syncopated rhythms and polyrhythms, which involved playing two or more contrasting rhythms simultaneously. Incorporating these rhythmic elements can help create an authentic Mayan musical texture.*
- 5. Vocal music: Mayan vocal music, such as chants and songs, could provide insight into their tonal harmony and scales. Unfortunately, due to the lack of preserved recordings or written music, it is difficult to study their vocal music in depth.*

It is important to note that the information available on Mayan music is limited, as much of their musical culture and knowledge has been lost over time. However, by incorporating the elements mentioned above, you can create a composition that reflects the essence of Mayan music while blending it with Western tonal harmony. You can also listen to contemporary Mayan music and Central American folk music to gain more insight into the musical traditions and styles of the region.

Svaret gav ganska mycket bra fakta att jobba vidare med men är något av det en hallucination? ChatGPT har dessutom påpekat att det finns begränsat med information om detta och föreslagit att jag söker mer information och hittar ljudande exempel på modern mayakulturell musik.

Orkestrering för Mayamusik

När det gäller orkestrering så nämner ChatGPT ett antal instrument och efter lite efterforskning så finner jag att lerflöjter så som ocarina använts bland annat av mesoamerikanska kulturer som Mayakulturen tillhörde, tillsammans med blockflöjter, trätrummor, handtrummor, lertrummor, olika former av idiofoner (skallror, rainsticks, etcetera) (Bourg, 2005).

Filmmusik för science fiction bestod tidigt av elektroniska instrument (theremin, synthesizer) och var influerade eller till och med komponerade av många av de tidiga avant-gardé kompositörerna såsom Stockhausen, Varèse, Schoenberg och Cage. Detta började ändras runt 60-talet och vid 70-talet med framför allt succén med John Williams tonsättning av *Star Wars: A new hope* så började orkestral musik bli populär igen (May, 2020). Idag används både orkestral musik såväl som elektronisk musik, och även blandningar av dessa.

Då jag bestämt mig att skriva en tonsättning som kan framföras av en orkester utan krav på elektroniska instrument väljer jag att bestämma orkestreringen till följande uppsättning:

- Träblås
 - Flöjt
 - Ocarina
 - Oboe
 - Klarinett (Bb)
 - Fagott
- Bläckblås/brass
 - Horn (F)
 - Trumpet (Bb)
- Slagverk
 - Timpani
 - Bongo (2)
 - Rainstick
 - Bell Tree
 - Shaker
 - Marimba
- Stråk
 - Fioler 1
 - Fioler 2
 - Altfioler
 - Celli
 - Kontrabas

De etniska elementen i denna uppsättning består av ocarina, flöjt, rainstick, bell tree, shaker och en marimba. Alla dessa instrument har någon motsvarighet inom Mayakulturen och borde finnas redan eller vara relativt enkla att införskaffa i en modern orkester.

Skalor och tonal harmoni

ChatGPT gav mig informationen att mayakulturell musik ofta använder en pentatonisk skala med fem toner, antingen i moll eller dur form. Vidare nämner den poly rytmer och synkopering vilket också är viktiga delar som jag kan använda mig av för att skapa en kontrast mot den västerländska formen. Dessa påståenden kunde jag verifiera något genom att hitta referenser till en artikel där författaren visar att genom analyser av arkeologiska fynd av instrument och hur dessa varit stämde kunnat verifiera att det varit vanligast med pentatoniska skalor (Stacy, 2014).

Det spelas modern mayakulturell musik även idag som har förts fram genom generationerna i hela den regionen som tidigare utgjorde Mayariket och även om det inte är möjligt att verifiera deras korrekthet gentemot den gamla formen så har jag ändå valt att lyssna på dessa för att få lite referens och inspiration.

Minimalism

Då Mayamusiken verkar vara uppbyggda av relativt enkla melodier och mycket rytmik så väljer jag en liknande form för den orkestrala delen, genom att jobba minimalistiskt med små rörelser och borduntoner. Dessa borduntoner återfinns också inom Mayamusiken (Stacy, 2014) vilket ger en ytterligare koppling mellan den västerländska inledningen och den etniska Mayakulturella delen.

Komponerande

När jag har mina byggstenar är det dags att skriva musiken och min metod här är att komponera direkt till bild. Detta innebär att jag i mitt notprogram kan koppla detta till det animerade bildspelet direkt och allt eftersom jag skriver och testar musiken så kan jag se bildspelet spelandes synkroniserat i takt till musiken.

Under det första blocket som jag identifierade, när skeppet åker genom hyperrymden och kommer ut till planeten så väljer jag att jobba med en minimalistisk progression med att börja lågt med basen i stråk och bygga på lager på lager. Tempot är sakta, Lento med 60 slag i minuten och kontrabasen spelar väldigt svagt och bygger sakta upp med halvtoner för att

skapa en pulserande rytmik i takt . Min tanke här är att skapa en lugn introduktion där både visuellt och auditivt presenteras miljön och situationen för åskådaren.

Jag för sedan in fler element för att bygga på ljudbilden lager för lager, först cello i takt fem som jobbar tillsammans med basen och sedan resten av stråksektionen i takt 9 (0:30) med oktaveringar på fioler och altfiol. Fortfarande har vi en lugn uppbyggnad som sätter tempot och tonen när vi är i rymden och på bryggan visuellt i historien.

I takt 13 (0:48) så kommer den första rytmiska förändringen, där jag nu tar in träblåset och låter klarinetten stöta korta åttondelsnoter på fjärdedels-slagen medan fagotten dubblerar kontrabasens grundlinje. Sedan kommer Oboen in på åttondelsnoter och skapar en ännu tydligare pulserande rytmik. Jag gillar dessa minimalistiska rytmer med upprepande toner vilket bildar en bordun ton. Då detta dessutom är en del även i Mayakulturell musik så vill jag här forma en koppling mellan det västerländska minimalistiska och den mer exotiska Mayakulturella musiken som kommer att komma senare.

I takt 21 (1:20) så ändrar övre stråket (fiol 1, fiol 2 och altfiol) till tremolo medan låga stråket gör en åttondels stackato med accentuering på slag 1 och 3. Detta för att ännu mer förstärka pulsen och i samband med timpanin som kommer in här och spelar fjärdedelar på takt 1 och 3 så förstärks det ytterligare.

Hela det introducerande segmentet i takt 1-33 (0:00-2:08) har jag konstruerat för att jag tycker det skapar en mjuk introduktion och eskalering, och efter det sista lagret i takt 25 till och med 28 (1:36-1:51) där flöjten spelar stötande stackato så tar jag i takt 29 (1:52) i stället bort ett antal lager. Detta gör jag för att förbereda lyssnaren att något kommer att ske, en auditiv cliffhanger så att säga.

Efter att tre takter av den tystare delen spelas i takt 29-31 (1:52-2:03), så växlar stråket till tremolo för alla, och börjar att sänka dynamiken i ett de-crescendo på övre stråket (fiol 1, fiol 2 och altfiol). Samtidigt görs en gradvis temposänkning till femtio slag i minuten. Det jag vill göra härnäst är ju att börja ta in de etniska elementen från Mayakulturen och nu har jag skapat en förväntan på att något skall hända.

Genom att behålla klarinetten i samma stackato som den spelat hittills och låga stråket (celli och kontrabas) i tremolo gör sitt decrescendo så finns ändå element kvar, och nu börjar jag lite försiktigt att med hjälp av bongo, shaker och marimba slå ut samma pulserande rytm som det var tidigare.

De första indikationerna på att något skall växla rytmiskt sker i takt 34(2:12), när de två relativt kaotiska instrumenten rain-stick(takt 34, 2:12) och bell-tree(takt 35, 2:17) inträder. Detta skapar kaskader av ljud som inte har någon intern rytm utan glittrar mot den pulserande rytmiken från övriga slagverket och bildar en auditiv övergång som motsvaras i manuset av att miljön nu är på djungelplaneten.

I takt 36–40 (2:22-2:41) för jag gradvis in den nya mayakulturellt inspirerade rytmiken med synkoperande bongotrummor mot en pulserande marimba.

Värt att notera är också att jag återinför stråket från takt 38 (2:32) och tar in bläckblås för första gången med hjälp av hornen från takt 40 (2:41). Dessa lager förs in gradvis i ett crescendo över flera takter för att komma in mjukt i bakgrunden.

Nu är den mayainspirerade musiken ett faktum i stycket och åskådaren ser även visuellt att detta sammanfaller med att de nu befinner sig på den nya planeten.

I takt 48 (3:01) börjar flöjter göra drillar vilket är en teknik som vi återfinner i ett antal moderna Mayamusikers framförande för att efterlikna fågelkvitter i djungeln (Stacy, 2014). I just detta fall låter jag drillarna gå högre och högre för att bygga spänning mot avslöjandet av Mayatemplet i takt 51 (3:34). Åskådaren och de utforskande astronauterna rör sig ju genom tät djungel så dessa drillar är tänkta att ge en koppling på samma vis till djungelfåglar som det återfinns i Mayakulturell musik.

Denna gradvisa manipulering genom att lägga till och ta bort lager av instrumentation skapar i mitt tycke en mjuk övergång mellan de olika segmenten och förändrar rytmiken på ett naturligt sätt mellan den starka pulserande rytmen i den första delen och den synkoperande Mayainspirerade rytmiken i den mer etniska delen.

I de sista segmenten som leder upp efter avslöjandet av Mayatemplet och dialogen i scen 13 (takt 52-63, 3:39-4:31) och sedan mot slutet av berättelsen så väljer jag att ta bort slagverket efter att ha låtit rain-stick och bell-tree ljuda och klinga ut. Kvar är då helnoter i tremolo för stråket och halvnoter i bläck- och större delen av träblåset medan flöjterna gör stigande trills. Det allra sista så leder det upp till ett crescendo i takt 63 (4:32) och sedan de-crescendo ner till tystnad.

Resultat

Under arbetets gång har jag med hjälp av AI-stödda språkgenerativa verktyg producerat ett manus för en kort berättelse som utspelar sig i rymden och på en främmande planet. Jag har, med hjälp av AI-stödda bildgenerativa verktyg, skapat en animerad storyboard som illustrerar berättelsen över tid som en liten kortfilm.

Jag har tonsatt och komponerat ett musikstycke till denna storyboardade kortfilm, och använt en virtuell orkester som använder sig av AI-stöd för att skapa en så naturlig uppspelning som möjligt och allt har resulterat i en filmartefakt med musiken och storyboardet tillsammans samt ett partitur för orkester. Den totala spellängden på stycket blev 4 minuter och 55 sekunder.

Diskussion

Syftet med uppsatsen var att utforska hur användandet av AI-stödda verktyg påverkade och förändrade min gestaltande process under framställandet av en konstnärlig artefakt. Jag ville undersöka hur processen påverkades av att byta ut vissa traditionella verktyg och metoder, såsom att jobba med en manusförfattare och illustratör eller att använda mig av färdiga bilder och skapa kollage (moodboards), till att använda mig av generativa AI-stödda verktyg för att skapa text, manus och även visualiseringar och illustrationer. Jag kommer i detta kapitel gå igenom de viktigaste frågeställningarna jag kommit fram till under mitt användande av AI-stödda verktyg.

Trovärdighet hos verktygen

Jag fann att jag kan snabbt få resultat både i text och i bildform, men att jag hela tiden behöver vara kritisk mot det resultat jag får. Om jag får något som det AI-stödda verktyget anser är ett svar med fakta inbakat eller en representation av något jag frågat efter, så kan jag inte lita blint på verktyget. Jag behöver verifiera och analysera om det jag har fått är ett korrekt resultat.

Det är särskilt viktigt att tänka på när jag använder mig av språkgenerativa LLM modeller för att analysera och generera svar på riktade frågor, så som jag gjorde när jag ville lära mig mer om Mayakulturens musikaliska beståndsdelar. I det fallet behövde jag leta ordentligt efter ifall den informationen jag fick ut var korrekt eller om det var en hallucination från det AI-stödda verktyget i sin strävan efter att generera ett trovärdigt svar.

Just denna strävan efter att alltid komma med ett svar är extra tydligt i ChatGPT, framför allt GPT-3.5 modellen (Ji, et al., 2023). GPT 4 har varningar inbyggda när jag försöker ställa frågor som går utanför vad modellen är tänkt att användas för. Exempel på detta kan ses när jag ställde frågor om Mayakulturell musik och harmonik då ChatGPT (GPT 4) inledde med

*AS AN AI LANGUAGE MODEL, I AM UNABLE TO DIRECTLY CREATE OR PROVIDE MUSICAL NOTATION
OR AUDIO FILES.*

Trots denna varning så producerade ChatGPT detaljerad information om den tonala harmonin och orkestreringen för Mayakulturell musik utan att ange några referenser. När jag sedan prövade att ställa samma fråga till Bing CoPilot, som också använder sig av GPT 4 modellen, så gav den inget svar alls då den inte ansåg att den hittade källor som var tydliga nog.

En annan fråga som jag stötte på uppstod när jag använde mig av bildgenerativa verktyg. Jag frågade efter Mayakulturella delar, kläder och smycken. Men hur korrekt var resultatet? Efter att ha jämfört med bilder från arkeologiska utgrävningar så verkar det som att resultatet snarare är ett hopkok av ett antal olika kulturer och etniciteter. Resultatet ser okej ut för en lekman, men är kanske inte representativt för en person från den kulturen. I detta arbete så bedömde jag det dock som bra nog, särskilt då dessa bilder endast skulle användas som inspiration och berättelsen utspelar sig på en annan planet. Om resultatet skulle användas för att representera mayakultur och presentera illustrationen som faktabaserad, så behöver resultatet stämmas av mot de arkeologiska fynd och forskning som är gjort på området.

Juridiska frågeställningar

När denna rapport skrivs under våren 2023 så verkar vi vara i ett uppsving av AI-utveckling. Trots att fältet funnits ända sedan 50-talet, så har det inte funnits datorkraft eller implementationer förrän nu. I dagsläget så översvämmas marknaden av nya verktyg och lösningar, och det känns som det utvecklas exponentiellt.

Samtidigt så hinner vare sig lagstiftning, samhälle, kultur eller sociala faktorer med att anpassa sig och kunna hantera utvecklingen. För att citera kaos-matematikern från filmen Jurassic Park (1993) ”They [scientists] were so preoccupied with whether or not they could that they didn't stop to think if they should.” så verkar utvecklarna vara mer intresserade av att få ut verktyg och modeller på marknaden och se vad som händer, än att göra de etiska och juridiska bedömningarna först. Har vi rättighet att använda det material vi tränar vår modell på? Är vårt verktyg verkligen moget att släppas fritt för alla?

Ytterligare frågor som dyker upp och som även behandlats i tidigare kapitel är vilket material som modellerna har tränats på och ifall vi har rätt att använda materialet för att träna modeller på? Ibland förekommer både signaturer och vattenmärkningar på de bilder som producerats från verktygen. Detta kan vara en indikation på att vi inte har rättighet att nyttja materialet, då vattenmärkningar framför allt läggs på för att förhindra olovligt användande.

När denna rapport skrivs så är det många pågående juridiska fall över hela världen där konstnärer, författare, skådespelare och andra kreatörer försöker få bort sina alster från att tränas med de här modellerna, alternativt få kompensation. Hur dessa fall kommer att bedömas kommer eventuellt ha stor påverkan på hur framtida modeller och verktyg tränas.

Användarvänlighet

För mig som kreatör så var det en ögonöppnare att använda mig av dessa verktyg i mitt gestaltande, det var enkelt att börja och att få snabba resultat. Jag kunde på kort tid få fram ett manus som var tillräckligt bra för att ha som utgångspunkt. Det var inte helt färdigt, men vilket manus är det i första revisionen? Minuter senare kunde jag få högkvalitativa illustrationer skapade av ett bildgenerativt AI-stött verktyg. Dessa bilder var inte perfekta, men var även de en bra startpunkt.

Trots en snabb start så blev det snart svårare och tungrott när det skulle förfinas. Särskilt det bildgenerativa verktyget jag valt visade sig svårt att styra med precision. Med tanke på att gränssnittet till stora delar baserades på att verktyget tolkade mitt språk för att identifiera viktade söktermer att basera en ny bild på, så är det kanske inte konstigt att det krävdes relativt många iterationer för att komma fram till något användbart.

Andra modeller kan ha mer möjligheter att detaljstyra viktning och vilka termer som används under framtagandet av referensbilder, men i många av dessa visar det sig att jag behövde göra referensillustrationer för att träna AI-modellen i vad jag sökte. Jag kan tycka att om jag behöver göra en illustration för att skapa en ny illustration, varför gör jag inte rätt illustration från början?

Representation och transparens

Hur representativa och inkluderande var de resultat jag fick från de AI-stödda verktygen? De bilder jag skapade av astronauter föreställde över lag vita män (när vi kunde se deras ansikten), och i de bilder där jag refererade Mayacivilisationen så gjordes som tidigare

nämnts något hopkok av ett antal kulturer och etniciteter. Hur representativt skulle detta vara för en person som kommer från denna kultur?

Eftersom det oftast inte går att få se hur modellen faktiskt tolkar mina frågor, och hur detta appliceras på det material som den är tränad på, så blir det väldigt svårt att kunna bedöma om den ger ett representativt och inkluderande svar. Om vi dessutom inte kan se vilket material som den är tränad på, hur kan vi då bedöma att materialet är neutralt viktat, samt representativt och inkluderande?

En stor fara som jag ser är att om materialet som modellen är tränad på har värderingar inlagt i form av urval och även kodifieringar som endast representerar en del av kulturella och sociala grupperingar, så kan fördomar och stereotyper förstärkas. Därför behöver vi få mer transparens i hur dessa modeller har tränats för att kunna göra en bedömning av hur representativt och inkluderande resultatet kommer att bli.

Tonsättningen

När jag hade tagit fram ett fungerande manus och ett bra bildmaterial till bildspelet så var det dags att skapa musik. Här hittade jag som jag tidigare beskrivit väldigt få verktyg som kunde hjälpa mig i komponerandet. Det enda som var vettigt enligt min personliga åsikt var just Noteperformer som jag också använde. Det verktyget är dock inte direkt kopplat till komponerandet utan snarare till framförandet, men det ger i alla fall en bättre uppskattning av hur stycket kommer att låta om det framförs av riktiga musiker i stället för en dator.

De AI-stödda verktyg för musikkomponerande som finns i skrivande stund är enligt min åsikt mer av en gimmick. Jag ser dock att det i framtiden kan utvecklas mer AI-stödda verktyg inuti de programvaror som vi redan använder för vårt komponerande. Om jag till exempel har skapat en melodi, så skulle ett AI-verktyg som tränats på olika former av harmonilära, eller varför inte kontrapunktregler, kunna hjälpa till att skapa olika kompletterande stämmor och harmonier till melodin. Vid tiden för rapportens författande så har jag inte kunnat finna några sådana fungerande verktyg, dock ser jag detta som en möjlig intressant utveckling som skulle kunna vara ett bra stöd vid komponering.

Slutsatser

För att återkoppla till mina frågeställningar, så var det absolut möjligt att ersätta de traditionella verktygen och metoderna med AI-stödda verktyg. Däremot var det inte utan att det uppstår andra frågeställningar och faktorer att beakta när dessa användes. Då verktygen är

lätta att komma i gång med och snabbt få resultat, så kan det vara lätt att glömma bort de frågeställningar som tidigare nämnts angående hur långt verktygen och resultaten de producerar går att lita på.

Verktygen är framför allt ett sätt att snabbt komma i gång med skapandet. Dock behöver användaren alltid vara på sin vakt och bedöma ifall resultaten som verktygen levererat stämmer överens med de värderingar och den representation som användaren förväntade sig.

Många av framför allt de språkgenerativa verktygen och modellerna är tänkta att vara ett gränssnitt mot en tjänst eller ett annat verktyg, inte ett verktyg i sig. Så om det förväntas att de AI-stödda verktygen skall bota cancer eller byta däck på en bil bara för att det skrivs en prompt att göra det, så är det nog inte så genomtänkt vad det är för verktyg som används. I bästa fall så kommer verktygen också att informera om att dessa frågor är utanför deras ramar.

Framtida forskning

AI-stödda verktyg är här för att stanna och kan vara en stor tillgång för produktiviteten. Men i vilken form kommer de att användas i framtiden och kommer detta att bidra eller motverka disinformation. Mer forskning behövs på hur träning av modeller kan göras icke-normativa och värderingsneutrala och även vilken påverkan modeller kan ha på exempelvis sociala och kulturella normer eller stereotyper.

Om vi nu får tillgång till hur och på vilken data som modellerna har tränats, hur kan vi göra en bedömning om hur inkluderande och representativt materialet är? Det kan vara gigantiska mängder data och kodifieringar som gjorts, och för en lekman så blir det omöjligt att kunna göra en sådan bedömning. Mer forskning på hur ett sådant material kan analyseras och bedömas eller graderas behövs och kanske skulle i framtiden kunna bli en form av certifiering eller skala på hur värderande/inkluderande träningsmaterialet är.

Eftersom dessa AI-stödda verktyg, framför allt språkgenerativa och bildgenerativa modeller som använts i detta arbete, kommit ut på marknaden väldigt nära den tidpunkt som denna rapport skrevs så är området väldigt nytt och relativt utforskat. Detta arbete var ett försök att med hjälp av en konstnärligt utforskande metod hitta frågeställningar som behöver belysas, och sådana frågeställningar hittades också. Men mer forskning skulle även behövas här och jag tror att just konstnärlig forskning kan vara en bra form för att hitta de frågeställningar som

inte nödvändigtsvis är direkt kopplad till de tekniska eller matematiska aspekterna av AI-stödda verktyg.

För att avsluta så vill jag säga att användandet av AI-stödda verktyg har varit väldigt lärorikt, både i att se hur snabbt jag kan komma fram med bra visuellt och språkligt material, men också att det behöver göras mer bedömningar och efterforskningar på vad som produceras. AI-stödda verktyg kan vara till stor hjälp men också ha potential att skada eller förvränga en process om de levererar material som inte är önskvärt.

Referenser

- Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzett, M., Caillon, A., . . . Frank, C. (2023). *MusicLM: Generating Music From Text*. Google Research.
- Anoyoha, R. (2017). The History of Artificial Intelligence. *Harward University - Special Edition on Artificial Intelligence*.
- Azlant, E. (1997). Screenwriting for the Early Silent Film: Forgotten Pioneers, 1897-1911. *Film History, Volume 9*, 228-256.
- Bourg, C. H. (2005). *Ancient Maya music now with sound*. LSU Master's Theses.
- Britannica, T. E. (den 29 March 2023). *Maya*. *Encyclopedia Britannica*. Hämtat från The Britannica: Maya. Encyclopedia Britannica.
- Canestrino, R., Magliocca, P., & Li, Y. (2022). The Impact of Language Diversity on Knowledge Sharing Within International University Research Teams: Evidence From TED Project. *Frontiers in Psychology*.
- Chang, H. (2008). *Autoethnography as Method*. Left Coast Press.
- Choudhary, S. (den 16 February 2019). *Design Thinking: Divergence and Convergence Cycles*. Hämtat från Medium.com: <https://medium.com/@i.shubhangich/design-thinking-divergence-and-convergence-cycles-3ce7a6f27815>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. Cornell University.
- Dubberly, H. (2004). *How do you design?* San Francisco: Dubberly Design Office.
- Elmansy, D. R. (den 09 02 2021). *The Double Diamond Design Thinking Process and How to Use it*. Hämtat från Designorate: <https://www.designorate.com/the-double-diamond-design-thinking-process-and-how-to-use-it/>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review* 114(4), 864–886.
- Exploring the Impact of Artificial Intelligence on the Creative Industry: Opportunities and Challenges*. (den 23 2 2023). Hämtat från Medium:

<https://medium.com/@hasaneslik/exploring-the-impact-of-artificial-intelligence-on-the-creative-industry-opportunities-and-3f6597cefd7>

Gress, J. (2014). *Visual Effects and Compositing*. New Riders Publishing.

Hannula, M., Suoranta, J., & Vadén, T. (2005). *Artistic Research - Theories, Methods and Practices*. Helsinki: Academy of Fine Arts.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., . . . Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys, Volume 55, Issue 12*, 1-38.

Jones, J. c., Mitchell, C. T., & Jones, T. E. (1992). *Design Methods*.

Katz, S. D. (1991). *Film directing shot by shot: visualizing from concept to screen*. Michael Wiese Productions.

Leavy, P. (2020). *Method meets art: arts-based research practice*. The Guilford Press.

Lewis, T. (den 11 August 2014). *Brainy Machines Need An Updated IQ Test, Experts Say*. Hämtat från LiveScience: <https://www.livescience.com/47296-turing-test-needs-an-update.html>

Li, C. (den 3 Juni 2020). *OpenAI's GPT-3 Language Model: A Technical Overview*. Hämtat från LambdaLabs: <https://lambdalabs.com/blog/demystifying-gpt-3>

Liliequist, E., Tubella, A. A., Danielsson, K., & Cocq, C. (2023). Beyond the Binary—Queering AI for an Inclusive . *Interactions*, 31-33.

May, A. (2020). *The Science of Sci-Fi Music*. Springer Cham.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners*. OpenAI.

Robertson, J., Hutton, D., Colburn, C., Harmansah, Ö., Kjellgren, E., Koontz, R., . . . Blackmun Visonà, M. (2021). *The History of Art: A Global View: Prehistory to the Present*. Thames & Hudson.

Selvi, J. (den 5 December 2002). *Exploring Prompt Injection Attacks*. Hämtat från NCCGroup: <https://research.nccgroup.com/2022/12/05/exploring-prompt-injection-attacks/>

- Silver, D., Huang, A., Maddison, C., & Guez, A. (2016). *Mastering the game of Go with deep neural networks and tree search*. Toronto: University of Toronto.
- Simon, H. A. (1996). *The Sciences of the Artificial*. Cambridge, Massachusetts: The MIT Press.
- Stacy, A. (2014). Of the Same Stuff as Gods: Musical Instruments among the Classic Maya. *The Collegiate Journal of Anthropology*.
- Turing, A. (1950). *Mind* 49, 433-460.
- Vincent, J. (den 15 February 2023). *Microsoft's Bing is an emotionally manipulative liar, and people love it*. Hämtat från The Verge:
<https://www.theverge.com/2023/2/15/23599072/microsoft-ai-bing-personality-conversations-spy-employees-webcams>
- Wallander, A. (u.d.). *Noteperformer*. Hämtat från Noteperformer:
<https://www.noteperformer.com/>
- Wikipedia. (den 10 Oktober 2022). *Vår tideräkning*. Hämtat från Wikipedia:
https://sv.wikipedia.org/wiki/V%C3%A5r_tider%C3%A4kning
- Wikipedia. (den 11 April 2023). *Erich von Däniken*. Hämtat från Wikipedia:
https://en.wikipedia.org/wiki/Erich_von_D%C3%A4niken
- Xiaojin, Z., Goldberg, A. B., Eldawy, M., Dyer, C. R., & Strock, B. (2007). *A Text-to-Picture Synthesis System for Augmenting Communication*. Madison: University of Wisconsin.
- Xu, L., Cuesta-Infante, A., Berti-Equille, L., & Veeramachaneni, K. (2022). *R&R: Metric-guided Adversarial Sentence Generation*. MIT LIDS.

Figurer

- Figur 1 /imagine a lovely sunset –q 5 –ar 3:2 –v 5 –s 750 16
- Figur 2 /imagine once upon a time, in a world filled with magic and mythical creatures, there lived a girl named ava. ava was an adventurous young girl who loved exploring the vast forests and mountains that surrounded her village. she lived with her pet dragon 19

Figur 3 /imagine an oil painting of darth vader in the style of leonardo da vinci. composition like the mona lisa. --ar 3:2 --s 750 --q 2 --v 5.1 --style raw.....	21
Figur 4 /imagine the deep space exploration ship emerged from the hyperspace jump with a jolt, its engines roaring as it came to a stop in the void of space. the green jungle planet expands beneath it as the ship enters orbit.....	28
Figur 5 /imagine the deep space exploration ship emerged from the hyperspace jump with a jolt, its engines roaring as it came to a stop in the void of space. the planet expands beneath it as the ship enters orbit. --q 5 --ar 3:2 --s 750.....	29
Figur 6 väldigt detaljerad prompt	31
Figur 7 samma prompt i midjourney version 5.....	32
Figur 8 /imagine a small shuttle exiting the hangar bays of a large spaceship --q 5 --ar 3:2 --s 750.....	33
Figur 9 /imagine a small landing craft exiting the hangar bays of a large spaceship into space. extreme long shot. 28mm. view from outside the ship. --q 5 --ar 3:2 --s 750	33
Figur 10 /imagine a small landing craft exiting the hangar bays of a large spaceship into space. extreme long shot. 28mm. view from outside the ship. --q 5 --ar 3:2 --s 750	34
Figur 11 /imagine a blue alien woman in Mayan outfit looking out from behind bushes. Extreme closeup on face, cinematic lighting. --q 2 --ar 3:2 --s 50--v 4.....	36
Figur 12 /imagine a blue alien woman in Mayan outfit looking out from behind bushes. Extreme closeup on face, cinematic lighting. --q 2 --ar 3:2 --s 50 --v 5.....	36

Appendix 1 – Manus

FADE IN

1 EXT. HYPERSPACE – CONTINUOUS

THE CELESTIAL SEEKER FLIES THROUGH HYPERSPACE.

2 INT. CELESTIAL SEEKER - BRIDGE - DAY

THE CREW IS BUSY AT THEIR STATIONS, PREPARING FOR THE END OF THE HYPERSPACE JUMP.

3 EXT. SPACE - CONTINUOUS

THE CELESTIAL SEEKER EMERGES FROM HYPERSPACE, REVEALING THE GREEN PLANET XG-472 IN THE DISTANCE.

4 INT. CELESTIAL SEEKER - BRIDGE - CONTINUOUS

CAPTAIN STARLING GAZES AT THE PLANET ON THE VIEWSCREEN, MARVELING AT ITS BEAUTY.

CAPTAIN STARLING

Prepare a landing party. We're going down there.

5 INT. CELESTIAL SEEKER - HANGAR BAY - DAY

THE LANDING PARTY, WEARING ASTRONAUT BIO-SUITS, BOARDS THE LANDING CRAFT.

6 EXT. CELESTIAL SEEKER - CONTINUOUS

THE LANDING CRAFT UNDOCKS FROM THE CELESTIAL SEEKER AND BEGINS ITS DESCENT TOWARDS XG-472.

7 EXT. XG-472 - ATMOSPHERE - DAY

THE LANDING CRAFT ENTERS THE THICK ATMOSPHERE, SURROUNDED BY VIBRANT GREEN.

8 EXT. XG-472 - JUNGLE CLEARING - DAY

THE LANDING CRAFT TOUCHES DOWN IN A JUNGLE CLEARING. THE HATCH OPENS, AND THE ASTRONAUTS EXIT THE CRAFT, TAKING IN THEIR LUSH SURROUNDINGS.

9 EXT. XG-472 - JUNGLE - DAY

THE ASTRONAUTS MOVE THROUGH THE JUNGLE, MARVELING AT THE ALIEN FLORA AND FAUNA.

10 EXT. XG-472 - JUNGLE - HIDDEN LOCATION - DAY

BLUE-SKINNED ALIENS FOLLOW THE ASTRONAUTS FROM A DISTANCE, QUIETLY OBSERVING THEM.

11 EXT. XG-472 - JUNGLE - DAY

THE ASTRONAUTS MOVE DEEPER INTO THE JUNGLE, THE SOUNDS OF STRANGE BIRDSONG ECHOING AROUND THEM.

12 EXT. XG-472 - JUNGLE CLEARING - DAY

THE ASTRONAUTS COME TO A CLEARING AND SEE A STRUCTURE IN THE DISTANCE.

ASTRONAUT 1

(whispers)

Look at that...

13 EXT. XG-472 - JUNGLE CLEARING - STRUCTURE - DAY

THE ASTRONAUTS APPROACH THE STRUCTURE AND WATCH IN AWE AS THEY REALIZE IT IS A MAYAN TEMPLE STRUCTURE.

CAPTAIN STARLING

(whispers)

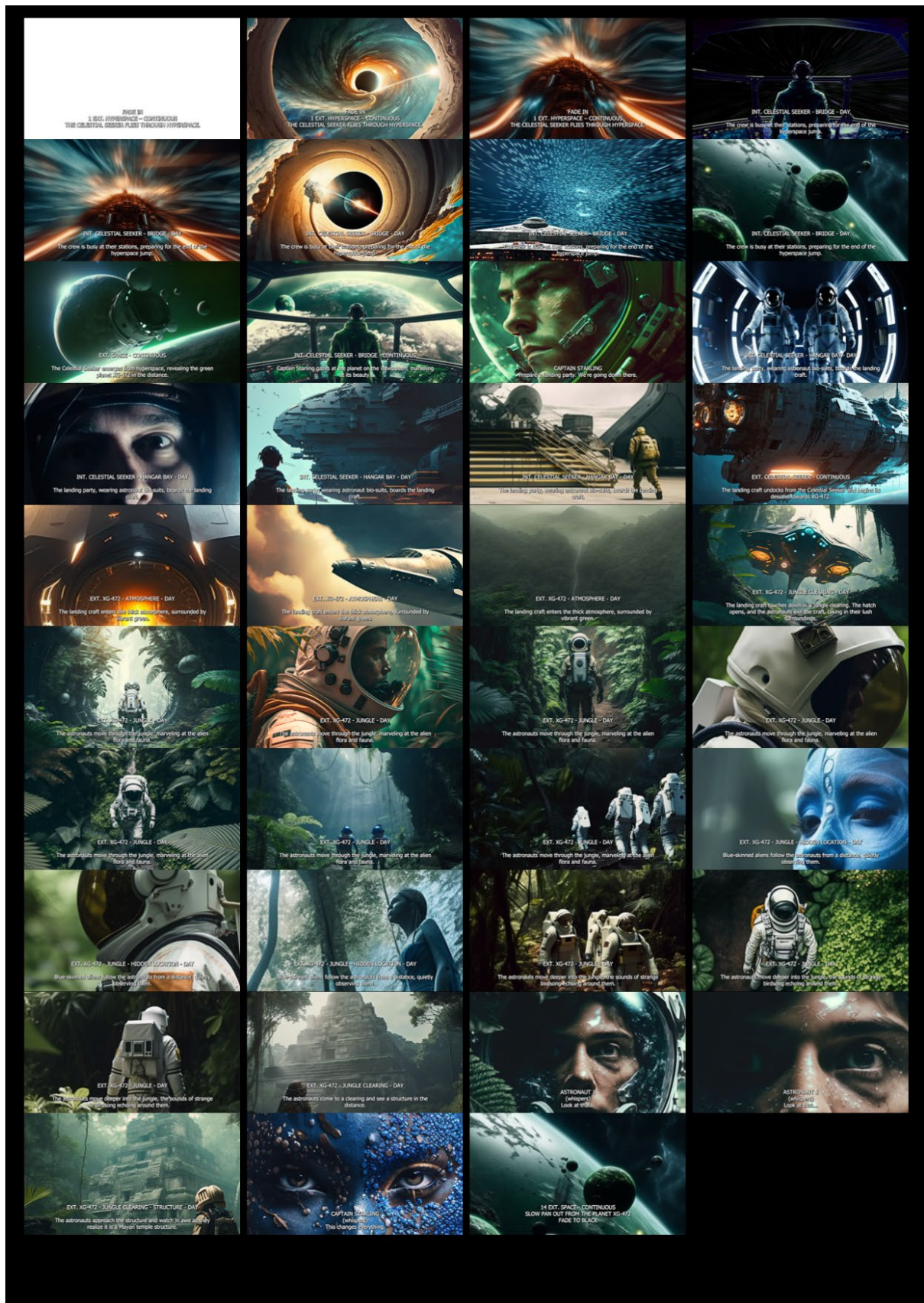
This changes everything.

14 EXT. SPACE - CONTINUOUS

SLOW PAN OUT FROM THE PLANET XG-472

FADE TO BLACK

Appendix 2 – Storyboard



Appendix 3 – Notbild

What a journey



A cinematic score for orchestra by
Henrik Bengtsson
2023

What a journey

Henrik Bengtsson

Lento ♩ = 60

This image shows a blank musical score for a symphony orchestra. The score is written in 4/4 time and includes dynamic markings such as *ppp* (pianissimo) and *p* (piano). The instruments are arranged in a standard orchestral layout, with woodwinds and brass in the upper staves, percussion in the middle, and strings in the lower staves. The score is divided into measures by vertical bar lines, and the instruments are labeled on the left side of the staves.

The instruments included in the score are:

- Flute
- Ocarina
- Oboe
- Clarinet in Bb
- Bassoon
- French Horn in F
- Trumpet in Bb
- Timpani
- Bongo 1
- Rainstick
- Shaker
- Marimba
- Violin I
- Violin II
- Viola
- Violoncello
- Double Bass

The score is a blank template, ready for a composer to write the musical notation.

10

Fl.

Oca.

Ob.

Cl. in B♭

Bsn

F. Hn in F

Tpt in B♭

Timp.

Bon. 1

Rain.

Shak.

Mar.

Vln I

Vln II

Vla

Vc.

D. B.

ppp

p

p

p

16

Fl.

Oca.

Ob.

Cl. in B♭

Bsn

F. Hn in F

Tpt in B♭

Timp.

Bon. 1

Rain.

Shak.

Mar.

Vln I

Vln II

Vla

Vc.

D. B.

The musical score for 'What a journey' begins at measure 16. The instrumentation includes a large woodwind section (Flute, Ocarina, Oboe, Clarinet in B-flat, Bassoon), brass (French Horn in F, Trumpet in B-flat), percussion (Timpani, Bongos 1, Rainsticks, Shakers, Maracas), and strings (Violin I, Violin II, Viola, Violoncello, Double Bass). The score is written in a single system with five measures. The woodwinds and strings play active parts, while the brass and percussion are mostly silent. The Oboe and Clarinet in B-flat play a melodic line, while the Bassoon and Violoncello play a bass line. The Violin I and II play a harmonic line. The Maracas play a rhythmic pattern. The Bongos 1, Rainsticks, and Shakers are silent.

21

Fl.

Oca.

Ob.

Cl. in B♭

Bsn.

F. Hn in F

Tpt in B♭

Tim.

Bon. 1

Rain.

Shak.

Mar.

Vln I

Vln II

Vla

Vc.

D. B.

mf

f

p

26

Fl.

Oca.

Ob.

Cl. in B♭

Bsn.

F. Hn in F

Tpt in B♭

Timp.

Bon. 1

Rain.

Shak.

Mar.

Vln I

Vln II

Vla

Vc.

D. B.

The musical score for 'What a journey' begins at measure 26. The instrumentation includes woodwinds (Flute, Ocarina, Oboe, Clarinet in B-flat, Bassoon), brass (French Horn in F, Trumpet in B-flat), percussion (Timpani, Bongos, Rain, Shakers, Maracas), and strings (Violin I, Violin II, Viola, Violoncello, Double Bass). The woodwinds and brass sections play melodic lines, while the percussion and strings provide a rhythmic foundation. The score is written for a full orchestra, with each instrument having its own staff. The key signature is one flat (B-flat), and the time signature is 4/4. The score is divided into measures, with measure 26 being the first measure shown. The instruments are listed on the left side of the page, and their corresponding staves are arranged in a vertical column. The Flute, Oboe, and Clarinet in B-flat staves are at the top, followed by the Bassoon, French Horn, and Trumpet. The percussion instruments (Timpani, Bongos, Rain, Shakers, Maracas) are in the middle, and the string instruments (Violin I, Violin II, Viola, Violoncello, Double Bass) are at the bottom. The score is written in a standard musical notation, with notes, rests, and other musical symbols used to represent the sound of each instrument. The Flute, Oboe, and Clarinet in B-flat staves show a melodic line of eighth notes. The Bassoon staff shows a sustained note. The French Horn and Trumpet staves show a melodic line of half notes. The Timpani staff shows a rhythmic pattern of eighth notes. The Bongos, Rain, and Shakers staves are empty. The Maracas staff is empty. The Violin I and Violin II staves show a melodic line of half notes. The Viola staff shows a melodic line of half notes. The Violoncello and Double Bass staves show a rhythmic pattern of eighth notes.

31 rit. $\text{♩} = 50$

Fl.

Oca.

Ob.

Cl. in B♭

Bsn

F. Hn in F

Tpt in B♭

Timp.

Bon. 1

Rain.

Shak.

Mar.

Vln I

Vln II

Vla

Vc.

D. B.

p

mf

p *f*

ppp

ppp

ppp

ppp

B. Tr.

36

Fl.

Oca.

Ob.

Cl. in B♭

Bsn.

F. Hn in F

Tpt in B♭

Timp.

Bon. 2

Bon. 1

B. Tr.

Shak.

Mar.

Vln I

Vln II

Vla

Vc.

D. B.

mp

f

ppp

39

Fl.

Oca.

Ob.

Cl. in B♭

Bsn

F. Hn in F

Tpt in B♭

Timp.

Bon. 2

Bon. 1

B. Tr.

Shak.

Mar.

Vln I

Vln II

Vla

Vc.

D. B.

ppp

Rain.

p

42

Fl.

Oca.

Ob.

Cl. in Bb

Bsn

F. Hn in F

Tpt in Bb

Timp.

Bon. 2

Bon. 1

To B. Tr.

B. Tr.

Shak.

Mar.

Vln I

Vln II

Vla

Vc.

D. B.

tr

ppp

ppp

ppp

ppp

p

p

ppp

ppp

ppp

p

45

This musical score page contains measures 45, 46, and 47. The instruments and their parts are as follows:

- Fl.**: Flute, rests in all three measures.
- Oca.**: Ocarina, plays a single note with a trill (*tr*) in measures 45 and 46, and a single note with a trill (*tr*) in measure 47.
- Ob.**: Oboe, plays a half note in measure 45, and a half note followed by a quarter note in measures 46 and 47. Dynamics: *p* in measure 46.
- Cl. in B♭**: Clarinet in B-flat, plays a half note in measure 45, and a half note followed by a quarter note in measures 46 and 47. Dynamics: *p* in measure 46.
- Bsn**: Bassoon, plays a half note in measure 45, and a half note followed by a quarter note in measures 46 and 47. Dynamics: *p* in measure 46.
- F. Hn in F**: French Horn in F, plays a half note in measure 45, and a half note followed by a quarter note in measures 46 and 47.
- Tpt in B♭**: Trumpet in B-flat, rests in all three measures.
- Timp.**: Timpani, rests in all three measures.
- Bon. 2**: Snare Drum 2, plays a continuous eighth-note pattern with accents in all three measures.
- Bon. 1**: Snare Drum 1, plays a continuous eighth-note pattern with accents in all three measures.
- B. Tr.**: Bass Drum, rests in all three measures.
- Shak.**: Shaker, plays a continuous eighth-note pattern with accents in all three measures.
- Mar.**: Maracas, plays a continuous eighth-note pattern in all three measures.
- Vln I**: Violin I, plays a half note in measure 45, and a half note followed by a quarter note in measures 46 and 47. Dynamics: *p* in measure 46.
- Vln II**: Violin II, plays a half note in measure 45, and a half note followed by a quarter note in measures 46 and 47. Dynamics: *p* in measure 46.
- Vla**: Viola, plays a half note in measure 45, and a half note followed by a quarter note in measures 46 and 47. Dynamics: *p* in measure 46.
- Vc.**: Violoncello, plays a half note in measure 45, and a half note followed by a quarter note in measures 46 and 47.
- D. B.**: Double Bass, plays a half note in measure 45, and a half note followed by a quarter note in measures 46 and 47.

[illegible]

[illegible]

[illegible]

61

Fl.

Oca.

Ob.

Cl. in Bb

Bsn

F. Hn in F

Tpt in Bb

Timp.

Bon. 2

Bon. 1

Rain.

B. Tr.

Shak.

Mar.

Vln I

Vln II

Vla

Vc.

D. B.

f

ppp

