

# AI-genererade parodiska låttexter

Hur upplevs parodilåttexterna som programmet Weird A.I. Yankovic maskingenererar och hur kan de förbättras?

## AI generated parody lyrics

How are the lyrics that the software Weird A.I. Yankovic produces regarded and how can they be improved?

Daniel Andersson

Institutionen för data-  
och systemvetenskap

Examensarbete 15 hp

Data- och systemvetenskap

Data- och systemvetenskap III (30 hp)

Höstterminen 2021

Handledare: Mirjam Palosaari Eladhari

English title: AI generated parodical lyrics



Stockholms  
universitet



# Abstract

This essay concerns AI generation of parody lyrics, based on an existing program called Weird A.I. Yankovic (hereafter WAIY) that is developed by Mark O. Riedl at Georgia Institute of Technology. It is a Python program that loads the powerful language models GPT-2 and XLNet and takes a short parody context description as well as an existing song text as input. While the syllable and rhyme structure are retained, the words are replaced with newly generated text based on the entered context. The result is new lyrics that can be sung to the original melody but are intended to gain comical qualities given the new content.

To get an overview of present and potential functionality, it has been investigated what approaches and functionality other AI programs that can generate lyrics use, and which of these could be implemented in WAIY. One of the identified potentials for improvement lies in additional training of the language models that generate the text content, as these are only generically trained in the original version. Accordingly, the main research question asks how the lyrics are affected by more specialized training of the language models.

In order to answer the research question, action research was performed where WAIY was further developed with new training of the GPT-2 language model on a few different specialized databases which are then evaluated. The evaluation concluded that a database of parody encyclopedia articles from Uncyclopedia gave the best results. The complete implementation of this version is called WAIY+.

To investigate whether WAIY+ is perceived as generating funnier lyrics, a survey was conducted in which participants were told to rate a number of lyrics' funniness and realness, in addition to answering background questions. The results of the survey showed that a significant difference was measured where lyrics generated by WAIY+ were perceived as more realistic than those of generated by the original version. WAIY+'s lyrics were also perceived funnier even though this difference was not statistically significant.

The survey also showed that a piece of real lyrics written by a famous comedy singer artist was regarded as considerably more fun and more realistic than any of the AI-generated lyrics. This indicates a great potential for future progress in the field, and several proposals for further research have been presented in this study.

Through the survey, a few differences between WAIY and WAIY+ have been confirmed and these indicate that some improvement can be achieved through the proposed implementation. The added code of WAIY+ is presented in its entirety in the essay to enable further work in the research field.

## Keywords

- Language technology
- Parody lyrics
- Perceived funniness
- AI
- Neural networks

# Sammanfattning

Denna uppsats behandlar AI-generering av parodilåttexter, med utgångspunkt i ett existerande program som heter Weird A.I. Yankovic (hädanefter WAIY) som utvecklats av Mark O. Riedl på Georgia Tech. Det är ett Pythonprogram som laddar de kraftfulla språkmodellerna GPT-2 och XLNet och tar en kort beskrivning av ett sammanhang för parodin samt en befintlig låttext som indata. Medan stavelse- och rimstrukturen behålls ersätts sedan orden med nygenererad text som bygger på det inmatade sammanhanget. Resultatet blir en ny låttext som kan sjungas i takt med den ursprungliga melodin men som förhoppningsvis låter komisk tack vare det nya innehållet.

Studieförfattaren har undersökt vilka angreppssätt och vilken funktionalitet som andra AI-program som kan generera låttexter utnyttjar och vilka av dessa som skulle kunna implementeras i WAIY. En identifierad förbättringspotential ligger i träningen av de språkmodeller som genererar textinnehållet, då dessa i originalprogrammet endast är generiskt tränade. Frågeställningen handlar därför om hur den upplevda kvaliteten förändras om språkmodellernas träning specialiseras.

För att svara på frågeställningen har aktionsforskning genomförts där WAIY vidareutvecklats med träning av GPT-2-språkmodellen på några olika specialiserade databaser som sedan utvärderats. Utvärderingen kom fram till att en databas över parodiska uppslagsverksartiklar från Uncyclopedia gav bäst resultat, och den kompletta implementationen av denna version benämns WAIY+.

För att undersöka om WAIY+ upplevs generera roligare låttexter så genomfördes en enkätundersökning, vars deltagare fick poängsätta ett antal låttexters rolighet och verklighetstrogenhet utöver bakgrundsfrågor. Resultatet av enkätundersökningen visade att en signifikant skillnad uppmättes där låttexter genererade av WAIY+ upplevdes mer verklighetstroga än originalversionens. Även roligheten verkade upplevas högre för WAIY+:'s låttexter även om denna skillnad ej var statistiskt signifikant.

Enkätundersökningen visade även att en verklig låttext skriven av en känd komedisångartist upplevdes betydligt roligare och mer verklighetstrogen än någon av de AI-genererade låttexterna. Detta tyder på en stor potential för framtida framsteg på området, och flertalet förslag för fortsatt forskning har presenterats i denna studie.

Genom enkätundersökningen har alltså vissa skillnader mellan de två jämförda programversionerna kunnat bekräftas och dessa tyder på att viss förbättring kan uppnås genom den föreslagna implementationen. Koden som lagts till i och med WAIY+ presenteras i sin helhet i uppsatsen för att möjliggöra fortsatt arbete inom området.

## Nyckelord

- Språkteknologi
- Parodilåtar
- Upplevd kvalitet
- AI
- Neurala nätverk

# Synopsis

## Bakgrund

Mark O. Riedl på Georgia Tech utvecklade ett Pythonprogram som använder modern språkteknologi för att generera parodilåttexter, och döpte programmet till Weird A.I. Yankovic efter en känd amerikansk komedisångare. Programmet laddar de kraftfulla språkmodellerna GPT-2 och XLNet och tar en kort beskrivning av ett sammanhang för parodin samt en befintlig låttext som indata. Medan stavelse- och rimstrukturen behålls ersätts sedan orden med nygenererad text som bygger på det inmatade sammanhanget. Resultatet blir en ny låttext som kan sjungas i takt med den ursprungliga melodin men som förhoppningsvis låter komisk på grund av innehållet.

## Problem

Weird A.I. Yankovic kombinerar två olika språkmodeller och några speciella tekniker för att åstadkomma en struktur med stavelser och rim som matchar ursprungslåten. I resultatet kan ibland den språkliga korrektheten vara lidande och de särskiljande egenskaperna hos paroditexter och låttexter kan saknas då språkmodellerna tränats generiskt på all typ av text.

## Frågeställning

- Hur förändras de parodilåttexter som WAIY genererar om språkmodellerna tränas på mer specialiserade data i form av låttexter eller parodiska texter?

Som beskrivet i problemformuleringen så lämnar utdatan i Weird A.I. Yankovic en del att önska vad gäller kvaliteten som parodilåttext. Då källkoden är allmänt tillgänglig så finns möjlighet att vidareutveckla programmet. En identifierad förbättringspotential ligger i träningen av de språkmodeller som genererar textinnehållet, då dessa i originalprogrammet endast är generiskt tränade. Då en mer specialiserad träning av språkmodellerna görs i syfte att förbättra den utdata som genereras så blir denna frågeställning central i det arbetet.

## Metod

Forskningsstrategin i denna studie är en kombination av aktionsforskning och undersökning. Aktionsforskning går ut på att lösa praktiska problem i verkligheten och samtidigt generera ny teoretisk kunskap. Arbetet är cykliskt och har målet att utvecklingen medför en användbar praktisk lösning. Efter aktionsforskningen så går strategin över till en undersökning för vilken en kvantitativ ansats med en enkät valts. Huvudsyftet är att ge ett konkret svar på huruvida den vidareutvecklade versionen av Weird A.I. Yankovic genererar roligare låttexter än originalprogrammet eller ej. Tack vare den kvantitativa ansatsen så kan denna studie dra konkreta slutsatser och svara på frågeställningen på ett tidseffektivt sätt.

## Resultat

Ett flertal möjliga förbättringsåtgärder som ej kunnat implementeras inom ramen för detta arbete har identifierats och presenterats. En ny version av Weird A.I. Yankovic har utvecklats där språkmodellerna tränas på en databas med parodiska upplagsverksartiklar från Uncyclopedia. Utdata

från denna version har presenterats, och viss skillnad har noterats. I en enkätundersökning fick deltagarna betygsätta roligheten och verklighetstrogenheten för en mängd olika lättexter genererade av båda programversionerna. En statistiskt signifikant skillnad syntes där den nya versionens lättexter ansågs något mer verklighetstroga. Även roligheten ansågs något högre i den nya versionen men denna skillnad låg inom den statistiska felmarginalen.

## Diskussion

Svaret på frågeställningen har delvis frångåtts genom att endast en av de två språkmodellerna som används av Weird A.I. Yankovic har kunnat tränas på de valda databaserna. Trots det har vissa skillnader kunnat bekräftas genom enkätundersökningen och dessa tyder på att viss förbättring kan uppnås genom den föreslagna implementationen. Samtliga underfrågor har besvarats i någon grad, med bland annat ett statistiskt säkert svar på att en verklig parodilåttext ansågs betydligt roligare och mer verklighetstrogen än någon av de AI-genererade lättexterna. Detta tyder på en stor potential för framtida framsteg på området, och flertalet förslag för fortsatt forskning har presenterats.

# Tack

Min handledare för engagerad assistans och feedback under examensarbetets gång.

Min mor, far, bror och faster för ert deltagande i pilotenkätundersökningen.

Alla som tog sig tid att svara på enkätundersökningen.

Och mest av allt min fru för ditt tålamod – speciellt under de två veckor då jag åkte ifrån er för att fokusera på skrivandet.

# Innehåll

<b>1</b>	<b>Introduktion</b>	<b>1</b>
1.1	Problemformulering	2
1.2	Frågeställning	3
<b>2</b>	<b>Bakgrund</b>	<b>5</b>
2.1	Weird A.I. Yankovic	5
2.2	Relaterad forskning	6
2.3	Studering av möjliga åtgärder	7
<b>3</b>	<b>Metod</b>	<b>12</b>
3.1	Forskningsstrategi	12
3.2	Litteraturgenomgång och tekniskt arbete	13
3.3	Applicering av implementation	15
3.4	Datainsamling	16
3.5	Dataanalys	18
<b>4</b>	<b>Resultat</b>	<b>20</b>
4.1	Implementation	20
4.1.1	Alternativ för träning av språkmodellen	20
4.1.2	Inför implementationstest	21
4.1.3	Resultat av implementationstest	23
4.1.4	Slutgiltig implementation	27
4.2	Enkätundersökning	30
4.2.1	Enkättagarnas demografiska data	31
4.2.2	Enkättagarnas bakgrundskunskap och engagemang	34
4.2.3	Jämförelse mellan WAIY+ och WAIY	37
4.2.4	Upplevd rolighet för verkliga parodiåttexter	40
<b>5</b>	<b>Diskussion</b>	<b>44</b>
5.1	Svar på frågeställningen	44
5.2	Begränsningar och framtida studier	46
5.3	Slutsats	47
	<b>Referenser</b>	<b>49</b>
	<b>Bilaga A – Enkät</b>	<b>51</b>
	<b>Bilaga B – Nyimplementerad kod</b>	<b>61</b>
	<b>Bilaga C – Reflektionsdokument</b>	<b>66</b>



# Figurer

Figur 1 Weird Al Yankovic under en turné 2010.....	1
Figur 2 WAIYs grafiska gränssnitt.....	5
Figur 3 Indata till To sing like a mockingbird för två låttexttrader. (Gatti et al. 2017, s.300) .....	8
Figur 4 Jämförelse av längd på de maskingenererade låttexterna och en testmängd med manuella låttexter. (Lau 2021, s.26) .....	9
Figur 5 Översikt över den låttextgenererande algoritmen i MABLE. (Singh 2018, s.15) .....	10
Figur 6 Översikt över arbetet med design och utveckling av artefakt enligt Johannesson och Perjons (2014, s.118).....	13
Figur 7 Översikt över WAIY:s algoritm för parodilåtttextgenerering (Riedl 2020, s.5).....	14
Figur 8 Exempel på utdatapresentationen från WAIY, med resultatet av indatan från Figur 2. ....	23
Figur 9 Könsfördelning för de respondenter vars enkätsvar analyserats.....	31
Figur 10 Åldersfördelning för de respondenter vars enkätsvar analyserats .....	32
Figur 11 Modersmålsfördelning för de respondenter vars enkätsvar analyserats .....	33
Figur 12 Fördelning av språkfärdighetsnivåer i engelska för de respondenter vars enkätsvar analyserats .....	33
Figur 13 Fördelning av boendeland för de respondenter vars enkätsvar analyserats .....	34
Figur 14 Kännedom om originallåtarna som enkätens parodilåttexter baseras på .....	35
Figur 15 Andel av enkätdeltagarna som lyssnade eller mindes melodin när roligheten bedömdes .....	36
Figur 16 Lådagram-jämförelse mellan upplevd rolighet för låttexter från WAIY+ respektive WAIY .....	37
Figur 17 Histogram över upplevd rolighet för låttexter från WAIY+ .....	37
Figur 18 Histogram över upplevd rolighet för låttexter från WAIY .....	37
Figur 19 Lådagram-jämförelse mellan upplevd verklighetstrogenhet för låttexter från WAIY+ resp. WAIY .....	39
Figur 20 Histogram över upplevd verklighetstrogenhet för låttexter från WAIY+.....	39
Figur 21 Histogram över upplevd verklighetstrogenhet för låttexter från WAIY .....	39
Figur 22 Upplevd rolighet för den äkta parodilåttexten till Smelling like Nirvana.....	41
Figur 23 Upplevd verklighetstrogenhet för den äkta parodilåttexten till Smelling like Nirvana .....	41
Figur 24 Jämförelse av upplevd rolighet för varje låttext i enkäten.....	43
Figur 25 Jämförelse av upplevd verklighetstrogenhet för varje låttext i enkäten.....	43

# Tabeller

Tabell 1 Sammanfattningar av Weird Al Yankovic parodiversioner av låtarna som används i indatan	15
Tabell 2 WAIYs stavelseanalys av de första fyra raderna i låttexten till Another one bites the dust ...	22
Tabell 3 Analys av det faktiska stavelseantalet och det av WAIY uträknade stavelseantalet.....	22
Tabell 4 Exempel över hur originallåttexten modifierats för att användas som indata .....	22
Tabell 5 Resultat av implementation av de tre alternativen på Another one bites the dust.....	24
Tabell 6 Resultat av implementation av de tre alternativen på Smelling like teen spirit .....	25
Tabell 7 Resultat av implementation av de tre alternativen på Lose yourself.....	26
Tabell 8 Alternativ 3 med tillhörande utdata, med gjorda justeringar inför enkäten .....	28
Tabell 9 Originalprogrammets utdata med gjorda justeringar inför enkäten .....	29
Tabell 10 Översikt över antalet fel vad gäller antal stavelser i utdatan som ska ingå i enkäten .....	30
Tabell 11 Skillnad i upplevd rolighet respektive verklighetstrogenhet mellan kvinnor och män. ....	31
Tabell 12 Skillnad i upplevd rolighet respektive verklighetstrogenhet beroende på ålder.....	32
Tabell 13 Skillnad i upplevd rolighet respektive verklighetstrogenhet beroende på språkfärdighetsnivå. .....	34
Tabell 14 Antal enkätmedtagare som gett respektive svar vad gäller hur roligheten bedömdes.....	36
Tabell 15 Paired samples t-test på den upplevda roligheten för låttexterna från WAIY+ och WAIY ..	38
Tabell 16 Paired samples t-test på skillnaden i den upplevda verklighetstrogenheten för WAIY+ och WAIY .....	40
Tabell 17 Paired samples t-test på skillnaden i den upplevda roligheten .....	42
Tabell 18 Paired samples t-test på skillnaden i den upplevda verklighetstrogenheten.....	42

# Förkortningar och termer

AI – Artificiell intelligens: särskilt i formen av en beståndsdel i ett datorprogram  
MIT license – En upphovsrättslicensmall med ursprung på Massachusetts Institute of Technology  
Mtoe – Mother tongue or equivalent (vad gäller språkfärdighet)  
Python – Ett programmeringsspråk  
Språkmodell – En programvara med artificiell intelligens för generering och/eller analys av text/tal  
TSLM – To Sing Like a Mockingbird, en AI utvecklad av Gatti et al. (2017)  
Vocaloid – Ett datorprogram som genererar syntetisk sång  
WAIY – Weird A.I. Yankovic (Mark O. Riedls program som genererar parodilåttexter)  
WAIY+ – Den i studien vidareutvecklade versionen av WAIY

# 1 Introduktion

”Weird” Al Yankovic är en amerikansk musiker som är känd som världens bäst säljande komedisångare och har fått många utmärkelser (Yankovic u.å). Som en hyllning till honom valde Mark O. Riedl (2020), en forskare på Georgia Tech, att döpa ett Pythonprogram som han utvecklade 2020 till namnet Weird A.I. Yankovic (benämns hädanefter WAIY). Programmet vars källkod är fritt tillgänglig och presenteras i en artikel i arXiv har en algoritm som tar en låttext som indata och producerar en parodi på låttexten med samma antal stavelser per rad och rim på samma ställen. Detta gjordes med anpassade språkmodeller som fungerar som neurala nätverk för språkgenerering både framlänges och baklänges, och en specialiserad samplingsstrategi för att styra antalet stavelser och rim. WAIY blev uppmärksammat i media i samband med att Riedl lagt upp en video med en parodiversion av Michael Jacksons låt Beat It tillsammans med instrumentellt originalljud, som togs ner av Twitter när en skivbolagsparaplyorganisation krävde det och hävdade upphovsrättsintrång (Canales 2020).



*Figur 1 Weird Al Yankovic under en turné 2010*

## **Tidigare forskning**

Till författarens kännedom har inga studier gjorts på Riedls (2020) tekniska lösning vad gäller hur programmet skulle kunna förändras och vilka effekter detta skulle ha på den upplevda kvaliteten på utdatan. Därför var det av författarens åsikt att det vore intressant att undersöka denna fråga som därmed utgör detta examensarbets huvudfokus. För att djupare förstå vilka typer av åtgärder och angreppssätt som skulle kunna öka den upplevda kvaliteten på de parodilåttexter som produceras, så har ett antal artiklar studerats. Artiklarna rör AI-genererad låttext och/eller poesi, där påträffat med

fokus på parodi. Syftet har varit att erhålla kunskap kring potentiella förbättringsåtgärder och teorin bakom AI-genererad låttext.

Tidigare relaterade studier utforskar språkteknologiska lösningar för att procedurellt skapa varierande, språkligt korrekta och kontextuellt relevanta texter. De tekniska lösningarna varierar och kan exempelvis involvera större befintliga databaser av låttexter eller av helt andra typer av texter som nyhetsartiklar. Dessa tekniker utforskas och presenteras för att ge en översiktlig bild av vilka åtgärder som skulle kunna implementeras i WAIY. Fokus ligger på hur språkmodellerna tränas inför textgenereringen vilket implementeras i en vidareutvecklad version av programmet, som benämns WAIY+.

## Mäta upplevd kvalitet

Ett programs kvalitet definieras och mäts som det genomsnittliga ”betyg” vuxna personer i allmänhet skulle ge i en enkät som svar på frågan *How funny do you think these parody lyrics are?* där svarsalternativen är en skala från 1 till 9. En förbättrad version i och med WAIY+ skulle alltså producera parodilåttexter som till en statistiskt signifikant grad erhåller högre betyg i en enkätundersökning än originalprogrammets utdata.

Syftet är att mäta kvaliteten på låttexterna som de parodier de är – då Riedl (2020) själv konkluderar att WAIY:s lyckade resultat sannolikt hänger ihop med parodins natur att den inte efterliknar normalt språk vilket inte heller har eftersträvat. Att parodiska textrader inte behöver förmedla meningsfullt innehåll eller vara språkligt konsekventa bör dessutom innebära ett tekniskt mindre utmanande arbete än för andra textgenrer. Därmed lämpar sig uppgiftens omfattning för ett examensarbete på kandidatnivå.

## Syfte

I detta examensarbete utvecklas en ny version av WAIY som benämns WAIY+. Såväl de exakta justeringar som gjorts som vilka effekter man kan se på de låttexter som genereras presenteras i detalj, för att ge läsaren möjlighet att bilda sig en uppfattning om respektive versions förmåga att producera parodilåttexter. En enkätundersökning genomförs för att se om det går att fastslå vad en tänkbar målgrupp anser om kvaliteten på parodilåttexterna, jämfört mellan de två versionerna samt jämfört med verkliga analogt skrivna parodilåttexter. Vad gäller studiens omfattning så är arbetet fokuserat på WAIY och teorin i övriga refererade artiklar som alla rör AI-genererad låttext eller poesi. De föreslagna förändringarna i den tekniska implementationen beskrivs och ny källkod presenteras i sin helhet.

Det primära syftet är att åstadkomma förhöjd upplevd kvalitet, med avseende på hur roliga de AI-genererade parodilåttexterna upplevs. Genom en förändrad upplevd kvalitet och däri bakomliggande orsaker och processer förväntas utvecklingen av WAIY+ bidra till ökad kunskap inom AI och dess applikationer.

## 1.1 Problemformulering

WAIY producerar parodilåttexter genom att låta två olika generiskt förtränade språkmodeller gissa en imaginär fortsättning på en av användaren inmatad kontextuellt inledande rad. De genererade textraderna efterföljer den stavelse- och rimstruktur som utlästs från en originallåttext som också ingår i indatan. Resultatet kan ibland innehålla bristande språklig koppling mellan textraderna som genererats av de olika språkmodellerna, en varierande stil då språkmodellernas förträning varit

generisk samt textrader som saknar sångens eller parodins särskiljande egenskaper från annan text. Då det finns andra forskare som också utvecklat AI-system för låttexter inklusive specifikt parodilåttexter med andra tekniska lösningar än Riedl så tros det finnas en potential att med avseende på dessa brister förbättra genom att vidareutveckla WAIY med en språkmodell som är mer specialiserat tränad.

Samtidigt har inte användarens subjektiva upplevelse av utdatan använts i utvecklingen av WAIY vilket ytterligare förstärker tesen om att det kan finnas utrymme för förbättring, och det saknas kunskap vad gäller denna upplevelse.

## 1.2 Frågeställning

### Huvudfrågeställning:

- Hur förändras de parodilåttexter som WAIY genererar om språkmodellerna tränas på mer specialiserade data i form av låttexter eller parodiska texter?

Som beskrivet i problemformuleringen så lämnar utdatan i WAIY en del att önska vad gäller kvalitet som parodilåttext. Då källkoden är allmänt tillgänglig så finns möjlighet att vidareutveckla programmet. En identifierad förbättringspotential ligger i träningen av de språkmodeller som genererar textinnehållet, då dessa i originalprogrammet endast är generiskt tränade. Då en mer specialiserad träning av språkmodellerna görs i syfte att förbättra den utdata som genereras så blir denna frågeställning central i det arbetet. Valet att begränsa frågeställningen till denna specifika förbättringsåtgärd härstammar från slutsatserna som dragits i den första, teoretiska delen av forskningsarbetet, vilken i sin tur utgår ifrån underfråga 1 nedan.

### Underfrågor:

1. Vilka tekniska lösningar skulle kunna implementeras i WAIY+?
2. Vad anser svarspersoner i en enkätundersökning om dessa parodilåttexter med avseende på hur roliga de är, jämfört med parodilåttexter som originalprogrammet genererat?
3. Vad anser svarspersoner i en enkätundersökning om datorgenererade parodilåttexter med avseende på hur roliga de är, jämfört med parodilåttexter som Weird AI Yankovic manuellt skrivit och publicerat?

Inför det tekniska arbetet med den vidareutvecklade versionen WAIY+ så behövs kunskap i vilka angreppssätt som liknande programvaror använt för att maximera kvaliteten i respektive utdata. Därför studeras tidigare forskning på området med fokus på tekniska lösningar i den första delen av forskningsarbetet, och denna process utgör förarbete inför underfråga 1. När kunskap erhållits kring dessa tekniska lösningar så analyseras genomförbarheten hos dessa i förhållande till strukturen i WAIY. Underfråga 1 används då som utgångspunkt i denna analys. Denna underfråga hör till diagnostik- och planeringsfaserna i det cykliska forskningsarbetet (Johannesson & Perjons 2014) som beskrivs i detalj i metodkapitlet.

En implementation i form av WAIY+ utvecklas där språkmodellerna tränas på specialiserade träningsdata i form av en textdatabas med lämpligt innehåll. Originalprogrammet kompletteras med kod för att läsa in vald databas, tokenisera textinnehållet samt finjustera de förtränade språkmodellerna genom kompletterande träning på denna tokeniserade data. För att konkretisera svaret på huvudfrågeställningen så görs ett försök att påvisa huruvida den upplevda kvaliteten förändrats i och med det vidareutvecklade WAIY+. Deltagare i en enkätundersökning får då svara på hur roliga de

anser att ett urval av färdiga parodilåttexter från två olika versioner av programmet är. Svaren från denna enkätundersökning analyseras då utifrån underfråga 2.

I samma enkätundersökning presenteras en äkta parodilåttext som Weird AI Yankovic skrivit, för att kunna jämföra den upplevda kvaliteten med de AI-genererade låttexterna. Detta utgör dels ett referensvärde för att undersöka huruvida den upplevda kvaliteten hos AI-genererade parodilåttexter kan bedömas på liknande sätt som äkta parodilåttexter, och ger också en insikt i det egentliga behovet av åtgärder som leder till ökad kvalitet i utdatan från WAIY. Denna jämförelse har formulerats som underfråga 3, vars svar precis som svaret på underfråga 2 utgörs av en statistisk analys. Den statistiska analysens centrala del utgörs av T-tester som kan ge konkreta svar på underfråga 2 och 3 (Denscombe 2014).

Såsom evident utifrån beskrivningen ovan så är underfrågorna alltså intimt relaterade till huvudfrågeställningen och utgör stommen i metod- och forskningsstrategivalet.

## 2 Bakgrund

Den artikel som beskriver programmet WAIY som utgör kärnan i detta examensarbete har studerats i detalj tillsammans med några andra relaterade studier, inom det aktuella forskningsområdet AI och maskininlärning. Dessa relaterade studier presenterar program som precis som WAIY med hjälp av AI genererar låttexter, men inte nödvändigtvis parodiska sådana. Denna genomgång görs som förarbete för att senare kunna besvara underfråga 1:

1. Vilka tekniska lösningar skulle kunna implementeras i WAIY+?

### 2.1 Weird A.I. Yankovic

WAIY producerar parodilåttexter utifrån en av användaren inmatad originallåttext tillsammans med en kort kontextuell inledning, som inte inkluderas i den producerade låttexten (Riedl 2020). Programmet börjar med att analysera originallåttextens stavelsestruktur, d.v.s. antal stavelser per rad, samt finna vilka rader som rimmar med varandra. Därefter genererar språkmodellerna låttexter en och en med målet att finna maximalt kontextuellt relevanta rader i förhållande till den korta inledningen samt de rader som redan genererats, från och med andra raden. Programmet använder sig av ett lexikon för approximativa rim, som programmet antingen laddar hem från en molntjänst eller skapar självt. Programmet kan sedan även avgöra om två rim är perfekta rim, vilket man som användare har möjlighet att kräva.

Språkmodellerna GPT-2 samt XLNet laddas in i som förtränade på stora mängder generiska data (Riedl 2020). GPT-2 används för att generera rader som varken behöver sluta med rim eller punkt medan XLNet används för att generera rader när det finns krav på att raden måste sluta med ett rim eller som en komplett mening. Kombinationen av två olika språkmodeller sägs ha valts då GPT-2 är den språkmodellen på marknaden som ger bäst resultat men den kan ej enkelt anpassas för att ställa krav på stavelse- eller rimstruktur. XLNet kan å sin sida anpassas för olika sådana krav och till exempel generera text från slutet och bakåt.

WAIY har förutom själva algoritmen för att skapa sångtexter ett enkelt grafiskt gränssnitt, ett avancerat läge där användaren ges möjlighet att påverka hur algoritmer jobbar med textgenereringen, samt en funktion för att producera en karaokevideo utifrån en ljudfil med en instrumentell version av originallåten samt data med timingen av varje låttextrad (Riedl 2020).

**Context prompt:**

**Enter lyrics here:**

```
My small rhino
Yeah, hey
I'm worse at what I do best
And for this gift I do feel blessed
Our little group always been
And always will until the end
```

**Use near rhymes?**

**Recontextualize?**

Figur 2 WAIYs grafiska gränssnitt.



## 2.2 Relaterad forskning

Det finns många studier som är gjorda på AI-system som genererar text och även ganska många på system som är specialiserade på att producera låttexter och/eller poesi. För denna studies syften så har forskning på de tekniska lösningarna vad gäller generering av låttexter och/eller dikter fokuserats på.

Gatti et al (2017) utvecklade ett program som liknar WAIY då det också automatiskt producerar parodier av befintliga låttexter, men med en helt annat angreppssätt. I stället för ett neuralt nätverk som bearbetar en ren textfil så kräver Gatti et als implementering mer manuellt arbete med indata där även själva musiken (tonerna, timingen etcetera) kodas tillsammans med texten, samtidigt som utdata kopierar textsnittar från verkliga databaser med nyhetsartiklar mm, i stället för att automatiskt generera ord för ord. Det skulle kunna vara intressant att jämföra resultatet av deras program med Riedls, men till författarens kännedom är inte deras programvara eller tillhörande källkod allmänt tillgänglig. Utifrån artikeln så har delar av implementeringen analyserats för att finna lösningar som kan tänkas vara användbara i en vidareutveckling av Riedls program.

Lau (2021) jämförde manuellt skriven musik samt maskingenererade låttexter med hjälp av verktyget *GPT-2 transformer architecture*. Lau lät då verktyget lära sig av en större databas med äkta engelska låttexter innan det användes för att skapa nya texter. Slutsatsen var att variationen hos de maskingenererade låttexterna berodde på mängden musik som tränades på, och författaren ansåg att det krävdes minst 15000 låtar för en rimlig variation i utdata. Det konstaterades också att de maskingenererade låttexterna innehöll betydligt fler ord än källmaterialet. Huruvida denna typ av källmaterial kan vara till användning även för skapandet av paroditexter har undersökts. Detta skulle till exempel kunna vara fallet genom medveten randomisering av ordvalen, eller en blandning av indata från äkta låttexter samt orelaterat material såsom nyhetsartiklar.

Oliveira (2012) utvecklade ett program kallat PoeTryMe för automatisk generering av poesi på portugisiska, som använder sig av återkommande mönster i riktig portugisisk poesi. Innan dess år 2007 så har Oliveira (2015) utvecklat programmet Tra-La-Lyrics för generering av låttexter, också på portugisiska. I denna artikel från 2015 beskriver han hur dessa två program sammanfogats för att åstadkomma låttexter med semantisk hänsyn vilket originalversionen av Tra-La-Lyrics inte klarade av. Det uppgraderade programmet som benämns *Tra-La-Lyrics 2.0* har sedan konstaterats producera låttexter som upplevs mer språkligt korrekt och meningsfulla än det gamla programmet, även om det sägs finnas gott om ytterligare förbättringspotential. De strategier som Oliveira använder i denna vidareutveckling har evaluerats ifall de skulle kunna implementeras i studiens vidareutveckling av WAIY. Det kan dock konstateras att de mekanismer som används är konstruerade för portugisiska samt spanska, och ingen del av funktionaliteten som beskrivs i artikeln och kan appliceras för generering av engelsk text har identifierats.

I en annan artikel skriver Oliveira (2020) om hur en modell för automatiserad låttextskrivning kan användas som hjälpmedel vid poesiskrivande, varför det förstås kan tänkas vara användbart även för parodilåtttextgenerering. Det vill säga inte bara vid generering av kompletta låttexter utan även som ett hjälpmedel där en låtskrivare justerar och färdigställer parodilåtttexter som har maskingenererats. Singh (2018) har å sin sida visat hur automatisk generering av poesi kan användas som hjälpmedel för låtskrivare att skriva sina egna låtar, inte minst för att lära sig de förmågor man behöver. Detta tyder på att en lyckad vidareutveckling i enlighet med examensarbetets mål kan komma att visa sig användbar inte bara för att generera färdiga låttexter att roa läsaren eller lyssnaren, utan även hjälpa låtskrivare bli bättre i sin kultur- eller yrkesutövning.

Singh (2018) har utvecklat ett AI-system namngett MABLE som är avsett för att generera låttexter för ballader som har en sammanhängande berättelse. Den bygger på MEXICA, ett annat AI-system som genererar berättelser. De viktigaste elementen i berättelsen identifieras och utgör efter anpassning till rytmen varannan textrad i balladlåttexten. Återstående textrader fylls ut med nyproducerad låttext med rim för att komplettera låttypens beståndsdelar. Den nyproducerade texten är tränad på en databas över kärlekssånger och kopierar föregående rads struktur. Dessa textrader som kompletterar berättelseraderna genereras genom att anropa Markov-modellen upprepade gånger och välja ut de rader som bedöms ha närmast rim, och sedan närmast korrekt antal stavelser. Systemet använder därefter attitydanalys för att bedöma textens attityd (till exempel positiv, neutral, dålig) och på så sätt åstadkomma en konsekvent varierande attityd. Slutligen korrigeras pronomenen för att få rätt perspektiv.

Singh (2018) pekar på systemets användbarhet speciellt i form av ett hjälpmedel för låttextrskrivare i deras arbete med att skriva egna låttexter, vilket också bör kunna sägas om andra program för maskingenererade låttexter inklusive denna studies slutprodukt i det fall att kvaliteten skulle anses vara tillfredsställande. MABLE har analyserats för att undersöka om det finns beståndsdelar som kan vara till fördel att implementera i WAIY+.

## 2.3 Studerande av möjliga åtgärder

De artiklar som tagits upp ovan har analyserats i detalj för att fastslå vilka åtgärder som skulle kunna implementeras i en vidareutvecklad version av WAIY. Analysen görs för att erhålla kunskap som kan användas för att besvara underfråga 1:

1. Vilka tekniska lösningar skulle kunna implementeras i WAIY+?

Tre av artiklarna från den relaterade forskningen har konstaterats innehålla beskrivningar av funktionalitet som kan användas i ett program för generering av låttexter, varav en artikel specifikt tar upp generering av parodilåttexter. De tekniska lösningar som har identifierats i dessa tre artiklar beskrivs och genomförbarheten diskuteras grundläggande.

### **To sing like a mockingbird**

Algoritmen som Gatti et al. (2017) utvecklat använder sig av indata som är mer innehållsrik än WAIY, med information om varje stavelses timing och korresponderande ton. Detta innebär att utdatan kopplat till en Vocaloid kan producera en komplett parodilåt med musik och syntetisk sång. Då parodilåttextens kvalitet så som definierad i frågeställningen ej avgörs av hur parodilåttexten presenteras så är denna funktion ej relevant att implementera i WAIY.

```

<song filename="AHARDDAY.m2a">
<key time="0">G major</key>
<chorus>
  <verse pvers="1">
    <token time="5040" orig-note="B" tone="3" interval="210">IT</token>
    <token time="5050" orig-note="B" tone="3" interval="210">'S </token>
    <token time="5280" orig-note="C" tone="4" interval="210">BEEN </token>
    <token time="5520" orig-note="B" tone="3" interval="210">A </token>
    <token time="5760" orig-note="D" tone="5" interval="810">HARD </token>
    <token time="6720" orig-note="D" tone="5" interval="570">DAY</token>
    <token time="6730" orig-note="D" tone="5" interval="570">'S </token>
    <token time="7440" orig-note="D" tone="5" interval="690">NIGHT</token>
  </verse>
  <verse pvers="2">
    <token time="8880" orig-note="C" tone="4" interval="212">AND </token>
    <token time="9120" orig-note="D" tone="5" interval="210">I</token>
    <token time="9130" orig-note="D" tone="5" interval="210">'VE </token>
    <token time="9360" orig-note="C" tone="4" interval="210">BEEN </token>
    <token time="9600" orig-note="D" tone="5" interval="210">WOR</token>
    <token time="9840" orig-note="F" tone="7" interval="930">KING </token>
    <token time="10800" orig-note="D" tone="5" interval="210">LI</token>
    <token time="11040" orig-note="C" tone="4" interval="210">KE </token>
    <token time="11050" orig-note="C" tone="4" interval="210">A </token>
    <token time="11280" orig-note="D" tone="5" interval="330">D</token>
    <token time="11640" orig-note="C" tone="4" interval="90">O</token>
    <token time="11760" orig-note="B" tone="3" interval="330">G</token>
  </verse>
  ...
</song>

```

Figur 3 Indata till *To sing like a mockingbird* för två låttexttrader. (Gatti et al. 2017, s.300)

För att detektera rim och approximativa rim i indatan så använder sig programmet av *CMU pronunciation dictionary* och en icke specificerad algoritm. Ingen information om algoritmen ges som tyder på att det finns någon skillnad mot WAIY, och den studeras därmed inte närmare.

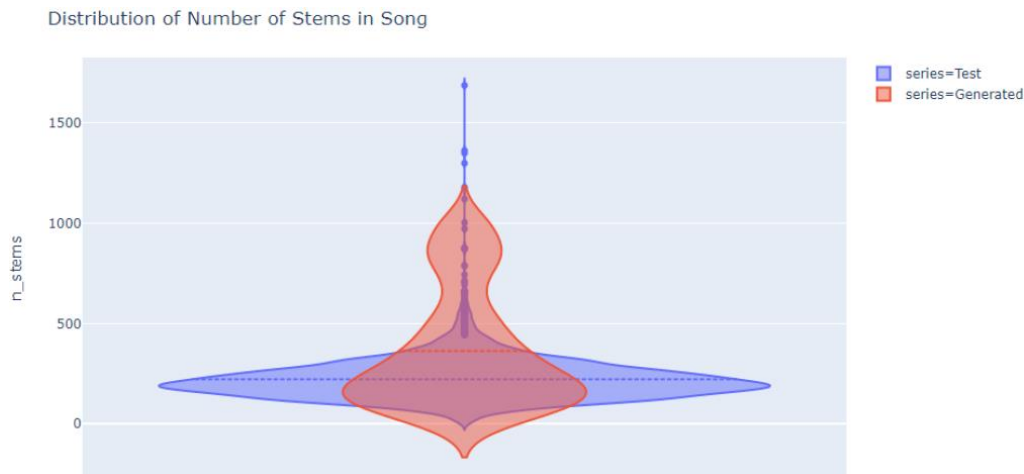
Jämfört med WAIY så har två signifikanta skillnader i *To sing like a mockingbirds* angreppssätt för låttextgenereringen identifierats. Dels görs en databassökning bland i första hand aktuella nyhetsartiklar för att finna koncept att ersätta ord i indatan med. Dels byts bara de av orden i originallåttexten ut som passar ihop med ord hämtade från dessa koncept, vilket gör att en stor del av utdatan kan vara oförändrat. I jämförelse så finner WAIY ersättningsord som har kontextuell närhet till originaltextens ord men i övrigt inga begränsningar vad gäller innehåll, samtidigt som samtliga ord byts ut.

För producerandet av utdatan så kan båda dessa skillnader övervägas att implementeras i en vidareutveckling av WAIY. Vad gäller ansatsen att endast byta ut vissa ord i originallåttexten så bedöms en stor potential finnas för gott resultat vad gäller kvalitet på de parodilåttexter som produceras, tack vare att det ursprungliga flytet och den språkliga riktigheten till stor del kan behållas samtidigt som igenkänningsfaktorn kan förväntas öka. Parodilåttexter som gör mindre förändringar i ursprungslåttexten är därför en metod som kan övervägas, men den kan innebära en ökad risk för upphovsrättsbrott om en del av den upphovsrättskyddade originallåttexten reproduceras och publiceras.

## Application of Machine Learning Model in Generating Song Lyrics

Lau (2021) har studerat egenskaper hos de låttexter som verktyget GPT-2 producerar beroende på vilka databaser som verktyget fått träna på. Det konstateras att om verktyget matas med träningsdata som består av minst 15000 låtar så erhålls en riklig variation i de nyproducerade låttexterna. WAIY använder sig av GPT-2 för de texttrader som inte innehåller rim men utan specifikt träningsdata på den generiskt förtränade språkmodellen. Vid generering utgås ifrån en kort inledning samt de rader som redan har genererats för att försöka åstadkomma ett visst sammanhang inom låttexten, men detta innebär inte att språkmodellen är ytterligare tränad. Även om Laus implementation är anpassad för vanliga låttexter och inte parodilåttexter så kan några olika alternativ för förbättringsåtgärder identifieras.

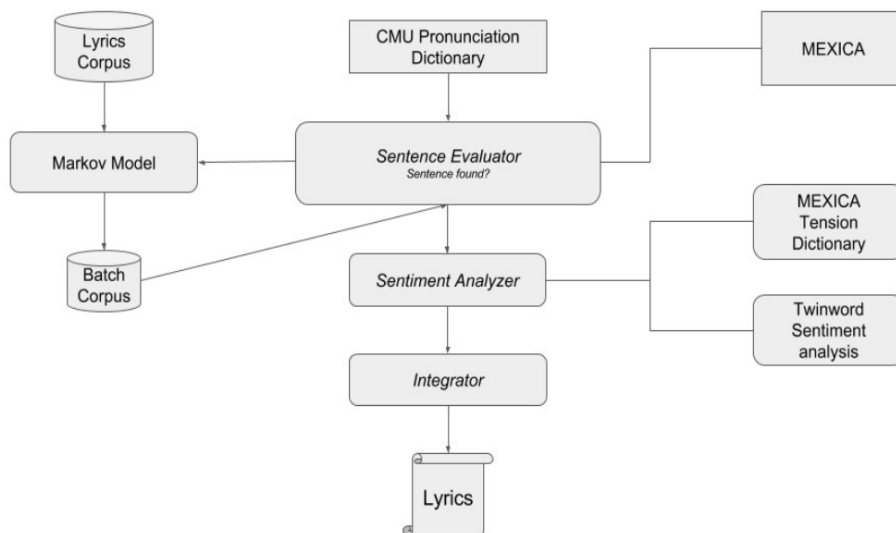
För att fokusera på utdatans kvalitet som låttext i allmänhet kan det vara relevant att använda en större databas med vanliga låttexter som träningsdata precis som Lau har gjort. För att fokusera på parodilåttexters särskiljande egenskaper så kan det vara relevant att använda en databas med endast parodilåttexter eller paroditexter som träningsdata. Ett ytterligare alternativ värt att beakta kan vara en kombination – att komplettera en databas med parodier med vanliga låttexter. Ett alternativ till att låta GPT-2-språkmodellen enbart träna på låttexter är att använda en brett förtränad språkmodell och finjustera med låttexter och/eller parodiska texter.



Figur 4 Jämförelse av längd på de maskingenererade låttexterna och en testmängd med manuella låttexter. (Lau 2021, s.26)

### Automatic lyrical narrative writing

Singhs (2018) system MABLE beskrivs precis som *To sing like a mockingbird* (Gatti et al. 2017) inte i detalj vad gäller de tekniska lösningarna, men det finns ändå allmänt beskrivna lösningar som kan beaktas. Det kan konstateras att metodiken i MABLE (Singh 2018) är säregen jämfört med andra här studerade ansatser på så sätt att rimraderna genereras av den så kallade Markov-modellen från vänster till höger och ingen kontroll görs i förväg att den genererade raden innehåller något rim, eller består av korrekt antal stavelser. Det kräver därmed ett stort antal anrop för att statistiskt sett åstadkomma en rimlig sannolikhet för att finna en rad som uppfyller dessa krav. Detta torde innebära en kompromiss i förhållande till hårdvarans resurstillgång och därmed en viss risk att ibland tvingas acceptera rader med relativt låg kravuppfyllnad. Att ersätta WAIY:s algoritmer för genererandet av låttextraderna med en algoritm som använder sig av Markov-modellen i stället skulle, särskilt då denna algoritm inte beskrivs i detalj av Singh, antagligen innebära ett omfattande arbete. Det noteras även att Lau (2021) skriver angående en annan implementation som använder sig av Markov-modeller: att kritiker menar att dessa saknar minne och därmed också förmåga att generera sammanhängande berättelser med långsiktig struktur. Ett möjligt alternativ skulle kunna vara att använda samma tillvägagångssätt men använda någon av de språkmodeller som WAIY redan utnyttjar: GPT-2 eller XLNet. Vid en sådan implementation behöver det beaktas den hårdvarukapacitet som krävs för att göra ett tillräckligt stort antal anrop av respektive språkmodell.



Figur 5 Översikt över den låttextragerande algoritmen i MABLE. (Singh 2018, s.15)

En annan särskiljande metodik i MABLE är attitydanalysen, *sentiment analysis*, som skulle kunna vara fördelaktig att implementera i andra program. Attitydanalysen görs genom anrop av en extern API från Twinword som bedömer om varje ord har en positiv, neutral eller negativ attityd. Varje rad kategoriseras därefter utifrån genomsnittet av samtliga ords attityd. Medan Singh valt att variera attityden genom att i de nyproducerade låttextraderna använda en motsatt attityd i förhållande till berättelseraderna, så skulle ytterligare överväganden behöva göras kring hur attitydvariationen bör struktureras, särskilt i ett program som genererar låttextror av andra genrer än kärleksballader. På grund av avsaknaden av en detaljerad beskrivning av hur attitydanalysen sker rent tekniskt så studeras den inte närmare här.

## Andra åtgärder

Riedl (2020) påpekar själv att GPT-2 producerar mer flytande resultat än XLNet som används för rader som slutar med rim. Det kan därmed antas att en algoritm som lyckas få GPT-2 att även producera rim ger bättre och jämnare resultat, speciellt då det skulle innebära att samtliga textrader produceras med samma språkmodell. Riedl skriver att anledningen att GPT-2 inte används för hela låttextrorna är att verktyget inte kan anpassas för att generera textrader från slutet och bakåt – vilket blir ett problem när en rad måste sluta med ett rim då det inte går att veta om ett rimord passar in efter de ord som genererats från början av raden.

En tänkbar lösning skulle kunna vara att bestämma rimordet först (eller en mängd kandidatord som alla tillhör samma ordklass och eventuellt tempus), sedan generera textrader med GPT-2 med rätt antal stavelser för att få plats med det utvalda rimordet på slutet. För att se till att rimordet passar in grammatiskt så kontrolleras om det ord som GPT-2 föreslår som följande ord är av samma ordklass och eventuellt tempus som det utvalda rimordet. Om så ej är fallet så förkastas ordraden och processen upprepas. För att undersöka om det utvalda rimordet är kontextuellt relevant tillsammans med resten av raden så kan befintliga funktioner användas för att bedöma relevansen, men då syftet är att skapa en parodisk textrad så är inte denna bedömning nödvändigtvis viktig eller ens önskvärd. Tänkbara nackdelar med denna ansats är potentiellt längre körtider om det krävs många försök innan en rad där det utvalda rimordet passar in genererats, samt mindre flyt i texten då rimordet inte valts ut av AI-verktyget att passa ihop med resten av raden.

För val av rimord så har Riedl (2020) skapat en algoritm vars syfte är att finna ett rimord som med relativt hög sannolikhet är kontextuellt relevant i den textrad som först i nästa steg genereras. Två olika angreppssätt diskuteras: *cosine similarity* och uppskattningen av sannolikheten för att varje kandidatord genereras naturligt. Slutsatsen är att *cosine similarity* resulterar i tråkigare val av ord, men då detta tycks vara ett helt subjektivt konstaterande så kan en ny bedömning av angreppssätt anses vara av intresse.

Riedl (2020) gör ett allmänt påpekande angående ansatsen för sitt program att som djupinlärningsforskare säger instinkten att man skulle kunna träna ett nytt eller anpassa ett befintligt neuralt nätverk för att skapa låttexter som slutar med rim. Det konstateras dock att det verkar som en svår uppgift varför Riedl själv avfärdade denna metod. För en framtida större studie eller projekt så kan denna möjlighet övervägas.

Utifrån Singhs (2018) inriktning mot ett program som används som hjälpmedel av en mänsklig låtskrivare, så skulle relativt enkelt ny funktionalitet kunna implementeras som bygger på Gatti et als (2017) ansats att enbart byta ut vissa ord. Detta genom att låta användaren vid inmatning bestämma vilka ord eller rader som inte ska ändras, och så genererar programmet rader som passar ihop med de rader som behållits.

Då implementationstesterna i 4.1.2 visar att detektionen av antal stavelser per rad ofta tar fel så hade en funktion för att manuellt ange antalet stavelser varit användbar för en användare som önskar högre precision till bekostnad av större arbetsinsats. Detta kan särskilt antas vara lämpligt för någon som ändå tänkt manuellt arbeta med resultatet så som föreslaget av Singh (2018). Då programmet inte tar hänsyn till kontexten i indatan så finns det dock en existerande lösning som innebär att användaren själv ersätter eventuella rader som feltolkas med en lättolkad ny rad. I den nya raden bör det sista ordet antingen behållas eller ersättas med ett ord som rimmar med det ursprungliga sista ordet – detta för att rimdetektionen inte ska påverkas.

# 3 Metod

## 3.1 Forskningsstrategi

Forskningsstrategin i denna studie är en kombination av aktionsforskning och undersökning. Aktionsforskning går ut på att lösa praktiska problem i verkligheten och samtidigt generera ny teoretisk kunskap (Johannesson & Perjons 2014). Arbetet är cykliskt och består av olika delar i en process.

Den alternativa forskningsstrategin som övervägdes men valdes bort är experiment. Anledningen till att det ansågs mindre lämpat än aktionsforskning för denna studie är delvis att syftet var ett bredare angreppssätt än vad som stipuleras i ett experiment (Denscombe 2014). Dessutom hade de höga kraven på validitet varit svåruppfyllda på grund av den stora variationen i AI-genererade parodisångtexter. Detta då utdatan skiljer sig relativt kraftigt vid varje körning även med samma indata, på ett slumpartat sätt. En fördel hade varit att experiment ofta anses vara det mest tillförlitliga angreppssättet.

Enligt Johannesson och Perjons (2014) så är kännetecknen för aktionsforskning förutom det cykliska arbetet och att resultatet bör medföra användbara praktiska lösningar, fokuset på det praktiska arbetet som studeras och hur det förändras. Aktivt deltagande i studien av studieobjekten är också ett kännetecken, men det är inte aktuellt för denna studie då tid ej finns för inblandning av externa användare under det cykliska arbetet med vidareutvecklingen av WAIY.

Det cykliska forskningsarbetet består av fem delar (Johannesson & Perjons 2014): de första två är diagnostisering samt planering, vilket i denna studie motsvaras av litteraturgenomgången i bakgrundskapitlet som ämnar att isolera relevanta tekniska lösningar som kan implementeras i vidareutvecklingen av WAIY. Nästa del är intervention, vilket innebär själva implementationen av den plan som upprättats. Implementationen innebär skrivandet av ny kod för att läsa in en databas som språkmodellerna kompletteringstränas på. Den fjärde delen är evaluering – d.v.s. det nya programmet körs upprepade gånger med olika indata för att ge en uppfattning om hur utdatan har förändrats. Sista delen är reflektion vilket bland annat innebär en bedömning av huruvida förändringarna ledde till nya problem eller inte. Med erhållen kunskap om resultatet så upprepas sedan cykeln med återgång till diagnostisering följt av övriga delar, tills att man under reflektionen drar slutsatsen att fortsatt arbete inte längre är meningsfullt och/eller att det inte finns mer tid.

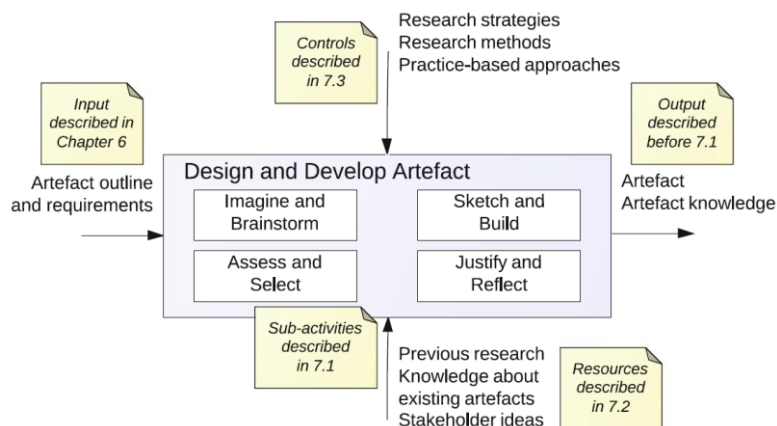
Efter att aktionsforskningsfasen är avslutad så tar en fas med forskningsstrategin undersökning vid. För undersökningen i denna studie så har en kvantitativ ansats valts då huvudsyftet är att ge ett konkret svar på frågeställningens underfråga 2:

2. Vad anser svars personer i en enkätundersökning om dessa parodilåttexter med avseende på hur roliga de är, jämfört med parodilåttexter som originalprogrammet genererat?

Den upplevda roligheten innebär i denna studie ett högre genomsnittligt omdöme i en enkätundersökning som svar på frågan *How funny do you think these parody lyrics are?* där svarsalternativen består av nio steg från *not funny at all* till *on par with the best comedy I've seen or heard*. En förbättrad version av WAIY producerar alltså parodilåttexter som till en statistiskt signifikant grad får högre omdömen än originalprogrammets utdata.

Vad gäller den kvantitativa ansatsen i undersökningsfasen så kan flera fördelar identifieras. De inkluderar att forskaren är avskild från själva datainsamlandet, och därmed är normalt sett inte påverkan på svaren särskilt stor (Denscombe 2014). Medan en kvalitativ ansats ger ett holistiskt perspektiv och skulle kunna svara på frågor som rör en viss grups uppfattningar eller egenskaper på ett djupare plan, men utan att göra några statistiska beräkningar eller jämförelser, så kan denna studie tack vare den kvantitativa ansatsen dra konkreta slutsatser och svara på frågeställningen. Kvantitativt analysarbete tar dessutom mindre tid vilket möjliggör större fokus på den tekniska implementationen i aktionsforskningsfasen. Med ett tekniskt analysverktyg så krävs inte heller någon djupare teoretisk eller teknisk kunskap om de statistiska metoderna, vilket är ännu en fördel med kvantitativ metod. En nackdel och en utmaning är att kvaliteten på datan inte nödvändigtvis syns i analysresultatet, vilket kan leda till ett resultat som inte speglar potentiellt bristande datakvalitet.

## 3.2 Litteraturgenomgång och tekniskt arbete



Figur 6 Översikt över arbetet med design och utveckling av artefakt enligt Johannesson och Perjons (2014, s.118)

Återkommande under aktionsforskningsfasen så studeras tidigare forskning på området i detalj tillsammans med Riedls (2019) programkod för WAIY för att få en uppfattning om vilka förändringar som kan göras av programmet för att öka kvaliteten på de parodilåttexter som produceras.

Som beskrivet i förra avsnittet så är vidareutvecklingen av WAIY ett cykliskt aktionsforskningsarbete. Som Johannesson och Perjons (2014) beskriver så är valet av forskningsstrategi vad gäller design och utveckling oftast av mindre betydelse då fokus ligger på att färdigställa en artefakt och inte på teorin bakom utvecklingen. Därför följer inte heller forskningsarbetet under vidareutvecklingen aktionsforskningens principer fullt ut och dokumentationen av metodiken under arbetet är något begränsad. Detta förhållningssätt gäller dock inte undersökningsfasen som är av mer teoretisk natur.

WAIY är skrivet i Python och använder sig av en databas med ett rimlexikon där även approximativa rim ingår. Programmet har ett enkelt grafiskt användargränssnitt där användaren själv matar in en låttext tillsammans med information om antal stavelser och rim. Raderna byggs utifrån de begränsningar som definierats vad gäller antal stavelser och rim, med kontextuellt relevanta ord som väljs genom att den externa språk-AI:n använder originallåttextern samt eventuella redan producerade rader som indata för generering av nästa rad. Rader med rim byggs från höger till vänster med hjälp av språk-AI:n XLNet medan rader utan rim byggs från vänster till höger med hjälp av språk-AI:n GPT-2 (Riedl 2020). Då WAIY är skrivet för att köras i Google Colab så utförs både kodskrivandet och alla körningar i denna miljö.



---

**Algorithm 1:** The lyric generation loop.

---

**Data:** *prompt* is a string; *scheme* is a list of lines where each line is a list of segments where  $segment_i = \langle s_j, r_j, e_j \rangle$  for  $j = 1 \dots n$  such that  $s_j$  is the number of syllables in the segment,  $r_j$  is rhyme index or string, and  $e_j$  is an optional signifier that this segment should end a sentence; and *recontextualize?* is a boolean.

**Result:** A list of strings constituting the lines of the lyrics

```
context ← prompt
for line in lines do
  if recontextualize? = true then
    | Insert prompt in context after last occurring period
  for segment in line do
    target_syllables ← number of syllables specified in segment
    rhyme_index ← rhyme index specified in segment
    end? ← true if segment specifies the segment ends in a period
    if rhyme_index is a string or rhyme_index in rhyme_map then
      end_targets ← pick rhyme words or use rhyme_index
      for target in end_targets do
        candidates ← candidates + generate_rhyme_lines(target, context,
          | target_syllables, end?)
      else
        if end? = true then
          candidates ← candidates + generate_terminal_non_rhyme_lines(context,
            | target_syllables)
        else
          candidates ← candidates + generate_non_rhyme_lines(context, target_syllables)
    best ← pick_best_candidate(candidates, context)
    context = context + best
    final_segments ← final_segments + best
  final_lines = final_lines + final_segments
```

---

Figur 7 Översikt över WAIY:s algoritm för parodilåtttextgenerering (Riedl 2020, s.5)

Under vidareutvecklingsarbetet så används Johannesson och Perjons (2014, s.125) riktlinjer för design och utveckling av artefakt:

- *Clearly Describe Each Component of the Artefact. Describe both the functionality and the structure of each component of the artefact.*
- *Justify Each Component of the Artefact. Explain the purpose of each component of the artefact, in particular which requirements it addresses.*
- *Describe the Use of the Artefact. Describe how the artefact and its components are to be used in its intended practice.*
- *Clarify the Originality. Describe the respects in which the artefact is different from existing ones, with respect to both functionality and structure.*
- *Specify the Sources of the Design of the Artefact. Describe the literature and the stakeholders that have contributed to components of the artefact and/or inspired the design of new components.*
- *Describe How the Artefact Has Been Designed and Developed. Explain what has been done to design and develop the artefact, in particular how the stakeholders have been involved and how existing solutions and research literature have been reviewed.*

Då det handlar om vidareutveckling av det befintliga fullt fungerande programmet WAIY så berörs endast förändringarna i arbetet enligt ovan. Dokumentation av hela processen inklusive eventuella ångrade förändringar presenteras i resultatdelen.

Johannesson och Perjons (2014) beskriver tre typer av bias vad gäller beslutsfattandet kring vilka designidéer som implementeras, som kortfattat kan beskrivas som tendensen att föredra de först tilltänkta idéerna, idéer som bekräftar forskarens uppfattningar och idéer som rättfärdigar tidigare beslut. Genom initial kännedom och upprepat ifrågasättande av fattade beslut utifrån att de är ett resultat av dessa biaser så minskas risken något att de påverkar resultatet. Risken för påverkan kan dock ej elimineras.

WAIY är licensierat i enlighet med *MIT license* vilket bland annat innebär att vem som helst får ändra i koden och distribuera vidare, så länge som alla vidareutvecklingar av programmet också ger användaren samma rättigheter. Detta är fullt förenligt med studiens mål och därmed krävs inget samtycke från originalprogrammets upphovsrättsinnehavare för studiens genomförande, eller för publicering av någon del av examensuppsatsen.

### 3.3 Applicering av implementation

När en teknisk implementation av en föreslagen uppgraderad version av WAIY är genomförd, så används både detta program och originalprogrammet för att generera ett mindre antal parodilåttexter av kända sånger. Programkörningarna görs på indata i form av ett till originallåttextern orelaterat ämne samt en populär låtttext. Då de utdata som jämförs genereras med samma indata så görs ingen djupare analys av effekterna av valt ämne. För att anknyta till programmets inspirationskälla och garantera att en parodisk version av respektive låtttext är möjlig att producera så utgörs ämnet av en kort sammanfattning av Weird Al Yankovic' parodiversion av respektive originallåt. Sammanfattningen är manuellt gjord av denna studies författare och presenteras nedan, vilket ger möjlighet att kontrollera sammanfattningens validitet i förhållande till den äkta parodilåttextern.

Originallåttitel	Parodilåttitel	Sammanfattning
Another One Bites the Dust	Another One Rides the Bus	Riding a cramped and hot bus
Smelling Like Teen Spirit	Smelling like Nirvana	Unintelligible and incoherent song
Lose Yourself	Couch Potato	Lazily watching tv from the sofa

Tabell 1 Sammanfattningar av Weird Al Yankovic parodiversioner av låtarna som används i indatan

Valet av sånger att låta de olika programversionerna parodiera är utifrån statistik över antal spelningar på Spotify uppdelat på decennium som låtarna släpptes (Sentireascoltare 2020). Denna källa används då det är den mest aktuella lista som funnits tillgängligt gratis och hittats genom sökning på Google. Att urvalet gjorts utifrån antal spelningar på Spotify är för att Spotify är den största tjänsten för musikströmning i skrivande stund (Statista 2021) och finns i ett stort antal länder i världen, vilket gör att låtarnas igenkänningsfaktor bland dem som svarar på enkäten antas vara hög. Det kan även noteras att såväl Spotify som att svara på enkäten kräver en viss grad av digital kunskap vilket skulle kunna innebära att målgrupperna överlappar väl. Innan slutligt val av låt per decennium görs kontrolleras att låttexterna innehåller minst ett rim per vers och är på korrekt standardengelska. Om topplistan innehåller en eller flera låtar som Weird Al Yankovic själv parodiserat så väljs en sådan, och bland de låtar som uppfyller kriterierna så väljs den mest spelade låten.

- Från 2000-talet har låttextern till låten *Lose yourself* av Eminem valts, som är den näst mest spelade låten från detta decennium, och förutom att uppfylla kraven på rimfrekvens så har Weird Al Yankovic skrivit en parodilåtttext på denna låt.
- Från 1990-talet har låttextern till låten *Smells like a teen sprit* av Nirvana valts, som är den näst mest spelade låten från detta decennium, och är också parodiserad av Weird Al Yankovic.
- Från 1980-talet har låttextern till låten *Another one bites the dust* av Queen valts, som är den tredje mest spelade låten från detta decennium, och är också parodiserad av Weird Al Yankovic.

För att undvika att den totala tiden för att slutföra enkäten blir lång och respondenterna därför riskerar tröttna, vilket bör undvikas (Denscombe 2014), så förkortades låttexterna till ungefär hälften av den

ursprungliga, med minst en refräng inkluderad för bibehållen igenkänningsfaktor. Detta innebar två verser och refrängen två gånger för *Smells like teen spirit* och *Another one bites the dust*, men en vers och refrängen för *Lose yourself*, vars låttext är betydligt längre.

### 3.4 Datainsamling

Datainsamlingen görs som en webbaserad enkätundersökning, för att svara på underfråga 2 i frågeställningen.

2. Vad anser svars personer i en enkätundersökning om dessa parodilåttexter med avseende på hur roliga de är, jämfört med parodilåttexter som originalprogrammet genererat?

Enkätundersökningar har enligt Blattacherjee (2012) fördelar som enkelheten att göra storskaliga studier och att många respondenter föredrar enkäter framför andra datainsamlingsmetoder i och med flexibiliteten att kunna fylla i enkäten när man själv har tid och lust. Nackdelar är bland annat risken för urvalsbias och låg svarsfrekvens, vilket påverkas av hur urvalet gjorts.

Ett alternativ hade kunnat vara en kvalitativ analys genom att intervjua ett antal personer om hur de upplever parodisångtexterna. Detta alternativ har valts bort på grund av att intervjuer associerade med vissa nackdelar som minskat användbarheten av denna studie. Det handlar främst om den relativt låga reliabiliteten som uppstår när analysen baserar sig på ett litet antal individer (Denscombe 2014). Att inte lika många individer kan ingå i en intervjuundersökning som i en enkätundersökning beror på att tidsåtgången för intervjuer är mycket högre.

En risk med webbenkäter är att svaren kan manipuleras om enkäten inte är skyddad från intrång eller att samma person svarar flera gånger etcetera. Om antalet svar skulle bli mycket stort eller oväntade mönster identifieras i svarsdatan så implementeras åtgärder för att minska risken för intrång och säkerställa att den som svarar på enkäten är en människa. Detta kan innebära att enkäten måste startas om och tidigare insamlade data raderas.

Vad gäller urvalsgruppen så definieras den brett: alla vuxna personer med relativt goda kunskaper i engelska för att enklare uppnå ett stort antal svarande. Enligt Blattacherjee (2012) så är en urvalsgrupp drabbad av bias om stickprovsvariabler inte är normalfördelade, eller om flera olika subgrupper inom urvalsgruppen ger varierande distributioner, vilket analyseras i analysdelen.

För att underlätta fler inkomna svar så är enkäten på engelska och innehåller frågor om respondentens kön, ålder, boendeland, modersmål och engelskkunskaper, samt frågor om upplevelsen av de parodilåttexter som presenteras. Vad gäller frågan om engelskkunskaper så korresponderar svarsalternativen mot CEFR-standardens tre nivågrupperna *basic* (CEFR A), *intermediate* (CEFR B) och *advanced* (CEFR C), samt inkludera *mother tongue or equivalent* (hädanefter benämnt *mtoe*). Vem som helst kan besvara webbenkäten och för att få in ett adekvat antal svar så annonseras enkäten privat i sociala medier.

Vid konstruktion av enkäten så tas hänsyn till enkättagarnas bakgrundskunskap i frågorna (till exempel att de känner till de originalsånger som valts ut) och generellt bör enkätfrågorna som används i studien uppfylla de kriterier som Blattacherjee (2012) presenterar för enkätundersökningar. En fråga verifierar ifall respondenten känner till den aktuella originallåttextern eller inte, med möjlighet att ange i vilken grad låten samt låttextern är bekant. Upplevd kvalitet mäts som svar på enkätfrågan *How funny do you think these parody lyrics are?* med svarsalternativen på en ordinalskala 1-9 med beskrivning av min- och maxvärdet.

Om praktiskt genomförbart så fanns en ambition att låta enkätdeltagarna få tillgång till en modifierad version av den ena eller båda versioner av WAIY. Detta för fri egen användning där data skulle inhämtas om all interaktion mellan användaren och programmet som senare analyseras tillsammans med svaren på ytterligare enkätfrågor angående hur man upplevde själva programmet och de eventuella parodilåttexter man själv låtit producera. Detta har visat sig ej kunna genomföras på grund av tekniska svårigheter: bland annat då programmet utan en större tidskrävande förändring i design måste initieras och köras på Google Colab-servrar där bara initieringen tar flera tiotals minuter, och genereringen av parodilåttexter tar ytterligare flertalet minuter per låttext.

Efter de demografiska frågorna så presenteras färdiga utskrifter med parodilåttexter på de utvalda originallåttexterna (som anges med artist och låtnamn, ej kompletta originallåttexter), som är producerade av WAIY samt av den i den tekniska implementationsdelen utvalda vidareutvecklade WAIY+. För att erhålla ett referensvärde som beskriver hur rolig respondenterna tycker att en äkta parodisång är så har en parodilåttext som Weird Al Yankovic skrivit på en av de valda låtarna också inkluderats i enkäten. För att möjliggöra jämförelse mellan svaren på de AI-genererade låttexterna och den äkta låttextern så presenteras Yankovics låttext på samma sätt, i detta fall som parodilåttext #2 på *Smelling like teen spirit*. Resultatet blir att den låten har tre parodilåttexter och de andra låtarna två, men då urvalet inte förklaras för enkätdeltagarna så bör detta inte påverka nämnvärt. För alla låttexter så presenteras texten från början till och med första refrängen, för att undvika att enkäten tar alltför lång tid att besvara.

Enkätdeltagaren ombeds betygsätta varje parodilåttext med hur rolig den är under frågan *How funny do you think these parody lyrics are?* och med svarsalternativ 1–9 där 1 innebär helt humorbefriad och 9 i klass med den roligaste humor man hört. Kompletterande frågor om varje låttext är *To answer the previous question, did you listen to the melody while reading?* (med avseende på frågan där man betygsätter roligheten), *Regarding these parody lyrics - are you familiar with the original song and its lyrics?* och *Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song?* Den sista frågan har svarsalternativ 1-9 där 1 innebär inte alls realistisk och 9 innebär att man inte ser eller hör någon skillnad. Ordningen på parodilåttextfrågorna har slumpats men är på grund av tekniska begränsningar i enkätverktyget lika för alla enkätdeltagare. Enkäten presenteras i sin helhet i Bilaga A.

Denscombe (2014) skriver att en lyckad enkät uppfyller tre villkor: hög svarsfrekvens, hög ifyllnadsgrad och svar som kan valideras. Detta tas hänsyn till i enkätkonstruktionen och analyseras efter datainsamlingen. Enkäten innehåller ej några känsliga frågor som skulle kunna påverka viljan att fylla i hela enkäten. Man kan svara *ej svar* eller hoppa över frågan vad gäller samtliga frågor för att inte riskera att respondenter väljer ett oärligt svar om de känner sig obekväma med eller osäkra på någon fråga. Längden minimeras så att respondenterna orkar fylla i ordentligt och inte svarar slumpvist eller *ej svar* på flertalet frågor. Svarsfördelningarna analyseras för att försöka upptäcka om svaren på någon fråga har låg kvalitet, till exempel genom att se om svarsfördelningarna på kvotskalevariabler är normalfördelade eller ej (Denscombe 2014).

För att öka validiteten och minimera ovanstående felkällor så utförs en pilotstudie där man samlar in enkätsvar från ett mindre antal personer och intervjuar dem för att verifiera att de svarar i enkäten på det sätt som enkätkonstruktören förväntat sig (Denscombe 2014). I intervjun ställs även frågor i avsikt att finna förbättrade formuleringar i den slutliga enkäten. Dessa tips förkastas dock generellt i det fall svaren på enkäten stämmer väl överens med enkätkonstruktörens förväntningar, för att inte riskera skapa nya problem. Om enkäten förändras väsentligt så bör en ny pilotstudie utföras i syfte att bekräfta

att den förändrade enkäten bättre än tidigare uppfyller syftet att mäta det som i frågeställningen önskas svar på.

Reliabilitet handlar om kvaliteten i undersökningen, mätt som förmågan att komma fram till samma resultat oavsett hur många gånger studien utförs eller av vem. Denscombe (2014) associerar reliabilitet i enkätundersökningar med svarsfrekvens. Hög svarsfrekvens i sin tur förutsätter hög kvalitet i enkätkonstruktionen så att respondenterna är motiverade och förstår innehållet väl.

Vetenskapsrådet (2014) skriver om fyra viktiga begrepp att tänka på vid hantering av forskningsmaterial: sekretess, tystnadsplikt, avidentifiering samt konfidentialitet. Vid användning av enkätundersökning som datainsamlingsmetod är den etiska aspekten om avidentifiering en viktig aspekt, d.v.s. att det inte finns någon koppling mellan respondentens svar på en enkät och deras identitet. Denna hantering underlättas i studien genom att helt undvika insamling av personuppgifter. Ett annat sätt att identifiera personer är om de tillhör en kategori med få respondenter. Då kan det även vara värt att ta bort dessa från undersökningen. Vad gäller specifika etiskt grundade åtgärder på datan så beskrivs dessa i analysdelen.

## 3.5 Dataanalys

Dataanalysen utförs i enlighet med vedertagna statistiska modeller så som beskrivet av Denscombe (2014) och Johannesson och Perjons (2014). Detta görs med hjälp av programvaran *IBM SPSS Statistics* som utför både de statistiska beräkningarna i analysen samt producerar de tabeller och figurer som används för att illustrera analysprocessen samt -resultaten. Först presenteras datan översiktligt med deskriptiv statistik, bland annat för att påvisa datans trovärdighet då källmaterialet ej är allmänt tillgängligt. Den följande analysprocessen som valts är inferentiell statistik – även kallat slutsatsstatistik – då det krävs för att kunna svara på frågeställningen som åsyftar den upplevda roligheten i en större målgrupp utifrån svaren från en mindre urvalsgrupp. I inferentiell statistik kan man dra slutsatser om olika variablers samband, vilket kan användas för att göra ytterligare konstateranden om det insamlade datan bortom själva frågeställningen.

Huvudfrågeställningen handlar generellt om skillnaderna mellan vidareutvecklingen i och med WAIY+ och originalversionen. Underfråga 2 specificerar att den upplevda roligheten jämförs. Med andra ord jämförs medelvärdet mellan två parade samplingar av samma urvalsgrupp, där den eventuella skillnaden och dess statistiska signifikans är relevant. Förutsatt att den upplevda roligheten ses som en kvotskala/intervallskala så är t-test med parat urval ett lämpligt statistiskt test (Denscombe 2014). Testet görs utifrån nollhypotesen:

- $H_0$ : Det finns inte någon statistiskt signifikant skillnad i upplevd rolighet mellan de låttexter som genereras av WAIY+ och WAIY.

Om  $p < 0,05$ , d.v.s. sannolikheten att skillnaden i medelvärde beror på en slump är mindre än 5% så kan nollhypotesen förkastas. Om så är fallet så jämförs de två medelvärdena och det kan fastställas huruvida den upplevda kvaliteten av den vidareutvecklade versionen av WAIY är bättre eller sämre än originalversionen.

De inledande frågorna handlar om respondentens bakgrund vilket åtminstone inkluderar kön, ålder och modersmål. De används för att undersöka ifall det finns något statistiskt signifikant samband mellan upplevd kvalitet av respektive programversion och dessa faktorer. Vad gäller skillnad mellan könen så kan ett chi-2-test utföras då det handlar om att undersöka det eventuella sambandet mellan den

nominala variabeln kön samt den ordinala variabeln upplevd kvalitet (Denscombe 2014). Om upplevd kvalitet ses som en kvotvariabel så kan även ett t-test med oberoende urval utföras, för att undersöka om den eventuella genomsnittliga skillnaden är signifikant. Då även modersmål är en nominal variabel så kan samma typ av test utföras för att undersöka om det finns ett samband mellan modersmål och upplevd kvalitet.

Vad gäller samband mellan ålder och upplevd kvalitet så kan Spearmans korrelationstest svara på om det finns ett statistiskt signifikant samband mellan intervallskalevariabeln ålder och den ordinala variabeln upplevd kvalitet. Om enkätdeltagarna delas in i två grupper utifrån ålder så kan även här ett t-test med oberoende urval utföras. I övrigt så övervägs ytterligare tester utifrån vilka bakgrundsfrågor som beslutas ingå i enkäten, och vilka svarsalternativ som erbjuds. Det skulle till exempel kunna handla om att undersöka ett eventuellt samband mellan ålder och kännedom om de olika originallåtarna från olika årtal.

# 4 Resultat

## 4.1 Implementation

Utifrån frågeställningen och konklusionerna kring genomförbarhet från bakgrundskapitlet för den funktionalitet som funnits i tidigare forskning, så har olika alternativ testats för den tekniska implementationen av en vidareutvecklad version av WAIY. Denna version kallas WAIY+ och bygger alltså på att språkmodellerna finjusteras med extra träning på en lämplig databas.

I första hand eftersöktes en databas över parodilåttexter men då denna genre är relativt liten och många låttextrdatabaser antingen inte klassificerar efter genre alls eller saknar komedi/parodi som genre så visade sig detta vara svårt (Open Lyrics Database u.å.; Azlyrics u.å.). På samma sätt som konstaterades vid arbetet med Alternativ 1 så finns få öppettillgängliga färdiga databaser över låttexter och ingen varifrån specifikt parodilåttexter kunde erhållas hittades. Likväl valdes tre alternativ ut för implementation i WAIY+. All tillförd kod återfinns i Bilaga B, och är tillagd efter inläsningen av språkmodellerna i början av sektionen *Load neural language models* i originalkoden till WAIY. Ingen annan del av koden är ändrad.

Då de två språkmodellerna som används i WAIY GPT-2 och XLNet tränas på olika sätt (Rusia 2019) så har försök gjorts att finna metodik för att träna båda språkmodellerna för ett likartat resultat. Dessa försök har dock varit fruktlösa och därför har det i denna studie valts att implementationen begränsas till att träna GPT-2.

### 4.1.1 Alternativ för träning av språkmodellen

#### Alternativ 1

Som första alternativ provades att träna språkmodellen på en databas över låttexter, då de publikt tillgängliga förtränade GPT-2-modellerna är tränade på alla typer av texter. Att finna färdiga databaser över låttexter som tillhandahålls utan kostnad var svårt, då de flesta moderna låttexter är skyddade av upphovsrätt och därmed inte kan spridas lagligt och fritt. Den databasen som valdes är skapad av Anderson (2019) och innehåller 164790 låttexter varav 55%, alltså ca 90000 låttexter är på engelska. Den totala storleken på databasen är ca 270 Mb.

Dessa ca 90000 låttexter kan ställas i relation till 15000 som Lau (2021) fann krävdes för att undvika en stor grad av repetition, och till 38900 som Lau använde när ett relativt bra resultat erhållits. Andersons databas innehåller dock låtar tillhörandes 6 olika genrer: rock, hip hop, pop samt tre brasilianska genrer, medan Laus slutsatser utgår ifrån att träningen sker på en specifik genre. Då WAIY ej har någon funktion för att välja musikgenre och implementerandet av någon sådan ny funktion ej övervägs så kan detta anses vara av mindre betydelse i förhållande till WAIY+. De brasilianska låtgenrerna har endast låtar på portugisiska och då låtar som inte är på engelska sällats bort så ingår inte dessa genrer i den data som läses in. Övriga genrer behövs dock och därmed tränas språkmodellen på låttexter från de tre ovan nämnda genrerna. Antalet låtar bör ses i ljuset av att genrerna är blandade och därför kan ej någon slutsats dras kring vilken variation som kan förväntas.

Rent tekniskt användes en algoritm utvecklad av St-Amant (2021) som anpassades för att ladda databasen från en monterad Google Drive och ta med alla låtgenrer. Låttexter på över 1024 tokens rensades bort då språkmodellsträningen inte klarar långa texter. Antalet epoker bibehölls på standardvärdet 5 för själva träningen av språkmodellen.

## Alternativ 2

Som andra alternativ tränas GPT-2-språkmodellen på en databas över skämttexter, för att försöka åstadkomma bättre parodiska inslag i de genererade parodilåttexterna. Databasen är producerad av Abhinav (2017) och innehåller 231656 korta skämt som är mellan 10 och 200 tecken långa. Den totala storleken på databasen okomprimerat är 23,5 Mb vilket alltså innebär en relativt liten databas jämfört med Andersons (2019). Då databasen verkar vara manuellt sammanställd och har ett konsekvent format så kan det tänkas att kvaliteten på textinnehållet är högre vilket skulle kunna väga upp för den mycket mindre textmängden jämfört med de andra alternativen. Detta har dock ej konstaterats genom någon utförlig analys.

Kodningen utgick från samma algoritm som i Alternativ 1, av St-Amant (2021), men justerades utifrån strukturen hos skämttextdatabasen. Då ingen skämttext var i närheten av maxantalet 1024 *tokens*, så har samtliga texter oförkortat tagits med. Även här bibehölls antalet epoker på 5.

## Alternativ 3

Som tredje alternativ för träning av språkmodellerna som testats är finjustering på en större databas över parodisk text, för att försöka öka kvaliteten på parodiaspekten i de genererade parodilåttexterna. Även om det fanns liknande svårigheter som vid eftersökandet av de andra databaserna så kunde en relativt stor databas på 908 Mb ändå implementeras. Det är en avbildning från det parodiska fritt redigerade nätuppslagsverket Uncyclopedia. Samtidigt som det relativt stora antalet artiklar som ingår i databasen är en fördel för träningsmodellen så kan en betydande nackdel identifieras i form av det faktum att artiklarna är skrivna och redigerade av ett stort antal okända skribenter. Kvaliteten på artiklarna kan därmed antas ha en hög variation. Ett annat frågetecken kring resultatet kan ställas angående typen av paroditext – huruvida parodiska uppslagsverksartiklar har egenskaper som lämpar sig i en parodisk låttext. Då Uncyclopedia beskrivs som [*Uncyclopedia*] is considered to be a rich source to build humor ontology (Jamshidlou & Moradi 2013) så bör den kunna anses som en bra representation av humor i allmänhet. Uppslagsverket som helhet bygger dessutom på uppslagsverket Wikipedia och många av dess artiklar har icke-parodiska motsvarigheter där. Därför kan det antas att Uncyplopedias artiklar generellt kan ses som parodier, som därmed bör lämpa sig som underlag för generering av paroditext.

För att läsa in Uncyplopedias avbildning i form av en xml-fil så skapades en ny funktion för att finna själva sidorna i xml-hierarkin, samt rensa bort alla sidor som inte var artiklar. Dessa ej inlästa sidor var exempelvis profiler hos Uncyplopedias registrerade användare, artiklars diskussionssidor samt hjälpsidor. I själva artiklarna togs bland annat länkar, faktarutor, innehållsförteckningar och referenslistor bort och slutligen rensades alla skiljetecken bort förutom komma, punkt och apostrof . Långa artiklar förkortades för att fungera med träningsalgoritmen till max 1024 tecken. I dessa fall ses det till att artiklarna förkortas vid en punkt, för att undvika avklippta meningar i träningsmaterialet. Antalet epoker vid träningen sänktes till 2 på grund den stora datamängden vilket innebär långa körtider.

### 4.1.2 Inför implementationstest

Vid testkörning av programmet visade det sig att analysen av stavelsestrukturen gjorde flertalet felaktiga bedömningar av antalet stavelser. Exempelvis missbedömdes de tre första raderna i *Another one bites the dust*:

Originallåttext	Antal stavelser	Av WAIY analyserat antal stavelser
Steve walks warily down the street	8	5



With the brim pulled way down low	7	6
Ain't no sound but the sound of his feet	9	8
Machine guns ready to go	7	7

Tabell 2 WAIYs stavelseanalys av de första fyra raderna i låttextern till Another one bites the dust

En komplett analys av originallåttextern gjordes för en översikt över hur ofta WAIY räknar fel på antalet stavelser. Endast den ursprungliga utdatan utan upprepningar av refräng och andra rader har räknats med i analysen.

Originallåtttext	<i>Another one bites the dust</i>	<i>Smells like teen spirit</i>	<i>Lose yourself</i>	<b>Totalt</b>
<b>Antal stavelser</b>	169	105	260	534
<b>Antal rader</b>	23	18	29	70
<b>Missade stavelser</b>	6	17	17	40
<b>Överräknade stavelser</b>	1	2	0	3
<b>Fel i totala antalet</b>	5	15	17	37
<b>Totala antalet fel</b>	7	19	17	43
<b>Antalet felaktiga rader</b>	5	8	7	20
<b>Andel felaktiga stavelser</b>	4,1%	18,1%	6,5%	9,6%
<b>Andel felaktiga rader</b>	21,7%	44,4%	24,1%	30,1%

Tabell 3 Analys av det faktiska stavelseantalet och det av WAIY uträknade stavelseantalet

Då WAIY saknar funktionalitet för att åsidosätta den automatiskt uträknade stavelsestrukturen och ersätta med manuellt inmatade värden så ändrades låttexterna för att åstadkomma ett korrekt resultat vad gäller hela indatanaalysen med stavelse- och rimstruktur. Genom att ändra ett ord i taget och sedan köra programmet så kunde orden vars antal stavelser inkorrekt analyserats, identifieras och därefter bytas ut. WAIY använder endast antalet stavelser per rad och för varje rads sista ord vilka ord som det rimmar med. Därför så påverkas inte utdatan av ändringar där ersättningsordet rimmar med det ursprungliga ordet, vad gäller varje rads sista ord, och inte alls av ändringar av övriga ord.

Originallåtttext	Justerad låtttext för indata
Steve walks warily down the street	Steve walks steadily down the street
With the brim pulled way down low	With the Jim pulled way down low
Ain't no sound but the sound of his feet	Isn't no sound but the sound of his feet
Machine guns ready to go	Machine guns ready to go

Tabell 4 Exempel över hur originallåttextern modifierats för att användas som indata

Hur ord bytts ut för att åstadkomma korrekt identifiering av stavelsestruktur visas i Tabell 4. Ord som *warily*, *brim* och *ain't* har alltså inte känts igen av programmet varför orden bytts ut mot andra ord vars antal stavelser räknas korrekt. Den fjärde exempelraden ovan behövde ej justeras då det identifierade antalet stavelser var rätt från början.

WAIY tar ej hänsyn till rader eller refränger som upprepas utan genererar helt nya rader oavsett sådan struktur. För att ytterligare efterlikna strukturen i originallåttexterna för ökad igenkänningsfaktor så har upprepade rader inklusive hela andra refrängen tagits bort från indatan, och i stället har de korresponderande raderna i utdatan kopierats för att åstadkomma originalets upprepande struktur. Detta har applicerats på tre rader i *Another one bites the dusts* refräng, samt hela dess andra refräng, och på de två rader som utgör övergången mellan verserna och refrängen, samt på hela andra refrängen vad gäller *Feels like teen spirit*. Inga upprepningar finns i den för studien förkortade versionen av *Lose yourself*.

### 4.1.3 Resultat av implementationstest

```
-----  
In particular this song  
Is sung by a human being.  
The song is sung by a human.  
The song is sung by an uneducated man who  
Has no knowledge of languages.  
The narrator is an orphan  
Who has no contact with human.  
The narrator goes on a quest  
Because the narrator is a bumann.  
Of this song he  
Is a human.  
The narrator.  
This song is a subhuman  
Creature.  
The narrator is human.  
He narrator of this song is albumin.  
The narrator of this song was.  
Also albumin is an author of the keying.  
-----
```

Figur 8 Exempel på utdatapresentationen från WAIY, med resultatet av indatan från Figur 2.

Implementationstesten gjordes med kompletterande träning av GPT-2-språkmodellen på respektive databaser, som benämns alternativ 1 till 3. Här presenteras utdatan med genomförd upprepning av de rader som upprepas i originallåttexterna enligt föregående stycke, först för *Another one bites the dust*. Vad gäller alternativ 1 så började rad 1 och 3 med skiljetecken (/ respektive <>) men dessa har raderats då stavelsestrukturen stämmer utan dem. Ett snedstreck innan sista ordet i rad 5 behövdes inte heller och raderades. I övrigt så har endast versaliseringen korrigerats där tillämpligt. Det kan konstateras att det förekommer fel vad gäller antal stavelser på vissa ställen i utdatan. Till exempel har rad 4 i både alternativ 1 och alternativ 3: 8 stavelser där originallåttexterna har 7. På samma sätt vad gäller rad 17 så har alternativ 2: 11 stavelser och alternativ 3: 12 stavelser medan originalet samt alternativ 1 har 10 stavelser. En mer noggrann kontrollräkning av antal stavelser, samt tillhörande åtgärd har gjorts på det alternativ som valts ut för implementation samt originalprogrammets version som både ingår i enkäten.

<b>Alternativ 1</b>	<b>Alternativ 2</b>	<b>Alternativ 3</b>
<p>And if it looks fun to play in The shade of that hot bus stop. And or if it sounds fun to play in Any shade of that hot bus stop Has got no name or meaning in this world if It looks fun to play in the hot bus shade. If it looks fun to play in the shade. If it looks fun to play on.</p> <p>If it is fun to play on. If it is fun to play on. If it is fun to play on the hot bus stop game. If it is fun to play on. The time was never too far If it is fun to play on.</p> <p>From an hour and twenty five minutes. His day was like a flame In the sun for three hours and twenty five. He worked his hardest because. He was one of the best in his class. He was in fire brigade Also and when the train got delayed Then he was always in Command.</p> <p>If it is fun to play on. If it is fun to play on. If it is fun to play on the hot bus stop game. If it is fun to play on. The time was never too far If it is fun to play on.</p>	<p>Am getting out of bed after A friend asked me if it was Cool and falling laughing tears laughter Shouting les petits deux pas. Jokes are always hilarious on reddit lol Was told by a customer that he lost Or tried missing her before going Out to go get her today.</p> <p>And then this happened to him. And then this happened to him. On one night he got into an argument with And then this happened to him. A man and he lost his wife. And then this happened to him.</p> <p>Am really sorry guys do things they do Not quite want to deal with. Do you say when someone asks for sex but they Do not want to go outside. Is this going to happen again. Just ok so am in aust Now but I'm still not good and there is plateauing. Try and get up to a rafter In this.</p> <p>And then this happened to him. And then this happened to him. On one night he got into an argument with And then this happened to him. A man and he lost his wife. And then this happened to him.</p>	<p>I think you have a little bit A little bit of your stuff Maybe just a little bit on it But you just get around and slough In place before me if anyone else gets On the bus that is going to do it. If anyone else comes onto it. Ave my hand down that</p> <p>In the world of magic is In the world of magic is The most awesome and beautiful thing ever made. In the world of magic is I think that is magical. In the world of magic is</p> <p>If you want to find out what magic means. What spell is to be laid. What spell is to be weaved and formed into one thing. And is it your best friend then That would be it or is your best friend. But is it a piece of shit. What is one thing in my little mit Ten and then you will quit Your name</p> <p>In the world of magic is In the world of magic is The most awesome and beautiful thing ever made. In the world of magic is I think that is magical. In the world of magic is</p>

Tabell 5 Resultat av implementation av de tre alternativen på Another one bites the dust

I *Smelling like teen spirit* finns en övergång mellan verserna och refrängen som innehåller fyra upprepningar av samma rad men där de sista två stavelserna i den sista upprepningen överröstas av att refrängen börjar och därför inte hörs. På samma sätt har de sista två stavelserna i programmets utdata tagits bort. I Alternativ 1 inleddes rad 3 av ett skiljetecken | som här har raderats.

<b>Alternativ 1</b>	<b>Alternativ 2</b>	<b>Alternativ 3</b>
<i>This song what it the you have. And how many things you have in View as many of these in your Home and workplace as you need but. The people that will be know you. The people that will be know you. The people that will be know you. The people that will be...</i>	<i>Qualified as any one This could be one of those things that One can take away from this song. It could be one of those things like. Qualified as any one this. Qualified as any one this. Qualified as any one this. Qualified as any...</i>	<i>Raises the ability With all its inhabitants to In general terms of playing A guitar with your hands at rest. Playing the guitar with hands at. Playing the guitar with hands at. Playing the guitar with hands at. Playing the guitar with...</i>
<i>Do you know most of the people. Do you know it the people have. And how many things you have are. Any good list of what you have And what you have. How many things. What much you have. Do in this song Mings The most</i>	<i>Just a couple of words like what. The title of this song was just. Sex has changed all over again Forever with every thrust And every. That you make it. Sell the phone to, someone emit The man</i>	<i>Plays the guitar with hands at rest. The guitar with your hands at rest. Ip was a musical word coined When the index finger was pressed Against the thumb. All you had to, do was to play. After a few To go</i>
<i>From its lyrics as if they Have crept in and are there to stay. Not all Mings the most from its lyrics Even if those lyrics pop in The people that will be know you. The people that will be know you. The people that will be know you. The people that will be...</i>	<i>Know that you have a good guy Hanging around this is a dry Was my mom telling me when she Realized what she was looking at. Qualified as any one this. Qualified as any one this. Qualified as any one this. Qualified as any...</i>	<i>Past the initial level. It are these common mistakes that bedevil Many guitar players today. Are famous for playing way too Playing the guitar with hands at. Playing the guitar with hands at. Playing the guitar with hands at. Playing the guitar with...</i>
<i>Do you know most of the people. Do you know it the people have. And how many things you have are. Any good list of what you have And what you have. How many things. What much you have. Do in this song Mings The most</i>	<i>Just a couple of words like what. The title of this song was just. Sex has changed all over again Forever with every thrust And every. That you make it. Sell the phone to, someone emit The man</i>	<i>Plays the guitar with hands at rest. The guitar with your hands at rest. Ip was a musical word coined When the index finger was pressed Against the thumb. All you had to, do was to play. After a few To go</i>

Tabell 6 Resultat av implementation av de tre alternativen på Smelling like teen spirit

I *Lose yourselfs* utdata från Alternativ 1 förekommer olika skiljetecken (|, \ och /) totalt 4 gånger på rad 3, 9 och 10. All utdata från Alternativ 1 i implementationstesterna uppvisar alltså en avvikande användning av skiljetecken. Totalt 7 skiljetecken som inleder respektive rader har identifierats i de tre utdatalåttexterna, och en genomgång av databasen som använts för Alternativ 1 visar att den för normal text innehåller onormalt många skiljetecken. Inget försök har gjorts för att rensa bort dessa under inläsningen av databasen. Även databasen som tillhör Alternativ 3 innehåller upprepad användning av vissa skiljetecken men då den tekniska lösningen för att läsa in denna databas byggts från grunden så inkluderades en teckenkontroll som raderar alla skiljetecken förutom komman, punkter och enkla apostrofer.

<b>Alternativ 1</b>	<b>Alternativ 2</b>	<b>Alternativ 3</b>
<p>I. I at one. From the Have you seen me before as A child or as an adult in the past. The other I. I at another. The other I at.</p> <p>It I at one but it is not here for your eyes To see at the present i am not here for your eyes now. I am the other one which is not here for your eyes now. I am the other one which is not here. The other I at is here for your eyes. I am the other one which is here to see. The other I at is here for your eyes now. I am here for your eyes to see how. The other I at is here for your eyes to see how. The other I at is here. Its but it is not here for your eyes to see it Is here for your eyes to see how in the present. The other I at is here for your eyes to see. In the present it is here for your eyes to see how and when it's. The other I at is here for your eyes to see. The other I at is here for your eyes to see how it.</p> <p>It's so let us supply we are ready to supply. It's so let us get we are ready to get. When and it's so let us hit we are ready to hit. It when and it's so let us hit is all about.</p> <p>Its so let us supply we are ready to supply. Its so let us get we are ready to get. When and its so let us hit we are ready to hit. It when and its so let us hit is all about.</p>	<p>Was All the guys Again. Was all the one and only Guys again but they were all together. And they were all. He was all the one. And all was the one.</p> <p>Says His name was Michael J of the NBA. Says his name was Michael J of the NBA and then. That they are all together again as Michael J team. And then they are all together again. It will be your favorite day of all Of your life in a million years forever. Can anyone tell how many cigarettes Do you smoke every day my bets That it is about five to six cigarettes daily. Your father was in jail for Two or three days when he came home so you have to vazquez When I call to tell you that my father is not. But my father was in jail for two days when his. His father and then that day when was in jail for another two. He was in jail for three days when his father came. My father was in jail for three days when his dad came home.</p> <p>Me coming downstairs to find you curled up in a ball Like some mercenary of the night was my moshe's. But my father has been in jail for four days you don't guez. You don't even consider it at all because.</p> <p>Me coming downstairs to find you curled up in a ball Like some mercenary of the night was my moshe's. But my father has been in jail for four days you dont guez. You dont even consider it at all because.</p>	<p>Is Not really Here yet. It's not really here yet but. Maybe its not even here yet but still. Obvious is Not so obvious. Opinion on that</p> <p>In Particular is not so clear anymore. I really am not so sure anymore about saying. Yes I need more time to be in touch with my feelings but. To read more about me and see how i Feel about this in the comments below. To read more about me and see how i feel. This in the comments below it will also. Yes and also it will also I see what i am going to say about you the Has more time than usual Which has more time for your emotions to come in. If you have any questions ask about it here. To read more about me in the comments below. To read more about me and see how i feel the most important. To make sure that you are in touch with your feelings. Make sure that there are plenty of people who will listen</p> <p>You and that there are plenty of people who will whiz Around so that your emotions will come in. What is most important to make sure that you are in Touch with your feelings is for that reason of course.</p> <p>You and that there are plenty of people who will whiz Around so that your emotions will come in. What is most important to make sure that you are in Touch with your feelings is for that reason of course.</p>

Tabell 7 Resultat av implementation av de tre alternativen på Lose yourself

#### **4.1.4 Slutgiltig implementation**

En pilotenkätundersökning med tillhörande samtal där 4 individer ingick genomfördes för att fastställa enkätfrågorna, tillhörande information och för att få enkel kvalitativ återkoppling på de tre implementationsalternativen. Den erhållna informationen tillsammans med författarens egen uppfattning låg till grund för att kunna fastställa alternativ 3 för implementation och inkludering i den slutgiltiga enkäten. För att säkerställa att låttexterna som presenteras i enkäten kan sjungas i takt med respektive melodi, gjordes en stavelseräkning.

Stavelseräkningens resultat presenteras sist i denna sektion, men i den framgår att utdatan från Alternativ 3 innehåller totalt 7 fel i de tre låttexterna. För att underlätta för respondenterna i enkäten att kunna sjunga låttexterna i rytm med melodin så har texten korrigerats för att åstadkomma korrekt antal stavelser. Justerade rader är markerade med avslutande asterisk. Valet av ord att lägga till eller ta bort har gjorts med målet att minimera den kontextuella förändringen.

De tre parodilåttexter som är genererade med alternativ 3 presenteras i Tabell 8 där låttexterna justerats enligt ovan. De rader där stavelseantalet stämmer har ej ändrats. För att se vilka ord som är tillagda eller borttagna på grund av fel antal stavelser, jämför med föregående sektionens presentation av den ojusterade utdatan från Alternativ 3.

<b>Another one bites the dust</b>	<b>Feels like teen spirit</b>	<b>Lose yourself</b>
<p><i>I think you have a little bit</i></p> <p><i>A little bit of your stuff</i></p> <p><i>Maybe just a little bit on it</i></p> <p><i>But you get around and slough*</i></p> <p><i>In place before me if anyone else gets</i></p> <p><i>On the bus that is going to do it.</i></p> <p><i>If anyone else comes onto it.</i></p> <p><i>Ave my hand down that</i></p> <p><i>In the world of magic is</i></p> <p><i>In the world of magic is</i></p> <p><i>The most awesome and beautiful thing ever made.</i></p> <p><i>In the world of magic is</i></p> <p><i>I think that is magical.</i></p> <p><i>In the world of magic is</i></p> <p><i>If you want to find out what magic means.</i></p> <p><i>What spell is to be laid.</i></p> <p><i>What spell is to be weaved into one thing.*</i></p> <p><i>And is it your best friend then</i></p> <p><i>That would be it or is your best friend.</i></p> <p><i>But is it a piece of shit.</i></p> <p><i>What is one thing in my little mit</i></p> <p><i>Ten and then you will quit</i></p> <p><i>Your name</i></p> <p><i>In the world of magic is</i></p> <p><i>In the world of magic is</i></p> <p><i>The most awesome and beautiful thing ever made.</i></p> <p><i>In the world of magic is</i></p> <p><i>I think that is magical.</i></p> <p><i>In the world of magic is</i></p>	<p><i>Raises the ability</i></p> <p><i>With all its inhabitants to</i></p> <p><i>In general terms of playing</i></p> <p><i>A guitar with your hands at rest.</i></p> <p><i>Playing the guitar with hands at.</i></p> <p><i>Playing the guitar with hands at.</i></p> <p><i>Playing the guitar with hands at.</i></p> <p><i>Playing the guitar with...</i></p> <p><i>Plays the guitar with hands at rest.</i></p> <p><i>The guitar with your hands at rest.</i></p> <p><i>Ip was a musical word coined</i></p> <p><i>When the index finger was pressed</i></p> <p><i>Against the thumb. All you had to, do was to play.</i></p> <p><i>After a few</i></p> <p><i>To go</i></p> <p><i>Past the initial level.</i></p> <p><i>It are these common mistakes that bedevil</i></p> <p><i>Many guitar players today.</i></p> <p><i>Are famous for playing way too</i></p> <p><i>Playing the guitar with hands at.</i></p> <p><i>Playing the guitar with hands at.</i></p> <p><i>Playing the guitar with hands at.</i></p> <p><i>Playing the guitar with...</i></p> <p><i>Plays the guitar with hands at rest.</i></p> <p><i>The guitar with your hands at rest.</i></p> <p><i>Ip was a musical word coined</i></p> <p><i>When the index finger was pressed</i></p> <p><i>Against the thumb. All you had to, do was to play. After a few</i></p> <p><i>To go</i></p>	<p><i>Is</i></p> <p><i>Not really</i></p> <p><i>Here yet.</i></p> <p><i>It's not really here yet but.</i></p> <p><i>Maybe its not even here yet but still.</i></p> <p><i>Obvious is</i></p> <p><i>Not so obvious.</i></p> <p><i>Opinion on that</i></p> <p><i>In</i></p> <p><i>Particular is not so clear anymore.</i></p> <p><i>I really am not so sure anymore about saying.</i></p> <p><i>Yes I need more time to be in touch with my feelings but.</i></p> <p><i>To read more about me and see how i</i></p> <p><i>Feel about this in the comments below.</i></p> <p><i>To read more about me and see how i feel.</i></p> <p><i>This in the comments below it will also.</i></p> <p><i>Yes and also it will also</i></p> <p><i>I see what i am going to say about you the</i></p> <p><i>Has more time than usual</i></p> <p><i>Which has more time for your emotions to come in.</i></p> <p><i>If you have any questions ask about it here.</i></p> <p><i>To read more about me in the comments below.</i></p> <p><i>To read more about me and see how i feel the most important.</i></p> <p><i>To make sure that you are in touch with your feelings.</i></p> <p><i>Make sure that there are plenty of people who will listen</i></p> <p><i>You and that there are plenty of people who will whiz</i></p> <p><i>Around so that your emotions will come in.</i></p> <p><i>What is most important to make sure that you are in</i></p> <p><i>Touch with your feelings is for that reason of course.</i></p> <p><i>You and that there are plenty of people who will whiz</i></p> <p><i>Around so that your emotions will come in.</i></p> <p><i>What is most important to make sure that you are in</i></p> <p><i>Touch with your feelings is for that reason of course.</i></p>

Tabell 8 Alternativ 3 med tillhörande utdata, med gjorda justeringar inför enkäten

Även originalprogrammets utdata för respektive låt analyserades för att verifiera stavelsestrukturen. Samtidigt korrigerades versaliseringen precis som tidigare gjorts på utdatan från Alternativ 1-3. Där stavelseantalet ej stämmer så har ord lagts till eller tagits bort som för utdatan från Alternativ 3, och dessa rader har markerats med en asterisk.

<b>Another one bites the dust</b>	<b>Feels like teen spirit</b>	<b>Lose yourself</b>
<i>The line is always coming out Of the first door opening. Nobody much to complain about. After the door reopening* The line is always coming out of first door. Not much to complain about either way. Coming out of first door opening. The line's always getting.*</i>	<i>In an open letter from Finally to your girlfriend who Is already engaged to you. In another letter from my Sister and friend who is not so Sister and friend who is not so Sister and friend who is not so Sister and friend who is...*</i>	<i>R To my left Has the Is my favorite of mine And for some reason it looks like she has A lot of friends. On the left side of And she is wearing</i>
<i>After first door opening. After first door opening. Come out of first door opening and let me tell After first door opening. You something more interesting. After first door opening.</i>	<i>Emotionally involved to you. I can only hope they know that The lyrics are very close to What they are feeling and not at All confusing. The only thing In your life that Needs to be sing** Is love.</i>	<i>Black. The left side of and she is wearing black shoes. The left side of and she is wearing black pants or something. Anonymous on April* 17 at 6 pm in Houston. That is beautiful in every way In the world you can't find anything else that Of that is even close to this beautiful Piece and it is very beautiful. It is a very beautiful piece by W. In every way but just Overall very beautiful and it is back** To see W on her right hand side now as It is an additional nice touch for the piece. Were there any previous pictures or video taken at The time you saw this piece and if so what was it. Anonymous on March*</i>
<i>Come out of first door opening and let. There is not much to dwell For us except the feeling of being here And having our first time. Come out of first door opening see. Just like spotting a blue jay Here and after the door reopening* Sticking up a little snout For me After first door opening. After first door opening. Come out of first door opening and let me tell After first door opening. You something more interesting. After first door opening.</i>	<i>The only thing in this world Inside of you that needs to be whirled* Is love and joy etc. One list of songs for your debut Sister and friend who is not so Sister and friend who is not so Sister and friend who is not so Sister and friend who is...  Plays the guitar with hands at rest. The guitar with your hands at rest. Ip was a musical word coined When the index finger was pressed Against the thumb. All you had to, do was to play. After a few To go</i>	<i>25 in Houston and thank you before that ends. Anonymous is here and thank you for the fact.*+** She is and cannot wait until you can get it back. She is in the process of getting it back now. 25 in Houston and thank you before that ends. Anonymous is here and thank you for the fact.*+** She is and cannot wait until you can get it back. She is in the process of getting it back now.</i>

Tabell 9 Originalprogrammets utdata med gjorda justeringar inför enkäten

En översikt över stavelseräkningen har sammanställts. I översikten benämns originalprogrammets utdata som OP och utdatan från Alternativ 3 som A3. Utifrån denna sammanställning kan det utläsas att andelen felaktiga stavelser i utdatan är betydligt mindre än det antal fel som syntes i analysen av indatan, med ett genomsnitt över de tre låttexterna på upp till 5% felaktiga stavelser vilka återfinns i 10% av raderna. Andelen fel verkar större för originalprogrammets utdata än för utdatan från Alternativ 3, men ytterligare analys av detta faktum, dess statistiska signifikans eller bakomliggande orsaker ingår ej i omfattningen av denna studie.



Låt	<i>Another one bites the dust</i>	<i>Smells like teen spirit</i>	<i>Lose yourself</i>	Totalt
<b>Antal stavelser</b>	169	105	260	534
<b>Antal rader</b>	23	18	29	70
<b>Fel antal stavelser OP</b>	5	2	16	23
<b>Felaktiga rader OP</b>	3	1	3	7
<b>Fel antal stavelser A3</b>	3	3	1	7
<b>Felaktiga rader A3</b>	2	1	1	4
<b>Andel felaktiga stavelser OP</b>	3,0%	1,9%	6,2%	3,7%
<b>Andel felaktiga rader OP</b>	13,0%	5,6%	10,3%	9,6%
<b>Andel felaktiga stavelser A3</b>	1,8%	2,9%	0,4%	1,7%
<b>Andel felaktiga rader A3</b>	8,7%	5,6%	3,4%	5,9%

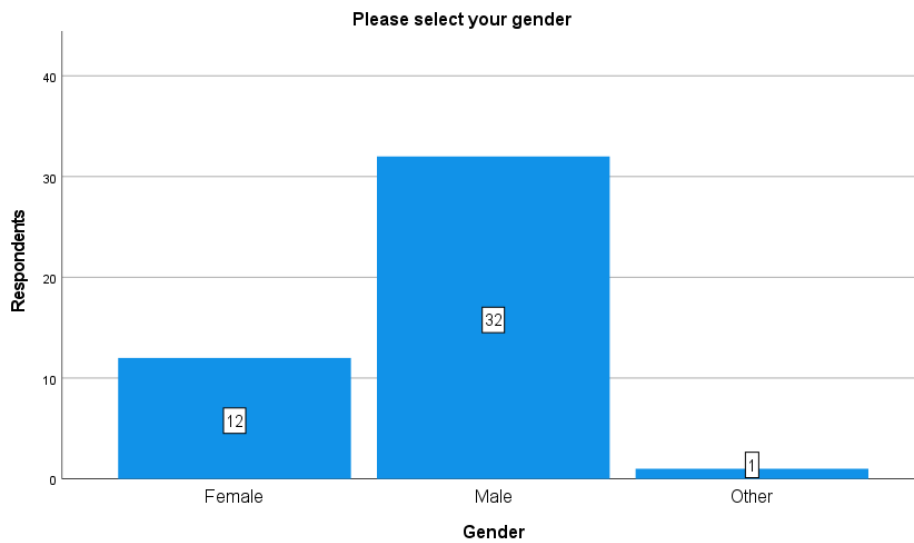
Tabell 10 Översikt över antalet fel vad gäller antal stavelser i utdatan som ska ingå i enkäten

## 4.2 Enkätundersökning

Enkätundersökningen besvarades av 52 personer. För att kunna jämföra den upplevda kvaliteten av WAIY respektive WAIY+ så sällades hela de enkätsvar bort som inte poängsatte roligheten för båda programversioners utdata från minst två av de tre låtarna. Detta resulterade i 45 användbara enkätsvar.

För att säkerställa konfidentialiteten hos enkättagarna så har ingen personlig information samlats in utöver de frågor som går igenom i 4.2.1. Då enkäten avslutades med möjligheten att skriva en kommentar i fritext vars innehåll skulle kunna anses personligt så återges ingen kommentar i sin helhet utan innehållet i kommentarerna diskuteras endast indirekt. Inget enkätsvar har raderats av etiska skäl men små demografiska grupper svar har ej analyserats. Vilka grupper detta gäller framgår av respektive stycke i 4.2.1.

### 4.2.1 Enkättagarnas demografiska data



Figur 9 Könsfördelning för de respondenter vars enkätsvar analyserats

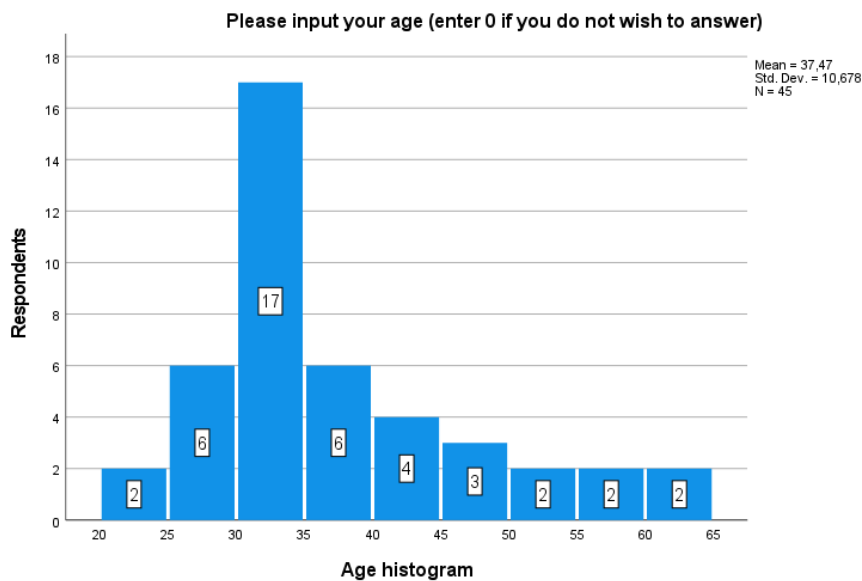
Könsfördelningen är något ojämn med en klar majoritet män som deltog. En tänkbar orsak kan vara att enkäten marknadsfördes delvis genom personlig kontakt från studieförfattaren till dennes vänner av vilka en klar majoritet var män. För att undersöka om den sneda könsfördelningen kan påverka resultatet så gjordes två t-test på den genomsnittliga skillnaden i upplevd rolighet respektive verklighetstrogenhet mellan WAIY+ och WAIY. Ett positivt värde innebär högre värdering för låttexterna som genererats av WAIY+.

	Respondent gender	N	Mean	Std. Deviation	Std. Error Mean
Diff_Funny_A_O	Female	12	,1944	,75823	,21888
	Male	30	,0944	,53010	,09678

	Respondent gender	N	Mean	Std. Deviation	Std. Error Mean
Diff_Real_A_O	Female	10	,4000	,82850	,26200
	Male	30	,2167	,60355	,11019

Tabell 11 Skillnad i upplevd rolighet respektive verklighetstrogenhet mellan kvinnor och män.

Resultatet av dessa t-test visade att det finns en viss skillnad där både män och kvinnor betygsatte låttexterna som genererats av WAIY+ högre både vad gäller rolighet och verklighetstrogenhet, men att denna skillnad var större bland kvinnors enkätsvar. Skillnaden mellan könen är dock inte statistiskt signifikant då p-värdet var 0,63 respektive 0,45.



Figur 10 Åldersfördelning för de respondenter vars enkätsvar analyserats

Vad gäller åldersfördelning så syns en tydlig obalans där åldersgruppen 30-35 år är klart överrepresenterade. T-tester utfördes på samma sätt som för könsfördelningen men med indelningen upp till 35 år och över 35 år.

**Group Statistics**

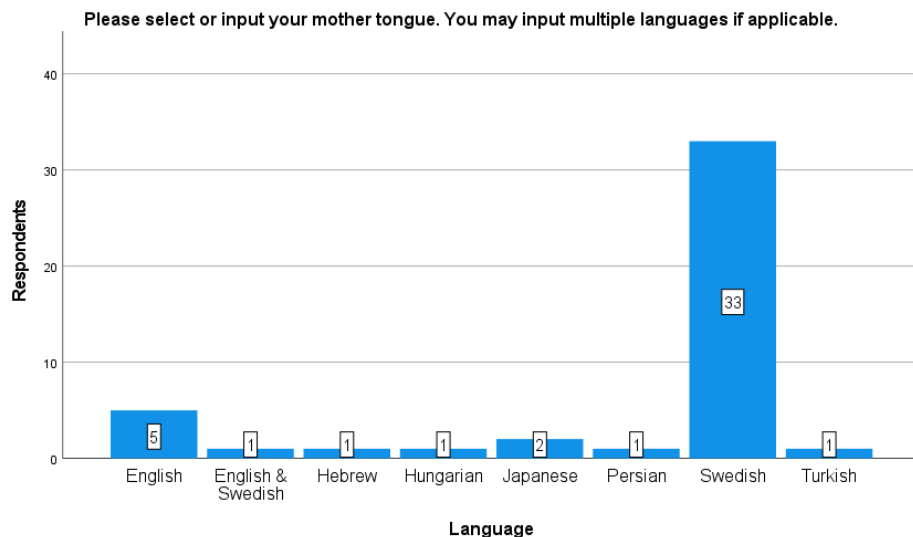
	Respondent age	N	Mean	Std. Deviation	Std. Error Mean
Diff_Funny_A_O	>= 36	17	,0196	,65850	,15971
	< 36	26	,2244	,56360	,11053

**Group Statistics**

	Respondent age	N	Mean	Std. Deviation	Std. Error Mean
Diff_Real_A_O	>= 36	17	,2157	,68406	,16591
	< 36	24	,2847	,64358	,13137

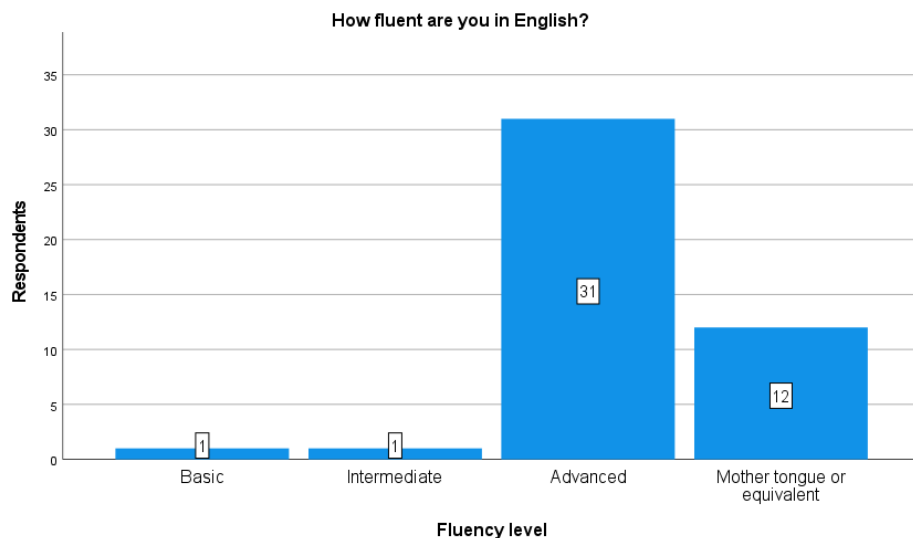
Tabell 12 Skillnad i upplevd rolighet respektive verklighetstrogenhet beroende på ålder.

Även här syns en viss skillnad i hur de två grupperna svarat, där yngre personer betygsatt WAIY+ högre med större marginal än äldre personer. Skillnaden tycks relativt liten vad gäller upplevd verklighetstrogenhet. Ingen av skillnaderna är dock statistiskt signifikanta med ett P-värde på 0,28 för upplevd rolighet och 0,74 för upplevd verklighetstrogenhet.



Figur 11 Modersmålsfördelning för de respondenter vars enkätsvar analyserats

Modersmål för drygt 70% av enkätdeltagarna är svenska, samtidigt som det näst största modersmålet engelska endast tillhör drygt 10%. Vid sidan av engelska så är inget språk representerat av fler än två individer och därmed är vidare analys av eventuella samband mellan modersmål och upplevelser av låttexter svåra att göra och skulle innefatta stora osäkerheter. Ingen sambandsanalys görs därför utifrån enkätdeltagarnas modersmål.



Figur 12 Fördelning av språkfärdighetsnivåer i engelska för de respondenter vars enkätsvar analyserats

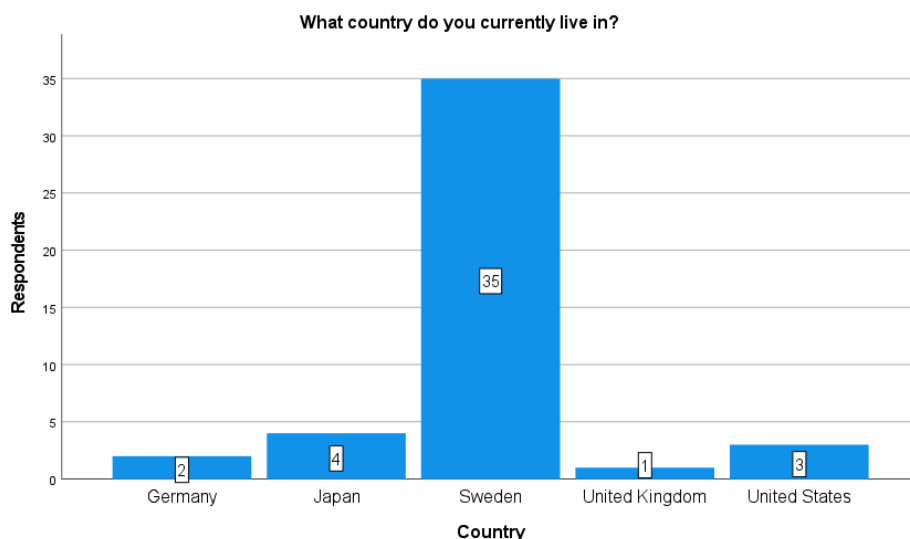
Språkfärdighetsnivån i engelska varierar i princip endast mellan dem som bedömer sig tillhöra de två högsta nivåerna. Ett t-test gjordes för att undersöka om det finns något samband mellan språkfärdighetsnivå och upplevd rolighet respektive verklighetstrogenhet, där de som svarat *mtoe* jämfördes med övriga.

Group Statistics					
English fluency level (4 = mother tongue or equivalent)		N	Mean	Std. Deviation	Std. Error Mean
Diff_Funny_A_O	>= 4	12	,2917	,83523	,24111
	< 4	31	,0860	,49229	,08842

Group Statistics					
English fluency level (4 = mother tongue or equivalent)		N	Mean	Std. Deviation	Std. Error Mean
Diff_Real_A_O	>= 4	10	-,0167	,75543	,23889
	< 4	31	,3441	,60390	,10846

Tabell 13 Skillnad i upplevd rolighet respektive verklighetstrogenhet beroende på språkfärdighetsnivå.

Detta test visade att inte heller språkfärdighetsnivå verkar påverka upplevd rolighet eller verklighetstrogenhet till en signifikant grad med p-värden på 0,44 respektive 0,13. Det kan dock nämnas att 0,13 för skillnad i upplevd verklighetstrogenhet är ett lägre p-värde än i något av de tidigare t-testen samtidigt som populationsstorleken för gruppen som svarat mtoe endast är 10 individer. Den skillnad som finns implicerar att de som bedömer sin egen engelskspråkfärdighet som högst betygsätter roligheten i WAIY+’s sångtexter högre i större grad än övriga. Samtidigt betygsätts verklighetstrogenheten högre i WAIY+ av gruppen som svarat *Basic* till *Advanced* medan de som svarat *mtoe* betygsätter de två versionerna ungefär lika.



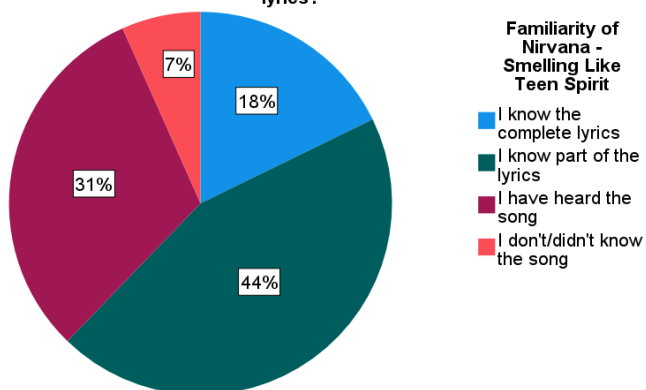
Figur 13 Fördelning av boendeland för de respondenter vars enkätsvar analyserats

De allra flesta enkätdeltagarna, 78%, bor i Sverige, och på 9% bor näst flest i Japan. På grund av den låga spridningen så anses det inte meningsfullt att djupare analysera skillnader mellan boende i de olika länderna.

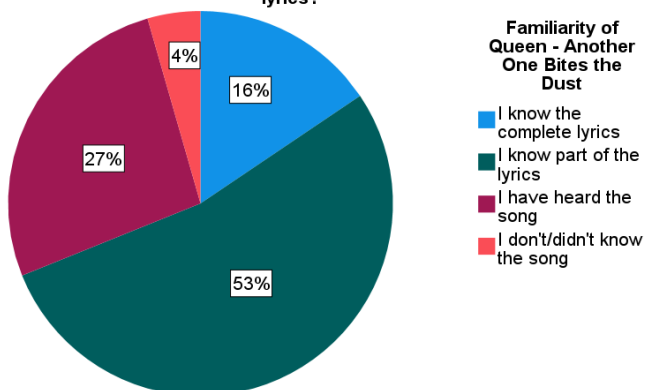
#### 4.2.2 Enkätdeltagarnas bakgrundskunskap och engagemang

Enkätdeltagarnas bakgrundskunskaper om originallåtarna undersöktes, då denna kännedom var ett mål med valet av låtar för undersökningen, för att underlätta bedömningen av låttexterna tillsammans med originalmusiken. Då samma tre originallåtar upprepas för de sju parodilåttexterna som ska bedömas, så presenteras här endast svaren på den aktuella frågan första gången varje låt dyker upp i enkäten.

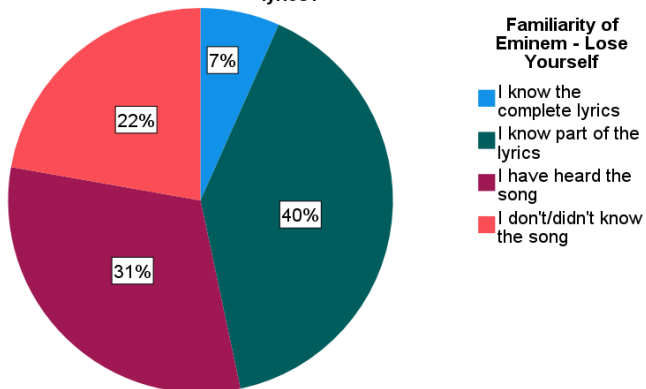
Regarding these parody lyrics - are you familiar with the original song and its lyrics?



Regarding these parody lyrics - are you familiar with the original song and its lyrics?



Regarding these parody lyrics - are you familiar with the original song and its lyrics?

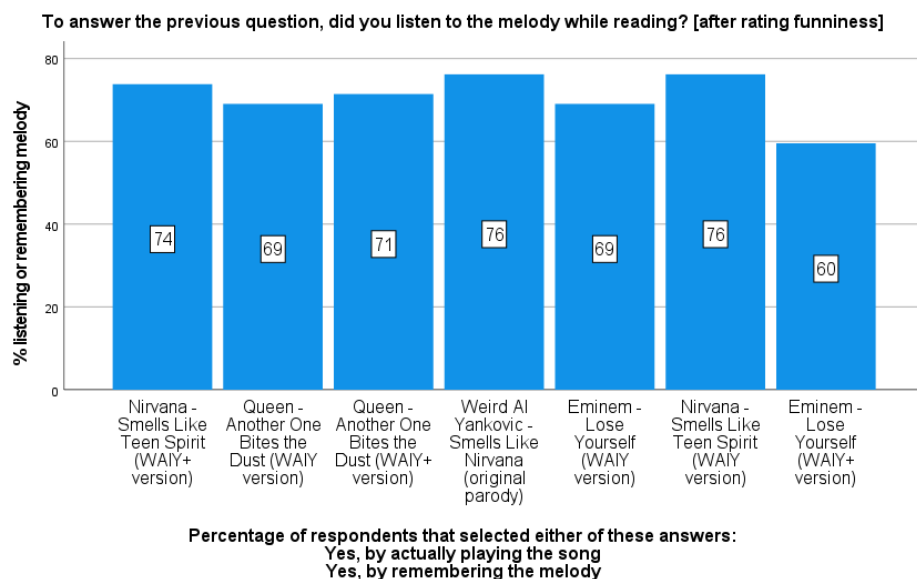


Figur 14 Kännedom om originallåtarna som enkätens parodilåttexter baseras på

Kännedomen är generellt hög med 78-96% som i någon grad kände till de respektive låtarna. Det finns dock en viss skillnad där *Lose Yourself* utmärker sig som något mindre igenkänd. Utöver kännedomen om originallåtarna mättes även enkätdeltagarnas engagemang i sitt deltagande genom att de fick ange hur de gått till väga när de bedömt de olika låttexternas rolighet.

Parodilåtttext	Nirvana - Smells Like Teen Spirit (WAIY+ version)	Queen - Another One Bites the Dust (WAIY version)	Queen - Another One Bites the Dust (WAIY+ version)	Weird Al Yankovic - Smells Like Nirvana (original parody)	Eminem - Lose Yourself (WAIY version)	Nirvana - Smells Like Teen Spirit (WAIY version)	Eminem - Lose Yourself (WAIY+ version)
Yes, by actually playing the song	11	10	10	11	10	9	5
Yes, by remembering the melody	22	21	22	23	20	24	20
No, but I checked the original lyrics	1	2	1	3	2	2	4
No, I only read the parody lyrics	11	12	12	7	13	9	14
Total	45	45	45	44	45	44	43

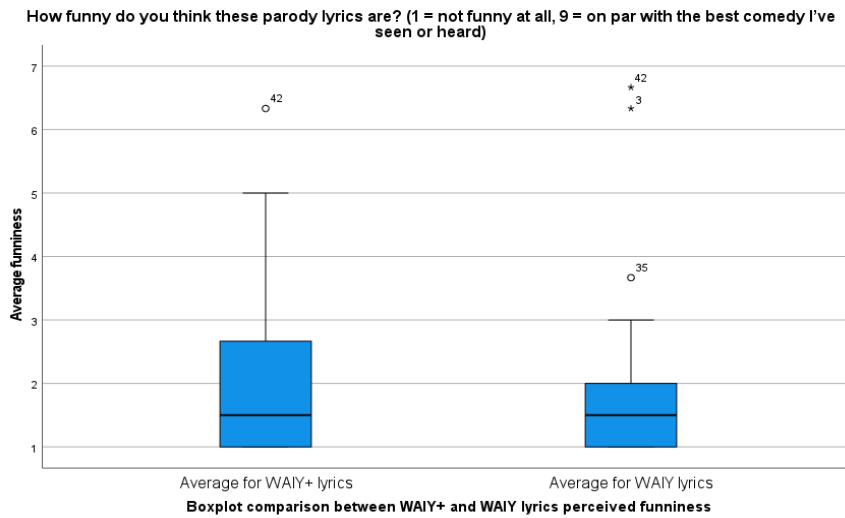
Tabell 14 Antal enkättagare som gett respektive svar vad gäller hur roligheten bedömdes



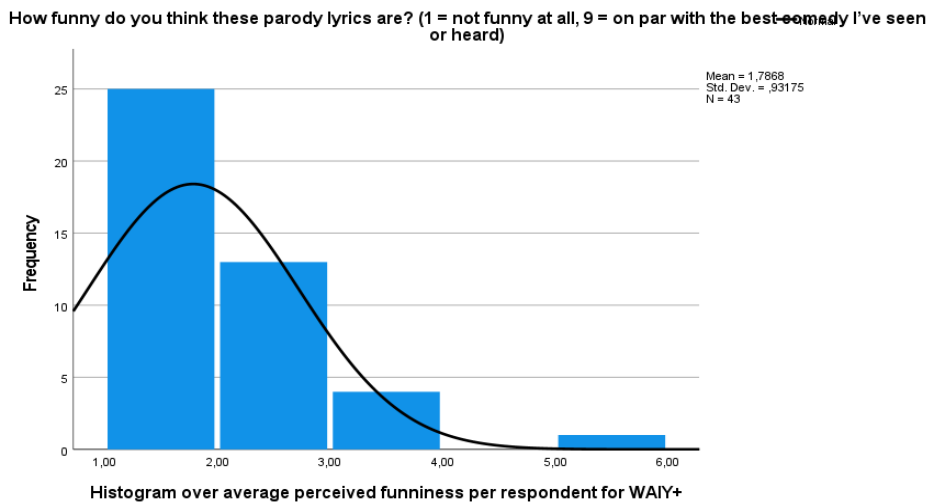
Figur 15 Andel av enkättagarna som lyssnade eller mindes melodin när roligheten bedömdes

För att erhålla ärliga och genomtänkta svar på enkätfrågorna antas det vara en fördel om enkättagarna lagt tid på att bedöma parodilåttexternas rolighet. Fördelen antas vara särskilt stor om enkättagarna föreställt sig låttextern sjungen genom att kombinera den med originalmelodin, då det är just låttextern som ska bedömas. Av Tabell 14 och Figur 15 framgår att det genomsnittliga engagemanget verkar ha varit högt då mellan 60% och 76% har lyssnat på eller kommit ihåg melodin medan de svarat på frågan om respektive parodilåttexterns rolighet.

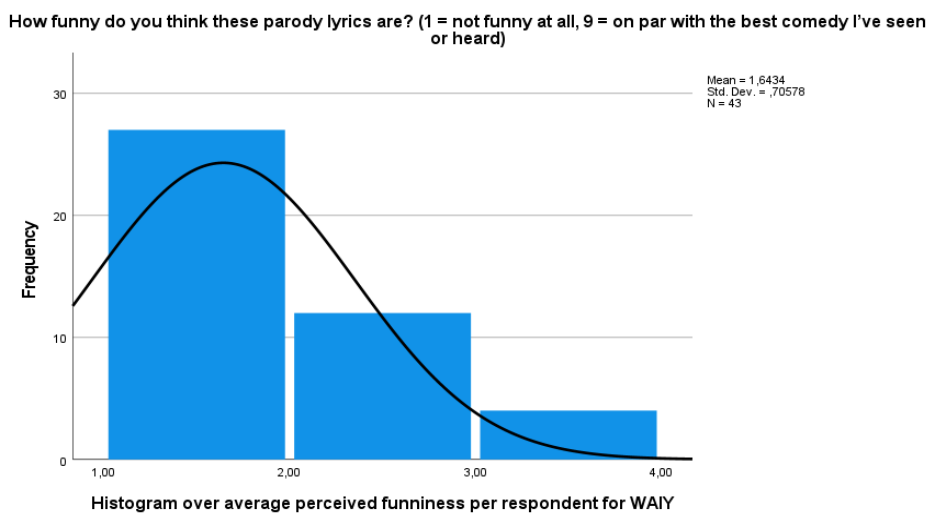
### 4.2.3 Jämförelse mellan WAIY+ och WAIY



Figur 16 Lådagram-jämförelse mellan upplevd rolighet för låttexter från WAIY+ respektive WAIY



Figur 17 Histogram över upplevd rolighet för låttexter från WAIY+



Figur 18 Histogram över upplevd rolighet för låttexter från WAIY



Figur 16, Figur 17 och Figur 18 visar en översikt över svarsfördelningen vad gäller upplevd rolighet för låttexterna från WAIY+ respektive WAIY. Varje enkättagares svar är uträknat som ett genomsnitt av de 3 låttexterna som hör till respektive programversion, eller av de 2 respektive låttexterna om personen inte svarat på samtliga frågor av denna typ. Två utläggare kan identifieras i svaren för låttexterna från WAIY, och dessa svar har därför sällats bort tillsammans med samma individers svar för låttexterna från WAIY+. Samtliga analyser rörandes upplevd rolighet inkluderar därför inte svaren från dessa individer – inte heller de analyser som gjorts tidigare i detta underkapitel. T-tester utfördes på återstående data för att få svar på underfråga 2:

2. Vad anser svarspersoner i en enkätundersökning om dessa parodilåttexter med avseende på hur roliga de är, jämfört med parodilåttexter som originalprogrammet genererat?

T-test förutsätter förutom avlägsnandet av utläggare att den beroende variabeln ska vara normalfördelad. Utifrån utseendet på histogrammen i Figur 17 och Figur 18 så kan en normalfördelad kurva delvis urskönjas men inte tydligt fastslås. Resultatet på t-testerna behöver därför läsas med viss reservation.

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	AvFunny_WAIYplus	1,7868	43	,93175	,14209
	AvFunny_WAIY	1,6434	43	,70578	,10763

Paired Samples Test									
		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	AvFunny_WAIYplus - AvFunny_WAIY	,14341	,60376	,09207	-,04240	,32922	1,558	42	,127

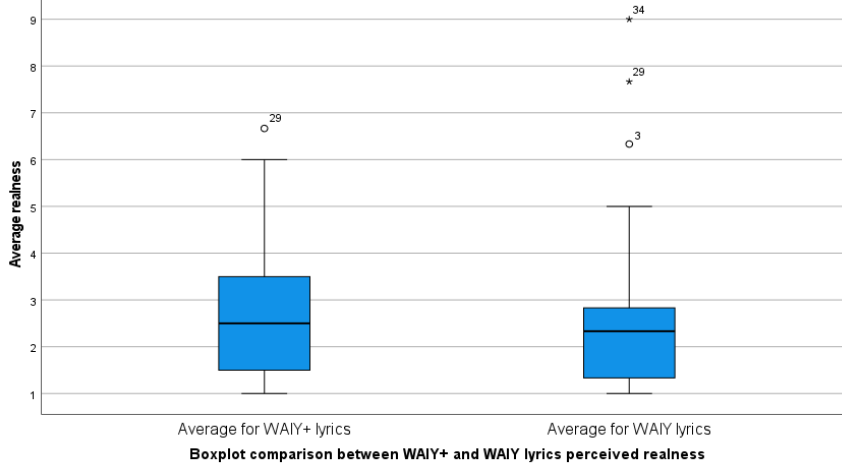
Tabell 15 Paired samples t-test på den upplevda roligheten för låttexterna från WAIY+ och WAIY

Samtidigt som det genomsnittliga värdet på den upplevda roligheten för låttexterna från WAIY+ är något högre än för låttexterna från WAIY så innebär den uträknade signifikansen på 0,127 att detta samband ej är statistiskt signifikant. Detta innebär att skillnaden i upplevd rolighet kan bero på slumpen. Nollhypotesen kan därmed inte uteslutas:

- $H_0$ : Det finns inte någon statistiskt signifikant skillnad i upplevd rolighet mellan de låttexter som genereras av WAIY+ och WAIY.

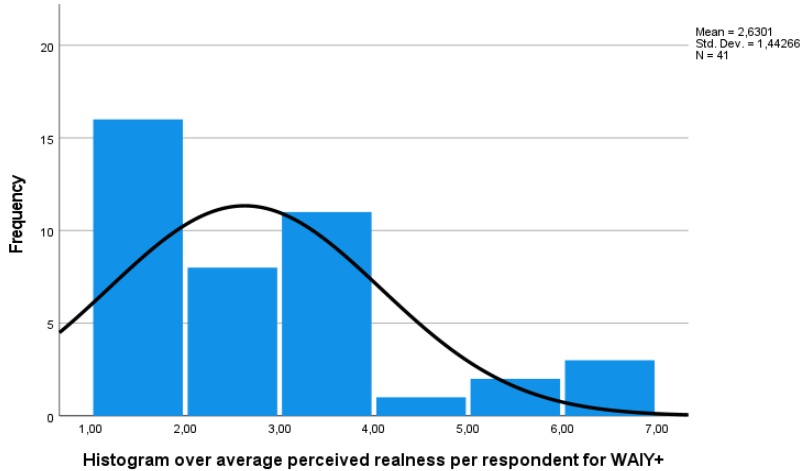
I nästa steg så jämförs den upplevda verklighetstrogenheten mellan de två programversionerna:

Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song?  
(1 = not realistic at all, 9 = I could not tell the difference)



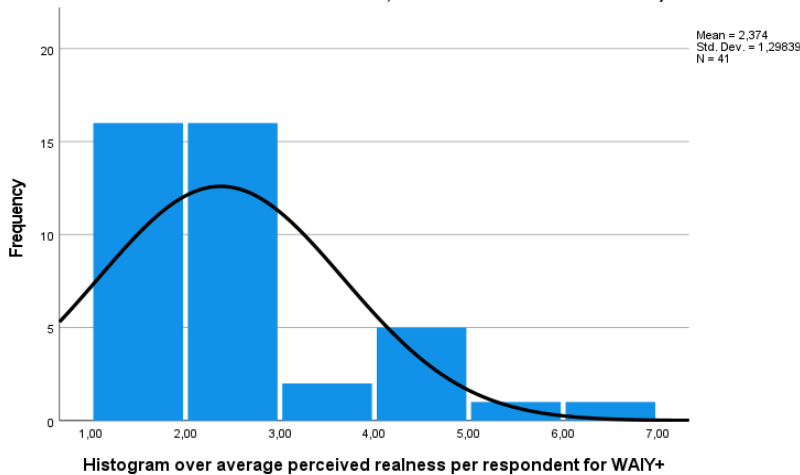
Figur 19 Lådagram-jämförelse mellan upplevd verklighetstrogenhet för låttexter från WAIY+ resp. WAIY

Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song? (1 = not realistic at all, 9 = I could not tell the difference)



Figur 20 Histogram över upplevd verklighetstrogenhet för låttexter från WAIY+

Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song? (1 = not realistic at all, 9 = I could not tell the difference)



Figur 21 Histogram över upplevd verklighetstrogenhet för låttexter från WAIY

Innan analysen gjordes sållades två enkätpersoners svarsdata bort då de inte betygsatt verklighetstrogenhet för båda programversionerna för minst 2 av de 3 låtarna. Detta är alltså utöver de sju som sorterades bort innan analysen av upplevd rolighet utfördes. Som kan utläsas från Figur 19 så finns två utliggare i datat för upplevd verklighetstrogenhet, vilka sorterats bort på samma sätt som för upplevd rolighet. Detta resulterar i 41 giltiga enkätsvar som kan analyseras. Figur 20 och Figur 21 visar den återstående svarsfördelningen vilken återigen har viss likhet med en normalfördelning utan att uppfylla alla dess kriterier. Resultaten från följande t-test behöver alltså tolkas med reservation.

		Mean	N	Std. Deviation	Std. Error Mean				
Pair 1	AvReal_WAIYplus	2,6301	41	1,44266	,22531				
	AvReal_WAIY	2,3740	41	1,29839	,20277				

		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	AvReal_WAIYplus - AvReal_WAIY	,25610	,65309	,10200	,04996	,46224	2,511	40	,016

Tabell 16 Paired samples t-test på skillnaden i den upplevda verklighetstrogenheten för WAIY+ och WAIY

Det genomsnittliga värdet som angivits för upplevd verklighetstrogenhet vad gäller låttexterna som genererats av WAIY+ är något högre än för de som genererats av WAIY. Denna skillnad har ett p-värde på 0,016 och är därmed statistiskt signifikant. En nollhypotes som påstår att det inte finns en skillnad i upplevd verklighetstrogenhet mellan låttexter från WAIY+ och WAIY skulle därmed kunna förkastas. En sådan nollhypotes skulle kunna formuleras som:

- $H_0$ : Det finns inte någon statistiskt signifikant skillnad i upplevd verklighetstrogenhet mellan de låttexter som genereras av WAIY+ och WAIY.

Då huvudfrågeställningen rör de skillnader som generellt kan observeras, så kan denna identifierade skillnad övervägas vad gäller huruvida den kan helt eller delvis utgöra ett svar på huvudfrågeställningen.

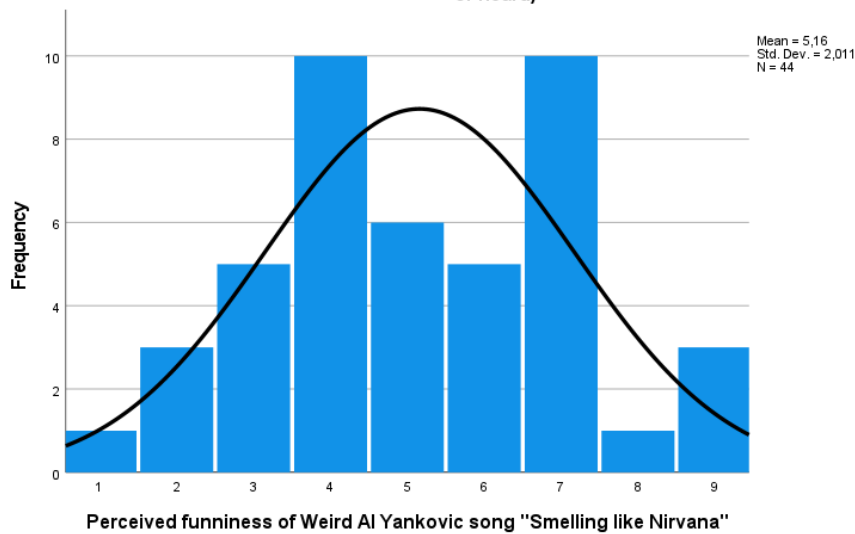
#### 4.2.4 Upplevd rolighet för verkliga parodilåttexter

Underfråga 3 lyder:

3. Vad anser svarspersoner i en enkätundersökning om datorgenererade parodilåttexter med avseende på hur roliga de är, jämfört med parodilåttexter som Weird Al Yankovic manuellt skrivit och publicerat?

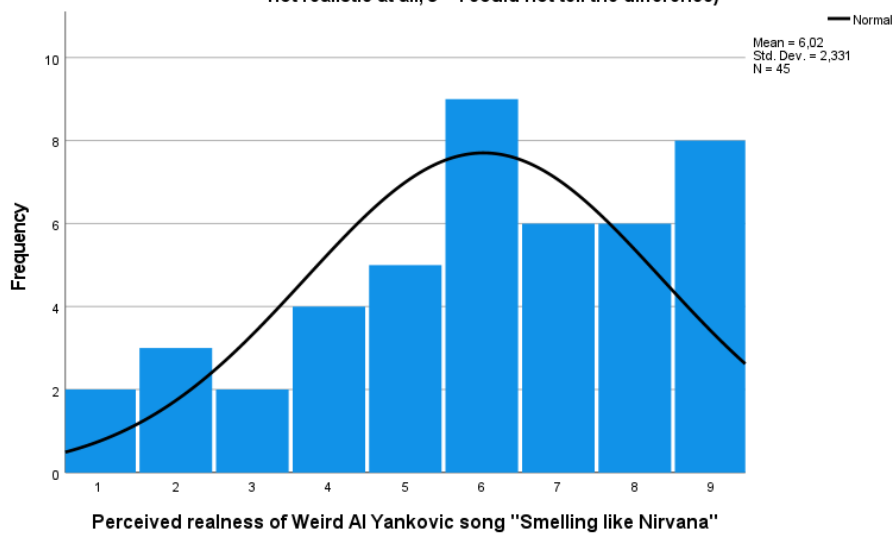
I denna enkät så utgörs Weird Al Yankovic manuellt skrivna parodilåttexter endast av ett exempel i form av *Smelling Like Nirvana* vilket är en parodi på *Smelling Like Teen Spirit*. Att detta enskilda ingångsvärde kan innebära stor osäkerhet vad gäller de slutsatser som dras om mängden av alla Weird Al Yankovic parodilåttexter, bör hållas i åtanke. Dessutom framgår i två av de manuellt skrivna kommentarerna att dessa enkätdeltagare kände igen denna låttext, vilket kan påverka analysen ytterligare. Svarsfördelningen för den upplevda roligheten presenteras nedan.

How funny do you think these parody lyrics are? (1 = not funny at all, 9 = on par with the best comedy I've seen or heard)



Figur 22 Upplevd rolighet för den äkta parodilåtttexten till Smelling like Nirvana

Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song? (1 = not realistic at all, 9 = I could not tell the difference)



Figur 23 Upplevd verklighetstrogenhet för den äkta parodilåtttexten till Smelling like Nirvana

Svaren är än en gång delvis normalfördelade och försiktighet bör därför iaktas vad gäller tolkningen av nedanstående t-test där skillnaden i upplevd rolighet och verklighetstrogenhet mellan denna äkta parodilåtttext och parodilåttexterna som genererats av WAIY+. Programversionen WAIY+ har valts för denna jämförelse då den genomsnittliga upplevda roligheten såväl som verklighetstrogenheten var något högre än för WAIY, samtidigt som dessa attribut för *Smelling Like Nirvana* är ännu högre. Detta innebär att om den uppmätta skillnaden kan sägas vara statistiskt signifikant, så gäller detta även för skillnaden gentemot WAIY samt mängden av alla AI-genererade låttexter i enkäten (WAIY+ och WAIY).

**Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Funny_Smelling_Like_Nirvana - AvFunny_WAIYplus	5,12	42	2,002	,309
	AvFunny_WAIYplus	1,8056	42	,93481	,14424

**Paired Samples Test**

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Funny_Smelling_Like_Nirvana - AvFunny_WAIYplus	3,31349	2,01391	,31075	2,68591	3,94107	10,663	41	,000

Tabell 17 Paired samples t-test på skillnaden i den upplevda roligheten

**Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Real_Smelling_Like_Nirvana - AvReal_WAIYplus	6,12	41	2,282	,356
	AvReal_WAIYplus	2,6301	41	1,44266	,22531

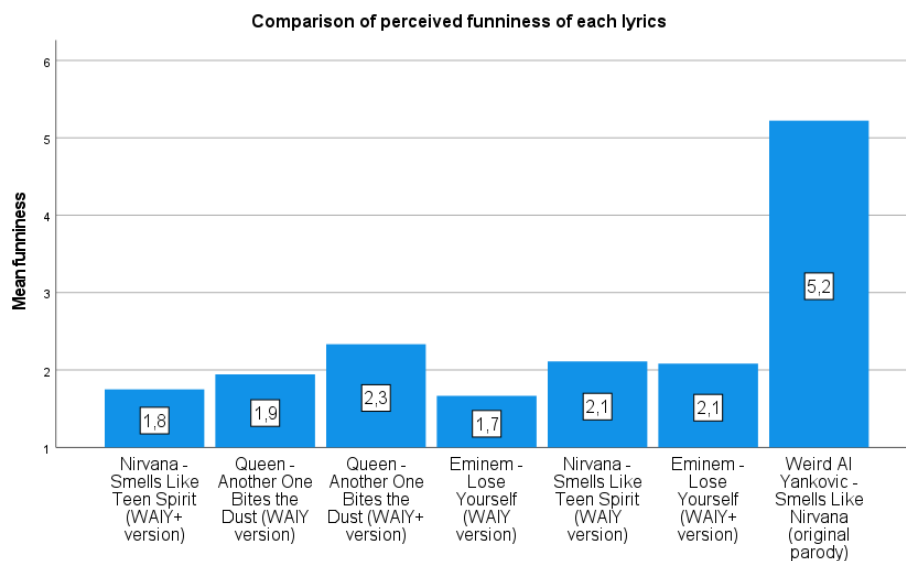
**Paired Samples Test**

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Real_Smelling_Like_Nirvana - AvReal_WAIYplus	3,49187	2,31629	,36174	2,76076	4,22298	9,653	40	,000

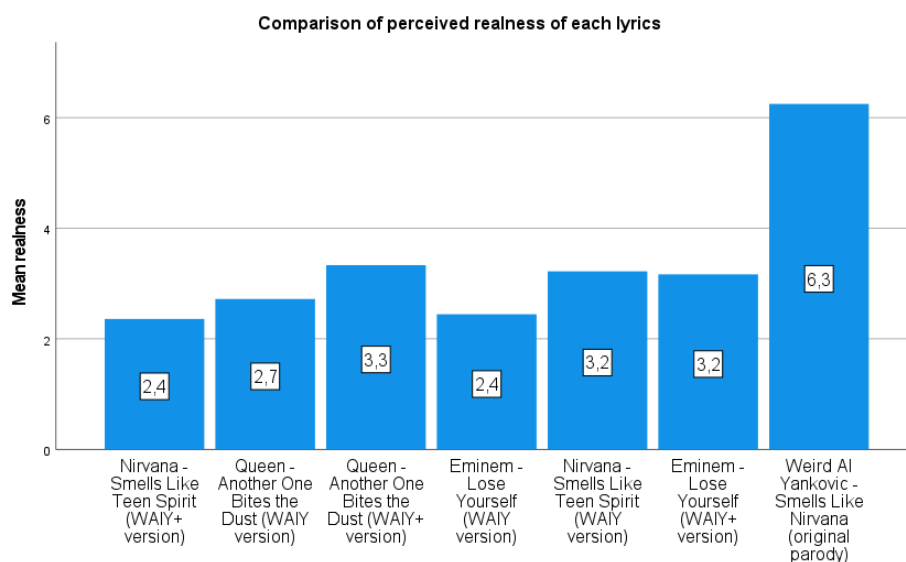
Tabell 18 Paired samples t-test på skillnaden i den upplevda verklighetstrogenheten

Testerna visar en tydligt signifikant skillnad i upplevd rolighet såväl som verklighetstrogenhet till den verkliga parodilåtttextens fördel. P-värdet är 0,000 vilket innebär minimal uppmätt sannolikhet att skillnaden beror på en slump, det vill säga mindre än 0,1% sannolikhet. Det kan utläsas att enkättagarna bedömde den verkliga låttexten som både roligare och mer verklighetstrogen än någon definierad mängd av AI-låttexterna, vilket ska tolkas i förhållande till underfråga 3 i frågeställningen.

Det genomsnittliga omdöme vad gäller rolighet respektive verklighetstrogenhet jämfördes för var och en av parodilåttexterna för att se variationen. Det kan konstateras att variationen bland de AI-genererade låttexterna är liten jämfört med skillnaden mellan någon av de AI-genererade låttexterna och den verkliga parodilåtttexten skriven av Weird Al Yankovic. Detta gäller för båda attributen rolighet och verklighetstrogenhet men skillnaden är mindre vad gäller verklighetstrogenhet.



Figur 24 Jämförelse av upplevd rolighet för varje låttext i enkäten



Figur 25 Jämförelse av upplevd verklighetstrogenhet för varje låttext i enkäten

Vad gäller omdömet om de AI-genererade parodilåttexterna så kan det konstateras att de anses vara roliga och verklighetstroga i en grad som är klart lägre än den verkliga parodilåttextern med titeln *Smells Like Nirvana* som Weird Al Yankovic skrivit. Medan *Smells Like Nirvana* har en genomsnittlig upplevd rolighet på 5,2 och verklighetstrogenhet på 6,3 så har ingen av de AI-genererade låttexterna omdömen som är i närheten. Den parodilåttextern med näst högst upplevd rolighet och verklighetstrogenhet är WAIY+-versionen av *Another One Bites the Dust* med 2,3 respektive 3,3 i genomsnittligt omdöme.

Det noteras att sambandet mellan upplevd rolighet och upplevd verklighetstrogenhet tycks starkt. Verklighetstrogenheten bedöms som 35-52% högre än roligheten för alla låttexter förutom *Smells Like Nirvana* vars verklighetstrogenhet bedöms 20% högre. Detta skulle kunna tyda på att högre verklighetstrogenhet i sig innebär högre rolighet, men ytterligare data skulle behövas för att kunna fastslå detta eventuella samband.

# 5 Diskussion

I följande sektioner diskuteras hur resultatet kan tolkas i förhållande till frågeställningen, vilka begränsningar som identifierats och hur fortsatt forskning på området kan bedrivas med denna studie som utgångspunkt.

## 5.1 Svar på frågeställningen

### Underfråga 1

1. Vilka tekniska lösningar skulle kunna implementeras i WAIY+?

Som antaget i inledningen så har litteraturgenomgången i bakgrunden (2.3) visat att det finns ett flertal olika angreppssätt för att konstruera ett program för låttextrgenerering som antingen använts för det syftet eller kan tänkas lämpa sig för uppgiften. Antalet funna studier om just generering av parodilåttexter är dock få och ingen har funnits som samtidigt beskriver funktionaliteten i detalj. Det måste därför sägas att kunskapsläget är lågt både vad gäller parodilåttextrgenerering i allmänhet och vad gäller metodik för träning av språkmodeller för denna användning i synnerhet, men trots det har värdefull kunskap erhållits kring tekniska lösningar som skulle kunna implementeras i WAIY+.

Den ursprungliga planen för examensarbetet var att utifrån de möjliga åtgärder som föreslås i bakgrundens litteraturgenomgång (2.3) undersöka med ett brett perspektiv vilka som skulle kunna implementeras i WAIY+ och därefter utveckla en eller flera sådana versioner. Brist på programmeringskunskaper inom AI samt tid innebar att själva implementationen (4.1) fokuserades på finjustering av en av de två språkmodellerna och hur de genererade parodilåttexterna därmed förändrades i jämförelse med originalprogrammet. Frågeställningens underfråga 1 har alltså besvarats i teorin men endast delvis kunnat testas i praktiken.

Bedömningen av de tre olika alternativen för implementation gjordes av författaren med viss hjälp från deltagarna i pilotenkätundersökningen. Detta innebär att valet av implementation för inkludering i enkäten ej förankrats i en konkret vetenskaplig process. Om resultatet hade visat på en statistiskt säkerställd förbättring i upplevd rolighet av utdatan från WAIY+ jämfört med WAIY så kunde detta möjligen ses som mindre betydelsefullt.

### Underfråga 2

2. Vad anser svarspersoner i en enkätundersökning om dessa parodilåttexter med avseende på hur roliga de är, jämfört med parodilåttexter som originalprogrammet genererat?

En enkätundersökning gjordes för att kunna dra slutsatser om resultatet av vidareutvecklingen i och med WAIY+, men oklarheter finns angående enkättagarnas representativitet (se nästa sektion). Då ingen konkret urvalsgrupp definierats för enkätundersökningen och därmed inga slutsatser dragits utifrån resultatet på någon population så anses dessa potentiella brister vara av mindre betydelse för studiens slutsatser. Med ett liknande urval av deltagare som de som deltog i denna studie kan därmed reliabiliteten anses vara godtagbar. Med reservation för dessa potentiella snedfördelningseffekter så har ingen ökning vad gäller upplevd rolighet för de låttexter som WAIY+ genererat kunnat konstateras. Det innebär att svaret på frågeställningens underfråga 2 blir att ingen signifikant skillnad kunnat observeras.

Det i inledningen formulerade problemet talar om potentialen för förbättring vad gäller bland annat struktur och konsekvens. Förutom att en parodilåtttext bör eftersträva att vara rolig så kan det också antas att det är en fördel om den är språkligt strukturerad och konsekvent. Ett försök att mäta detta har gjorts i och med enkätfrågan som rör upplevd verklighetstrogenhet som låtttext. Den upplevda verklighetstrogenheten har också kunnat konstateras vara högre för de låttexter som genererats av WAIY+ jämfört med WAIY, till en statistiskt signifikant grad. Detta bör kunna anses delvis ha löst grundproblemet.

Pilotenkätundersökningen vars deltagare intervjuades efteråt tillsammans med den feedback som 12 av deltagarna i huvudenkäten gav tyder på att svaren överensstämmer med vad enkäten var menad att undersöka. Ingen ansåg att frågorna var svåra att förstå och endast en deltagare kommenterade att det var otydligt vad som skulle göras. Denna person hade heller inte svarat på frågorna om låttexternas rolighet eller verklighetstrogenhet och därför sällades hela enkätsvaret bort i enlighet med villkoren som ställts upp i resultatets sektion enkätundersökning. Med detta i åtanke kan enkätundersökningen anses ha hög validitet.

### **Underfråga 3**

3. Vad anser svarspersoner i en enkätundersökning om datorgenererade parodilåttexter med avseende på hur roliga de är, jämfört med parodilåttexter som Weird AI Yankovic manuellt skrivit och publicerat?

Jämförelsen mellan omdömena av de AI-genererade låttexterna och den verkliga parodilåttextern *Smelling Like Nirvana* gör gällande att det är en statistiskt signifikant skillnad som innebär att den verkliga låttextern bedöms både roligare och mer verklighetstrogen. Skillnaden tycks stor där skillnaden i rolighet mellan den högst bedömda AI-genererade låttextern och den verkliga låttextern är betydligt större än skillnaden i rolighet mellan den lägst och den högst bedömda AI-genererade låttextern. Detta utgör svaret på frågeställningens underfråga 3.

Det kan verka motsäga det konstaterande som Gatti et al. (2017) gör att 6 av 10 parodilåttexter som deras To Sing Like a Mockingbird (hädanefter TSLM) genererar anses vara roliga. En viktig skillnad är dock att låtarna som genereras av TSLM behåller en stor del av originalsångtexten vilket bör underlätta igenkänningsfaktorn, höja den grammatiska korrektheten och möjliggöra en högre kontextuell konsekvens. Det kan dock nämnas att inga uttryckliga försök gjorts i vare sig TSLM eller WAIY+ att presentera igenkännliga låttexter. Samtidigt saknas data över intervjudeltagarnas bakgrund eller deras kunskaper om de låtar som de fått höra vad gäller TSLM.

Sammanfattningsvis kan sägas att frågeställningen och samtliga underfrågor har kunnat besvaras i någon grad och att resultatet kompletterar och ej motsäger resultat i andra studier inom ämnet. Den största bristen i studien anses vara att det blev nödvändigt att smalna ner det tekniska arbetet på grund av tidsbrist under examensarbetets gång, vilket också gjorde det svårare att dra slutsatser utifrån enkätundersökningen. Detta berodde delvis på bristande förkunskaper hos författaren inom ett relativt tekniskt avancerat fält, och delvis på att programvaran som hanterats varit hårdvarumässigt mycket krävande och varje körning tagit lång tid. Identifierade begränsningar i forskningsarbetet diskuteras i närmare detalj i nästa sektion.



## 5.2 Begränsningar och framtida studier

Även om de undersökta tidigare studierna saknat detaljerad beskrivning av funktionaliteten samt att den tillhörande programvaran inte har kunnat testas, så är denna inte nödvändigtvis helt otillgänglig. Inga försök har gjorts att nå studiernas upphovsmän för efterfrågan av ytterligare material samt programvaran i sig, och om resurserna tillåter så bör detta undersökas vidare i framtida forskning.

Det kan konstateras att antalet insamlade enkätsvar inte är särskilt stort vilket antas bero på de begränsade resurserna vad gäller marknadsföring av enkäten samt den korta tid som enkäten var öppen, vilket i sin tur berodde på att enkäten inte kunde färdigställas förrän implementationen var slutförd, vilket drog ut på tiden. Den demografiska fördelningen bland enkätdeltagarna är därutöver mycket snedfördelad med en överrepresentation av män, unga personer, personer med hög självbedömd engelskkunskap, svensktalande och personer som bor i Sverige.

Av dessa attribut har sambanden mellan kön, ålder och engelskkunskap och bedömningen av de presenterade låttexternas rolighet och verklighetstrogenhet analyserats. Resultatet av den analysen tyder på att en jämnare könsfördelning hade kunnat påvisa en större upplevd skillnad mellan låttexterna från WAIY+ och WAIY samtidigt som en jämnare åldersfördelning hade lett till en mindre upplevd skillnad.

Vad en fördelning av enkätdeltagare med jämnare varierande självbedömda engelskkunskaper hade inneburit är svårt att säga då underlaget av enkätdeltagare som bedömt sin språkfärdighet som *basic* eller *intermediate* är mycket liten på sammanlagt två individer. Inget underlag har heller tagits fram vad gäller fördelningen av personer med olika engelskkunskaper i en bestämd population. Inga relevanta analyser kunde heller göras på sambanden mellan modersmål och/eller boendeland och övriga svar på grund av små grupper. För framtida forskning som önskar klarlägga dessa samband och kunna dra slutsatser om en viss population bör urvalsgruppen tydligt definieras och de insamlade enkätsvarens representativitet för urvalsgruppen verifieras. Detta skulle öka reliabiliteten och framför allt generaliserbarheten som på grund av den vagt definierade urvalsgruppen i denna studie är relativt låg.

Tänkbara orsaker till att parodilåttexterna som genererats av WAIY+ uppfattades som mer verklighetstroga men inte roligare har inte författaren lyckats identifiera med mer underlag än ren spekulering. Den databas som användes med artiklar från Uncyclopedia borde å ena sidan vara bättre lämpad för komeditext än låttexter i allmänhet än till exempel den databasen som innehöll verkliga låttexter. Å andra sidan var Uncyclopediadatabasen betydligt större än låttextdatabasen och det kan tänkas att antalet låttexter helt enkelt var för litet för att åstadkomma ett verklighetstroget resultat. Lau (2021) konstaterade att minst 15000 låtar skulle behövas för en god variation. Då ingen av implementationerna som testats i denna studie uppvisat en märkbart låg variation så kan det varken sägas att Laus resultat bekräftats eller motbevisats.

Om en parodilåttexts viktigaste attribut anses vara dess rolighet, vilket har antagits i denna studie i och med valet av huvudfrågeställning, så kan inte den föreslagna implementationen rakt av användas på det sätt som var det primära syftet, det vill säga utan modifikation för att generera parodilåttexter som är roligare än de som WAIY genererar. Huruvida de andra alternativen för implementation skulle ha givit ett annat resultat kan ej utläsas i denna studie men fortsatta studier för att undersöka detta är fullt möjliga då all kod är färdigställd och finns tillgänglig i bilaga B.

Det ska noteras att urvalet av låttexter som ingår i enkäten är litet och det går ej att säga om urvalet är representativt för alla möjliga parodilåttexter som WAIY och WAIY+ kan generera, och det har inte

heller undersökts om valet av *Smelling Like Nirvana* som verklig parodilåtttext är representativt för mängden av de låtar som Weird AI Yankovich skrivit, eller för parodilåttexter i allmänhet. För att minska osäkerheten i urvalets representativitet skulle fler låtar behöva inkluderas i undersökningen. Samtidigt om enkäten upplevs lång av respondenterna så riskerar kvaliteten på den insamlade datan minska (Denscombe 2014). Detta dilemma kan lösas genom att slumpmässigt tilldela respondenterna ett mindre antal av det totala urvalet av parodilåttexter som ingår i undersökningen. Det skulle då kräva en mer avancerad teknisk plattform för enkätundersökningen, samt ett större antal respondenter så att det ändå går att statistiskt identifiera eventuella utliggare.

Fortsatt forskning på området bör vidare analysera effekterna av träning på mer specialiserade data. Fokus bör ligga på att implementera finjustering även på den språkmodell som ej berörts i denna studie, XLNet, alternativt att modifiera programmet så att det endast använder GPT-2. En sådan implementation bör sedan undersökas antingen objektivt med hjälp av till exempel lingvistiska modeller eller med kvantitativ metod. En kvantitativ undersökning behöver innehålla ett större urval av exempelutdata och troligen även ett betydligt större antal enkätsvar.

## 5.3 Slutsats

Huvudfrågeställningen löd:

- Hur förändras de parodilåttexter som WAIY genererar om språkmodellerna tränas på mer specialiserade data i form av låttexter eller parodiska texter?

Denna fråga har delvis frångåtts genom att arbetet tvingats fokusera på endast GPT-2 som är en av de två språkmodellerna som används. Frågan, om den omformuleras för att gälla en språkmodell, har besvarats genom att exempelutdata från de två olika programversionerna presenterats i 4.1.3 och 4.1.4. Ingen objektiv analys av dessa skillnader har kunnat göras då tiden ej räckt till varken för att teoretiskt och praktiskt fastslå en metodik för denna typ av analys, eller för att producera en mängd utdata som är tillräckligt stor för att kunna anses vara representativ för all utdata som programmen kan generera.

Den analys som gjorts med hjälp av en enkätundersökning fastslår att det finns en statistiskt signifikant skillnad i verklighetstrogenhet till utdatan från den vidareutvecklade versionen WAIY+’s fördel. Denna skillnad måste dock sägas innehålla stora osäkerheter och därför kan det inte fastslås huruvida den observerade skillnaden faktiskt beror på den vidareutveckling som gjordes, och därmed skulle gälla generellt för utdata från WAIY+.

Underfrågorna i frågeställningen har besvarats helt eller delvis. En litteraturgenomgång har gjorts för den vetenskapliga förankringen och denna ger en översikt över vilka tekniska möjligheter som finns för utvecklandet av program som använder AI för att generera parodilåttexter. Den kompletteras med ett genomförbarhetsresonemang som föreslår några potentiella åtgärder som skulle kunna användas för att öka den upplevda roligheten för de låttexter som genereras av WAIY. Tre föreslagna åtgärder implementerades varav en valdes ut för närmare analys med hjälp av en enkätundersökning.

Den föreslagna åtgärden verkade öka den upplevda roligheten hos de genererade låttexterna något, men ej till en statistiskt signifikant grad. Dock upplevdes verklighetstrogenheten större hos låttexterna från WAIY+, vars orsak är oklar. Ett samband mellan en låts upplevda rolighet och verklighetstrogenhet har funnits men ej heller kunnat bekräftas empiriskt. Om ökningen av verklighetstrogenheten och sambandet mellan rolighet och verklighetstrogenhet kan fastslås så skulle implementationen ändå kunna tänkas leda till roligare låttexter, även utan en direkt observerad ökning

av den upplevda roligheten. Att den verkliga parodilåttexten som inkluderades i enkätundersökningen ansågs vara betydligt roligare och mer verklighetstrogen än någon av de AI-genererade låttexterna tyder på att det fortsatt finns en stor förbättringspotential för denna typ av applikationer.

WAIY+ är ett program vars primära syfte är att generera parodilåttexter och andra ändamål har inte behandlats i denna studie eller av Riedl (2020). Men som Oliveira (2015) vidareutvecklar sitt program som genererar poesi PoeTryMe till Tra-la-Lyrics för låttextrgenerering kan en omvänd process tänkas där WAIY eller WAIY+ anpassas för att generera någon annan typ av låttextr eller poesi istället. Och då Singh (2018) konstaterar att denna typ av applikation kan utgöra hjälpmedel för låttextrskrivare, så kan funktionalitet som specifikt riktar sig till sådana användare göra programmet användbart i fler sammanhang. Dessa faktum tillsammans med det faktum att all kod är fritt tillgänglig för redigering och distribution för icke-kommersiell användning, innebär att studien har en tydlig och användbar vidareförbarhet.

Analysen av enkätundersökningen antyder att det finns en betydande skillnad i hur AI-genererade parodilåttexter och analogt skrivna parodilåttexter upplevs. Detta i kombination med en forskningsmetodik som säkerställer utomståendes konfidentialitet och integritet såsom hänsyn tagits i denna studie gör att integritetsrelaterade etiska dilemman har kunnat undvikas. Vid framtida studier inom området i det fall att teknologin möjliggör generering av parodilåttexter eller annan typ av text vars verklighetstrogenhet närmar sig eller är i nivå med vad som är manuellt skrivet, bör ytterligare etiska överväganden göras.

# Referenser

- Abhinav, M. (6 februari 2017). *Short jokes*. Kaggle.  
<https://www.kaggle.com/abhinavmoudgil95/short-jokes> (åtkomst 2021-12-11)
- Anderson, N. (18 november 2019). *Song lyrics from 6 musical genres*. Kaggle.  
<https://www.kaggle.com/neisse/scrapped-lyrics-from-6-genres> (åtkomst 2021-12-03)
- Azlyrics (u.å.). *Weird AI Yankovic Lyrics "Another One Rides The Bus"*  
<https://www.azlyrics.com/lyrics/weirdalyankovic/anotheroneridesthebus.html> (åtkomst 2021-12-28)
- Canales, K. (24 juli 2020). A researcher created a 'Weird A.I. Yancovic' algorithm that generates parodies of existing songs, and now the record industry is accusing him of copyright violations. *Business insider*. <https://tinyurl.com/2vz375rm> (åtkomst 2021-11-03)
- Denscombe, M. (2014). *The Good Research Guide (5 uppl.)*. Open University Press.
- Gatti, L., Özbal, G., Stock, O., & Strapparava, C. (2017). *To Sing like a Mockingbird*. EACL.  
<https://aclanthology.org/E17-2048/> (åtkomst 2021-11-03)
- Jamshidlou, P. & Moradi, M. (2013). Building Humor Ontology Using [www.uncyclopedia.co](http://www.uncyclopedia.co). *Conference: the 5th International Conference of Cognitive Science* (s. 220).  
<https://tinyurl.com/2xr5saxf> (åtkomst 2022-01-04)
- Johannesson, P. & Perjons, E. (2014). *An introduction to design science*. Springer International Publishing.
- Lau, W. (2021). *Application of Machine Learning Model in Generating Song Lyrics*. UCLA.  
<https://escholarship.org/uc/item/77d7c3hh> (åtkomst 2021-11-03)
- McCormick, C. (19 september 2019). *XLNet Fine-Tuning Tutorial with PyTorch*.  
<https://mccormickml.com/2019/09/19/XLNet-fine-tuning/> (åtkomst 2021-12-03)
- Oliveira, H.G. (2012) PoeTryMe: a versatile platform for poetry generation. *Proceedings of the ECAI 2012 Workshop on Computational Creativity, Concept Invention, and General Intelligence* (s. 18-23).  
<https://tinyurl.com/ycxrmesh> (åtkomst 2021-12-30)
- Oliveira, H.G. (2015). Tra-la-Lyrics 2.0: Automatic Generation of Song Lyrics on a Semantic Domain. *Journal of Artificial General Intelligence* 6. <https://sciendo.com/article/10.1515/jagi-2015-0005> (åtkomst 2021-11-03)
- Oliveira, H.G. (2020). *WeirdAnalogyMatic: Experimenting with Analogy for Lyrics Transformation*. ICCI. <https://api.semanticscholar.org/CorpusID:228097273> (åtkomst 2021-11-03)
- Open Lyrics Database (u.å.). *Github: Open Lyrics Database*. <https://github.com/Lyrics/lyrics> (åtkomst 2021-12-28)
- Riedl, M. (2020). *Weird AI Yankovic: Generating Parody Lyrics*. arXiv.  
<https://arxiv.org/abs/2009.12240> (åtkomst 2021-11-03)
- Rusia, A. (30 juni 2019) *XLNet speaks. Comparison with GPT-2*. Aman Rusia.  
<https://amanrusia.medium.com/xlnet-speaks-comparison-to-gpt-2-ea1a4e9ba39e> (åtkomst 2022-01-04)

- Sentireascoltare (15 oktober 2020). *Most Played Songs on Spotify by Decade*.  
<https://datastudio.google.com/u/0/reporting/74cda77c-f120-40f7-932f-5cdca9d2120a/page/VrXI>  
(åtkomst 2021-12-27)
- Singh, D. (2018). *AUTOMATED LYRICAL NARRATIVE WRITING*. Master's Projects. 617.  
[https://scholarworks.sjsu.edu/etd\\_projects/617/](https://scholarworks.sjsu.edu/etd_projects/617/) (åtkomst 2021-11-03)
- St-Amant, F. (8 maj 2021). *How to Fine-Tune GPT-2 for Text Generation*.  
<https://towardsdatascience.com/how-to-fine-tune-gpt-2-for-text-generation-ae2ea53bc272> (åtkomst 2021-12-03)
- Statista (juli 2021). *Share of music streaming subscribers worldwide in the 1st quarter of 2021, by company*. <https://www.statista.com/statistics/653926/music-streaming-service-subscriber-share/>  
(åtkomst 2021-12-27)
- Uncyc.org (u.å.). *Index of/*. <https://download.uncyc.org/> (åtkomst 2021-12-28)
- Vetenskapsrådet (2017). God forskningssed. Vetenskapsrådet.
- Yankovic, A. (u.å.). *Biography*. <https://www.weirdal.com/archives/biography-2/> (åtkomst 2021-11-03).

# Bilaga A – Enkät

## **Samtyckesblankett:**

### **About the survey:**

The survey is part of a bachelor's thesis about perceived quality of computer generated parody lyrics. Anyone with the link to this survey may participate, but only once per person. If you are under 18 years old, you should not participate.

### **About participation:**

Participation and answering of each question is voluntary (there is always an option to not answer the question). The survey takes 5-15 minutes depending on how quickly you read the lyrics you will be presented. For full contribution to the research, please read the questions and lyrics thoroughly and think carefully about your answer. However, if you really can't decide on an answer or don't want to answer honestly, please tick no answer instead of answering randomly. No data except your answers in the survey will be kept, and your answers will not be publicly available, except as part of a summary of the results.

### **About the study results:**

The thesis will be publicly available when passed. The survey results will present general demographic information about the respondents, general results from the lyrics questions and some analysis about relationships between demographics and lyrics judging. No analysis will be performed for demographic groups of less than 3 individuals.

### **Contact information:**

Questions about the survey may be sent to the author Daniel Andersson by e-mail [daniel54n@gmail.com](mailto:daniel54n@gmail.com) .

By proceeding, you accept that you agree to these conditions. If you change your mind while filling out the survey, please close the window/tab. Your answers will not be collected unless completing the whole survey and pressing *Submit*.

# Enkät:

## 1. Please select your gender

- Female
- Male
- Other
- No answer

## 2. Please input your age (enter 0 if you do not wish to answer)

## 3. Please select or input your mother tongue. You may input multiple languages if applicable.

- English
- Chinese
- Spanish
- French
- Japanese
- Swedish
- Danish
- Norwegian
- No answer
- Other, please specify

## 4. How fluent are you in English?

- Basic
- Intermediate
- Advanced
- Mother tongue or equivalently
- No answer

## 5. What country do you currently live in?

- Denmark
- Germany
- Japan
- Norway
- Spain
- Sweden
- United Kingdom
- United States
- No answer
- If other, please specify

# Parody lyrics evaluation

You will now be presented 7 different parody lyrics for 3 different songs (e.g. each song has several different parody versions), which may or may not be AI generated. The original lyrics or melodies are not provided here but feel free to Google them or listen to karaoke versions on Youtube or similar.

## **Nirvana - Smells Like Teen Spirit (parody #1)**

Raises the ability  
With all its inhabitants to  
In general terms of playing  
A guitar with your hands at rest.

Playing the guitar with hands at.  
Playing the guitar with hands at.

### **Chorus:**

Plays the guitar with hands at rest.  
The guitar with your hands at rest.  
It was a musical word coined  
When the index finger was pressed  
Against the thumb. All you had to, do was to play. After a few  
To go

Past the initial level.  
It's these mistakes that bedevil  
Many guitar players today.  
Are famous for playing way too

Playing the guitar with hands at.  
Playing the guitar with hands at.

### **Chorus:**

Plays the guitar with hands at rest.  
The guitar with your hands at rest.  
It was a musical word coined  
When the index finger was pressed  
Against the thumb. All you had to, do was to play. After a few  
To go

## **6. Regarding these parody lyrics - are you familiar with the original song and its lyrics?**

- I know the complete lyrics
- I know part of the lyrics
- I have heard the song
- I don't/didn't know the song
- No answer

## **7. How funny do you think these parody lyrics are? (1 = not funny at all, 9 = on par with the best comedy I've seen or heard) Select 0 if you do not wish to answer.**

## **8. To answer the previous question, did you listen to the melody while reading?**

- Yes, by actually playing the song
- Yes, by remembering the melody
- No, but I checked the original lyrics
- No, I only read the parody lyrics

## **9. Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song? (1 = not realistic at all, 9 = I could not tell the difference) Select 0 if you do not wish to answer.**



**Queen - Another One Bites the Dust (parody #1)**

The line is always coming out  
Of the first door opening.  
Nobody much to complain about.  
After the door reopening  
The line is always coming out of first door.  
Not much to complain about either way.  
Coming out of first door opening.  
The line's always getting.

**Chorus:**

After first door opening.  
After first door opening.  
Come out of first door opening and let me tell  
After first door opening.  
You something more interesting.  
After first door opening.

Come out of first door opening and let.  
There is not much to dwell  
For us except the feeling of being here  
And having our first time.  
Come out of first door opening see.  
Just like spotting a blue jay  
Here and after the door reopening  
Sticking up a little snout  
For me

**Chorus:**

After first door opening.  
After first door opening.  
Come out of first door opening and let me tell  
After first door opening.  
You something more interesting.  
After first door opening.

**10. Regarding these parody lyrics - are you familiar with the original song and its lyrics?**

- I know the complete lyrics
- I know part of the lyrics
- I have heard the song
- I don't/didn't know the song
- No answer

**11. How funny do you think these parody lyrics are? (1 = not funny at all, 9 = on par with the best comedy I've seen or heard) Select 0 if you do not wish to answer.**

**12. To answer the previous question, did you listen to the melody while reading?**

- Yes, by actually playing the song
- Yes, by remembering the melody
- No, but I checked the original lyrics
- No, I only read the parody lyrics

**13. Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song? (1 = not realistic at all, 9 = I could not tell the difference) Select 0 if you do not wish to answer.**

**Queen - Another One Bites the Dust (parody #2)**

I think you have a little bit  
A little bit of your stuff  
Maybe just a little bit on it  
But you get around and slough  
In place before me if anyone else gets  
On the bus that is going to do it.  
If anyone else comes onto it.  
Ave my hand down that

**Chorus:**

In the world of magic is  
In the world of magic is  
The most awesome and beautiful thing ever made.  
In the world of magic is  
I think that is magical.  
In the world of magic is

If you want to find out what magic means.  
What spell is to be laid.  
What spell is to be weaved into one thing.  
And is it your best friend then  
That would be it or is your best friend.  
But is it a piece of shit.  
What is one thing in my little mit  
Ten and then you will quit  
Your name

**Chorus:**

In the world of magic is  
In the world of magic is  
The most awesome and beautiful thing ever made.  
In the world of magic is  
I think that is magical.  
In the world of magic is

**14. Regarding these parody lyrics - are you familiar with the original song and its lyrics?**

- I know the complete lyrics
- I know part of the lyrics
- I have heard the song
- I don't/didn't know the song
- No answer

**15. How funny do you think these parody lyrics are? (1 = not funny at all, 9 = on par with the best comedy I've seen or heard) Select 0 if you do not wish to answer.**

**16. To answer the previous question, did you listen to the melody while reading?**

- Yes, by actually playing the song
- Yes, by remembering the melody
- No, but I checked the original lyrics
- No, I only read the parody lyrics

**17. Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song? (1 = not realistic at all, 9 = I could not tell the difference) Select 0 if you do not wish to answer.**

**Nirvana - Smells Like Teen Spirit (parody #2)**

What is this song all about?  
Can't figure any lyrics out  
How do the words to it go  
I wish you'd tell me, I don't know

Don't know, don't know, don't know, oh no  
Don't know, don't know, don't know

**Chorus**

Now I'm mumblin', and I'm screamin'  
And I don't know what I'm singin'  
Crank the volume, ears are bleedin'  
I still don't know what I'm singin'  
We're so loud, and incoherent. Boy, this oughta, bug your parents  
Yeah

It's unintelligible  
I just can't get it through my skull  
It's hard to bargle nawdle zouss  
With all these marbles in my mouth

Don't know, don't know, don't know, oh no  
Don't know, don't know, don't know

**Chorus**

Well, we don't sound like Madonna  
Here we are now, we're Nirvana  
Sing distinctly, we don't wanna  
Buy our album, we're Nirvana  
A garage band, from Seattle. Well, it sure beats, raisin' cattle  
Yeah

**18. Regarding these parody lyrics - are you familiar with the original song and its lyrics?**

- I know the complete lyrics
- I know part of the lyrics
- I have heard the song
- I don't/didn't know the song
- No answer

**19. How funny do you think these parody lyrics are? (1 = not funny at all, 9 = on par with the best comedy I've seen or heard) Select 0 if you do not wish to answer.**

**20. To answer the previous question, did you listen to the melody while reading?**

- Yes, by actually playing the song
- Yes, by remembering the melody
- No, but I checked the original lyrics
- No, I only read the parody lyrics

**21. Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song? (1 = not realistic at all, 9 = I could not tell the difference) Select 0 if you do not wish to answer.**

**Eminem - Lose Yourself (parody #1)**

R  
To my left  
Has the  
Is my favorite of mine  
And for some reason it looks like she has  
A lot of friends.  
On the left side of  
And she is wearing

Black.  
The left side of and she is wearing black shoes.  
The left side of and she is wearing black pants or something.  
Anonymous, completely anonymous on April  
17 at 6 pm in Houston.  
That is beautiful in every way  
In the world you can't find anything else that  
Of that is even close to this beautiful  
Piece and it is very beautiful.  
It is a very beautiful piece by w.  
In every way but just  
Overall very beautiful and it is back  
To see w on her right hand side now as  
It is an additional nice touch for the piece.  
Were there any previous pictures or video taken at  
The time you saw this piece and if so what was it.  
Anonymous, yes completely anonymous on march

**Chorus:**

25 in Houston and thank you before that ends.  
Anonymous is here, thank you for the fact.  
She is and cannot wait until you can get it back.  
She is in the process of getting it back now.

25 in Houston and thank you before that ends.  
Anonymous is here, thank you for the fact.  
She is and cannot wait until you can get it back.  
She is in the process of getting it back now.

**22. Regarding these parody lyrics - are you familiar with the original song and its lyrics?**

- I know the complete lyrics
- I know part of the lyrics
- I have heard the song
- I don't/didn't know the song
- No answer

**23. How funny do you think these parody lyrics are? (1 = not funny at all, 9 = on par with the best comedy I've seen or heard) Select 0 if you do not wish to answer.**

**24. To answer the previous question, did you listen to the melody while reading?**

- Yes, by actually playing the song
- Yes, by remembering the melody
- No, but I checked the original lyrics
- No, I only read the parody lyrics

**25. Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song? (1 = not realistic at all, 9 = I could not tell the difference) Select 0 if you do not wish to answer.**

**Nirvana - Smells Like Teen Spirit (parody #3)**

In an open letter from  
Finally to your girlfriend who  
Is already engaged to you.  
In another letter from my

Sister and friend who is not so  
Sister and friend who is not so

**Chorus:**

Emotionally involved to you.  
I can only hope they know that  
The lyrics are very close to  
What they are feeling and not at  
All confusing. The only thing, in your life that, needs to be sing  
Is love.

The only thing in this world  
Inside you that needs to be whirled  
Is love and joy etcetera.  
One list of songs for your debut

Sister and friend who is not so  
Sister and friend who is not so

**Chorus:**

Emotionally involved to you.  
I can only hope they know that  
The lyrics are very close to  
What they are feeling and not at  
All confusing. The only thing, in your life that, needs to be sing  
Is love.

**26. Regarding these parody lyrics - are you familiar with the original song and its lyrics?**

- I know the complete lyrics
- I know part of the lyrics
- I have heard the song
- I don't/didn't know the song
- No answer

**27. How funny do you think these parody lyrics are? (1 = not funny at all, 9 = on par with the best comedy I've seen or heard) Select 0 if you do not wish to answer.**

**28. To answer the previous question, did you listen to the melody while reading?**

- Yes, by actually playing the song
- Yes, by remembering the melody
- No, but I checked the original lyrics
- No, I only read the parody lyrics

**29. Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song? (1 = not realistic at all, 9 = I could not tell the difference) Select 0 if you do not wish to answer.**

**Eminem - Lose Yourself (parody #2)**

Is  
Not really  
Here yet.  
Its not really here yet but.  
Maybe its not even here yet but still.  
Obvious is  
Not so obvious.  
Opinion on that

In  
Particular is not so clear anymore.  
I really am not so sure anymore about saying.  
Yes i need more time to be in touch with my feelings but.  
To read more about me and see how i  
Feel about this in the comments below.  
To read more about me and see how i feel.  
This in the comments below it will also.  
Yes, also it will also  
I see what i am going to say about you the  
Has more time than usual  
Which has more time for your emotions to come in.  
If you have any questions ask about it here.  
To read more about me in the comments below.  
To read more about me and see how i feel the most important.  
To make sure that you are in touch with your feelings.  
Make sure that there are plenty of people who will listen

**Chorus:**

You and that there are plenty of people who will whiz  
Around so that your emotions will come in.  
What is most important to make sure that you are in  
Touch with your feelings is for that reason of course.

You and that there are plenty of people who will whiz  
Around so that your emotions will come in.  
What is most important to make sure that you are in  
Touch with your feelings is for that reason of course.

**30. Regarding these parody lyrics - are you familiar with the original song and its lyrics?**

- I know the complete lyrics
- I know part of the lyrics
- I have heard the song
- I don't/didn't know the song
- No answer

**31. How funny do you think these parody lyrics are? (1 = not funny at all, 9 = on par with the best comedy I've seen or heard) Select 0 if you do not wish to answer.**

**32. To answer the previous question, did you listen to the melody while reading?**

- Yes, by actually playing the song
- Yes, by remembering the melody
- No, but I checked the original lyrics
- No, I only read the parody lyrics

**33. Disregarding they being supposed to be a parody, do the lyrics look or sound like the lyrics of a real song? (1 = not realistic at all, 9 = I could not tell the difference) Select 0 if you do not wish to answer.**

**34. If you have any comments to this survey, please input here. Avoid disclosing personal information. Write preferably in Swedish, English or Japanese.**

# Bilaga B – Nyimplementerad kod

## Klass för att tokenisera sångtexter från en sångtextdatabas

```
import pandas as pd
from transformers import GPT2LMHeadModel, GPT2Tokenizer
import numpy as np
import random
import torch
from torch.utils.data import Dataset, DataLoader
from transformers import GPT2Tokenizer, GPT2LMHeadModel, AdamW, get_linear_schedule_with_warmup
from tqdm import tqdm, trange
import torch.nn.functional as F
import csv

### Import data from Drive
lyrics = pd.read_csv('/content/drive/MyDrive/Colab/lyrics-data.csv',
quoting=csv.QUOTE_MINIMAL)
#lyrics = pd.read_csv('lyrics-data.csv', quoting=csv.QUOTE_MINIMAL)
lyrics = lyrics[lyrics['Idiom']=='ENGLISH']

#Drop the songs with too long lyrics (more than 1024 tokens)
df = lyrics[lyrics['Lyric'].apply(lambda x: len(x.split(' ')) < 350)]

class SongLyrics(Dataset):
    def __init__(self, control_code, truncate=False, model_type="gpt2",
max_length=1024):

        self.tokenizer = GPT2Tokenizer.from_pretrained(model_type)

        self.lyrics = []

        testcount = 0
        for row in df['Lyric']:
            if testcount < 100:
                testcount += 1
                print(row)

                self.lyrics.append(torch.tensor(
                    self.tokenizer.encode(re.sub(r"[^a-zA-Z0-9,.\']+", '', row))
                ))
            if truncate:
                self.lyrics = self.lyrics[:20000]
        self.lyrics_count = len(self.lyrics)

    def __len__(self):
        return self.lyrics_count

    def __getitem__(self, item):
        return self.lyrics[item]

dataset = SongLyrics(df['Lyric'], truncate=False, model_type="gpt2")
```



## Klass för att tokenisera skämttexter från en skämttextdatabas

```
import re

### Import data from Drive
jokes = pd.read_csv('/content/drive/MyDrive/Colab/shortjokes.csv',
quoting=csv.QUOTE_ALL)
df = jokes.drop(columns='ID')

#Drop the jokes with too long text (after more than 1024 tokens, does not
work)
df = df[df['Joke'].apply(lambda x: len(x.split(' ')) < 350)]

class Jokes(Dataset):
    def __init__(self, control_code, truncate=False, model_type="gpt2",
max_length=1024):

        self.tokenizer = GPT2Tokenizer.from_pretrained(model_type)

        self.jokes = []

        for row in df['Joke']:
            self.jokes.append(torch.tensor(
                #Remove all special characters except ,.
                self.tokenizer.encode(re.sub(r"[^a-zA-Z0-9,.\']+", ' ', row))
            ))
        if truncate:
            self.jokes = self.jokes[:20000]
            self.jokes_count = len(self.jokes)

    def __len__(self):
        return self.jokes_count

    def __getitem__(self, item):
        return self.jokes[item]

dataset = Jokes(df['Joke'], truncate=False, model_type="gpt2")
```

## Funktion för att läsa in Uncyclopedia som xml till en lista med strängar

```
from torch.utils.data import Dataset, DataLoader
import xml.etree.ElementTree as ET

#Load xml file from Drive
tree = ET.parse('/content/drive/MyDrive/Colab/uncyclo-pages.xml')
root = tree.getroot()          #Get uppermost level of xml hierarchy

xmlpages = []

#Convert xml contents of interest to list of pages
def read_xml(retained):
    for element in root:
        if not "page" in element.tag:      #Only keep pages
            continue

        titletext = ""
        for subel in element:
            if "title" in subel.tag:
                titletext = subel.text
                break

        #Discard certain categories of pages
        if titletext == "" or "Uncyclopedia" in titletext or "User" in
titletext or "File" in titletext or "Help" in titletext or "MediaWiki" in
titletext or "Talk" in titletext or "Category" in titletext:
            continue:

        #Discard first row that only contains formatting properties
        firstrow = True
        elementtext = []

        #Extract body text from page
        for subel in element:
            if "revision" in subel.tag:
                for subsubel in subel:
                    if "text" in subsubel.tag:
                        if not subsubel.text:
                            break
                        elementtext = subsubel.text.split('\n')

        #Check each row for links, references and other unrelated contents
        for row in elementtext:
            if firstrow:
                firstrow = False
                continue
            if "References" in row or "See also" in row or "External link" in
row or "Category" in row:
                break
            if "|" in row or "http" in row or row.strip() == "":
                continue
            row = re.sub("\\'\\'\\'", "'", row)
            row = re.sub("\\'\\'", "'", row)
            if re.sub(r"^[a-zA-Z]+", "", row) == "":
                continue

        retained.append(row)
    return retained
```

## Klass för att tokenisera en färdig lista med textsidor

```
class Uncyclo(Dataset):
    def __init__(self, control_code, truncate=False, model_type="gpt2",
max_length=1024):

        self.tokenizer = GPT2Tokenizer.from_pretrained(model_type)

        self.textpages = []

        countpages = 0          #For progress feedback
        print("Total number of pages: " + str(len(xmlpages)))
        for page in xmlpages:
            countpages += 1
            if countpages % 10000 == 0:
                print("Pages processed: " + str(countpages))

            #Remove special characters except ,.
            tokenizestring = re.sub(r"^[a-zA-Z0-9,.\']+", ' ', page)

            #Crop long articles to keep the first part but make sure
            #that the retained part ends with a period
            if len(tokenizestring) > max_length:
                for i in range(max_length-1, max_length/2, -1):
                    if tokenizestring[i] == ".":
                        tokenizestring = tokenizestring[0:i]
                        break

            #Double check that a long article was actually cropped, otherwise
            #discard article
            if len(tokenizestring) > max_length:
                continue

            self.textpages.append(torch.tensor(
                self.tokenizer.encode(tokenizestring)
            ))

        if truncate:
            self.textpages = self.textpages[:20000]
            self.textpages_count = len(self.textpages)

    def __len__(self):
        return self.textpages_count

    def __getitem__(self, item):
        return self.textpages[item]

xmlpages = read_xml(xmlpages)
dataset = Uncyclo(xmlpages, truncate=False, model_type="gpt2")
```

## Funktion för att träna GPT-2 på ett tokeniserat dataset

```
#Accumulated batch size (since GPT2 is so big)
def pack_tensor(new_tensor, packed_tensor, max_seq_len):
    if packed_tensor is None:
        return new_tensor, True, None
    if new_tensor.size()[1] + packed_tensor.size()[1] > max_seq_len:
        return packed_tensor, False, new_tensor
    else:
        packed_tensor = torch.cat([new_tensor, packed_tensor[:, 1:]], dim=1)
        return packed_tensor, True, None

#Train GPT-2 on dataset
def gpttrain(
    dataset, model, tokenizer, batch_size=16, epochs=5, lr=2e-5,
    max_seq_len=400, warmup_steps=200, gpt2_type="gpt2", output_dir=".",
    output_prefix="wreckgar", test_mode=False, save_model_on_epoch=False,
):
    acc_steps = 100
    device=torch.device("cuda")
    model = model.cuda()
    model.train()

    optimizer = AdamW(model.parameters(), lr=lr)
    scheduler = get_linear_schedule_with_warmup(
        optimizer, num_warmup_steps=warmup_steps, num_training_steps=-1
    )

    train_dataloader = DataLoader(dataset, batch_size=1, shuffle=True)
    loss = 0
    accumulating_batch_count = 0
    input_tensor = None

    for epoch in range(epochs):

        print(f"Training epoch {epoch}")
        print(loss)
        for idx, entry in tqdm(enumerate(train_dataloader)):
            (input_tensor, carry_on, remainder) = pack_tensor(entry,
input_tensor, 768)

            if carry_on and idx != len(train_dataloader) - 1:
                continue

            input_tensor = input_tensor.to(device)
            outputs = model(input_tensor, labels=input_tensor)
            loss = outputs[0]
            loss.backward()

            if (accumulating_batch_count % batch_size) == 0:
                optimizer.step()
                scheduler.step()
                optimizer.zero_grad()
                model.zero_grad()

            accumulating_batch_count += 1
            input_tensor = None
        if save_model_on_epoch:
            torch.save(
```

# Bilaga C – Reflektionsdokument

Jag anser att mitt examensarbete väl svarar mot kursens mål. Den introducerar ämnet på ett tydligt sätt som intresserar läsaren och formulerar ett konkret problem. En frågeställning presenteras vars möjligheter att lösa problemställningen förklaras utförligt. Genom uppsatsen ges sedan återkommande återkoppling på frågeställningen vilket gör att läsaren naturligt kan följa hur arbetet har fortskridit med hänsyn till det ursprungliga problemet.

Metodvalet är mycket utförligt beskrivet och diskuterat. Hur metoderna i de olika delarna av forskningsarbetet interagerar är beskrivet tillsammans med djupgående resonemang kring forskningsstrategi och -metoder som i sin tur har en tydlig förankring i den vetenskapliga metoden. Alternativa metoder som valts bort har presenterats med motivering, och en konkret förklaring ges hur de valda metoderna appliceras i examensarbetet. Etiska överväganden är också väl beskrivna.

Tillämpningen har skett i enlighet med det motiverade valet av metod och forskningsstrategi. Hur arbetet har utförts har beskrivits i detalj tillsammans med komplett programmeringskod och utdata från programmet vilket möjliggör vidare forskning. Uppställda krav har uppfyllts och etiska överväganden har kunnat tillämpas i praktiken.

Resultatet är väl anknytet till frågeställningen som till stor del kan besvaras. De problem som upptäckts under arbetets gång och de beslut som därigenom fattats har tydligt beskrivits och motiverats. Slutsatserna och diskussionen är förankrad i den tidigare forskningen och resultatet med en omfattande analys av hur resultatet kan tolkas. Osäkerheter och potentiella felkällor har noggrant beskrivits.

Planeringen för examensarbetet gick förhållandevis bra och tidsplanen har i stort sett kunnat hållas. Det som försköt tidsplanen var att det tekniska programmeringsarbetet drog ut på tiden vilket gjorde att det blev bråttom med att få färdigt enkäten inför juledigheten. Enkäten öppnades därigenom också några dagar senare än planerat och först på kvällen innan julafton, vilket förkortade tiden som datainsamlingen kunde ske och möjligen var timingen precis vid jul olycklig för att erhålla maximalt antal enkätsvar. Det avslutande arbetet med slutsats, diskussion med mera har dock kunnat utföras enligt plan (delvis genom förkortad juledighet) och därmed har uppsatsen kunnat slutföras i tid. Genom mer förkunskaper inom AI-programmering eller kanske genom deltagande i någon form av professionell intensivkurs så hade troligen tiden för det tekniska arbetet kunnat kapas och kanske hade även omfattningen av det tekniska arbetet kunnat utökas med ett förbättrat resultat. Den ursprungliga planen som dock frångicks tidigt innebar implementation av fler justeringar av Riedls program än endast träningen av den ena språkmodellen. Att inte den andra språkmodellen, XLNet, tränades på samma data som GPT-2 berodde också på kunskaps- och tidsbrist vilket var olyckligt.

Examensarbetet är nära relaterat till min utbildning inom Data- och systemvetenskap. Introduktionskursen där grundläggande Pythonprogrammering lärts ut samt de efterföljande programmeringskurserna har varit användbara för arbetet med den tekniska implementationen. En AI-kurs har dock saknats och hade troligen varit till stor hjälp för att snabbare och mer effektivt kunna genomföra det tekniska arbetet. METOD-kursen som tagits omedelbart före examensarbetskursen har varit till mycket stor hjälp för metodvalsprocessen och genomförandet samt analysen av enkätundersökningen. Slutligen så har metodkursen tillsammans med de två tidigare examensarbeten

inom andra ämnesområden som jag skrivit varit viktig för generell förståelse för den vetenskapliga metoden och praktiska strategier för skrivandet av uppsatsen.

Examensarbetet skulle dels vara mycket värdefullt för eventuella framtida studier inom systemvetenskap i allmänhet och inom AI i synnerhet. Det kommer också vara användbart i mitt arbetsliv då jag dels lärt mig mer om Python och maskininlärning som används mer och mer i industrin. Den generella arbetsgången, rapportskrivande och muntliga presentationer som också är viktigt i arbetslivet har också inneburit viktiga erfarenheter.

Jag är ganska nöjd med genomförandet och resultatet av examensarbetet. Jag hade förstås en förhoppning om att faktiskt kunna säga att de modifieringar av Riedls program som gjorts under examensarbetet innebär en signifikant skillnad i hur rolig den genererade utdatan upplevs, och företrädesvis att den upplevs roligare. Samtidigt så har en signifikant skillnad konstaterats i det andra kvalitetrelaterade måtvärdet som analyserats.

Den enskilda omständigheten som känns mest olycklig är just att jag inte lyckades träna båda språkmodellerna som ingår i programmet på de databaser som jag valt ut, utan träningen har enbart utförts på den ena (GPT-2). Samtidigt så har jag utförligt analyserat vilka strategier som använts för att AI-generera låttexter och hittat en del intressanta skillnader som kan vara till användning i framtida forskning. Jag har också kunnat dra slutsatser kring den generella upplevda kvaliteten av de parodilåttexter som Riedls program genererar jämfört med en äkta parodilåt, vilket visar på potentialen i framtida förbättringsarbete.