

AI musiker - “Är det framtidens melodi?”

Alexander Johansson

Kandidatavhandling i datateknik

Handledare: Kristian Nybom

Fakulteten för naturvetenskaper och teknik

Åbo Akademi

2023

Abstrakt

TODO:

... Det är viktigt att veta hur artificiell intelligens (AI) har blivit till vad det är idag, därför förklaras bakgrunden så att man får en inblick i hur AI har utvecklats under årens lopp och hur det kan appliceras i musikgenerering. I bakgrunds kapitlet förklaras också vad dessa algoritmer använts till och vilka som är i bruk ännu idag. ...

TODO:

Nyckelord: TODO:

Innehållsförteckning

1 Inledning	4
2 Bakgrund	4
2.1 Hur AI har utvecklats under årens lopp	4
2.2 Datormusikens uppkomst samt utveckling.....	5
2.3 AI algoritmer som används än i dag.....	6
3 Musikgenerering med AI.....	7
3.1 AI algoritmer som används för att generera musik	8
3.2 Möjligheter för musikgeneration med hjälp av AI.....	9
4 Utvärdering	9
4.1 Musikala korrektheten.....	9
4.2 Hur bra musiken egentligen är.....	10
5 Sammanfattning.....	10
Litteraturförteckning.....	10

1 Inledning

TODO:

2 Bakgrund

2.1 Hur AI har utvecklats under årens lopp

Tänkande maskiner med möjligheten att skapa konst samt musik förutsågs redan under 1840-talet av Ada Lovelace, när hon nämnde att en maskin kommer i framtiden att kunna komponera olika komplexa låtar [1, p. 7]. Artificiell intelligens (AI) är ett begrepp som etablerades redan i slutet av 1940-talet i Alan Turings skrifter samt tal. I en skrift som Alan Turing skrev [2], ställer han frågan ”Can machines think?”, vilket inspirerade många att börja med utveckling av AI. För Alan Turing var det dock omöjligt att skapa dessa ”tänkande maskiner”, eftersom det under denna tidsperiod inte fanns några datorer som man kunde spara kommandon till, utan det gick bara att köra dem. AI behöver nämligen minne för att fungera på rätt sätt [1, p. 22].

Ett halvt årtionde senare, år 1956 utvecklades datorprogrammet Logic Theorist av Herbert Simon, Allen Newell och Cliff Shaw [3]. Detta är alltså världens första AI-program, som tog världen med storm. Logic Theorist var ett program som bevisade matematiska problem. Programmets framgång berodde på dess förmåga att lösa problem på ett liknande sätt som en människohjärna skulle [3]. Logic theorist är en av orsakerna till att utvecklingen av AI blev allt snabbare med åren, eftersom regeringen började investera i forskning av AI. Det tog inte länge förrän militären blev involverad, speciellt i USA, där Defense Advanced Research Projects Agency (DARPA) började finansiera utvecklingen av AI [4].

Flera mål som försöktes nå under kommande årtionden misslyckades och populariteten av AI i nyheter och i människoögon försvann. År 1997 hände ändå något som ansågs vara omöjligt. Nämligen IBM skapade en dator som besegrade den dåtida

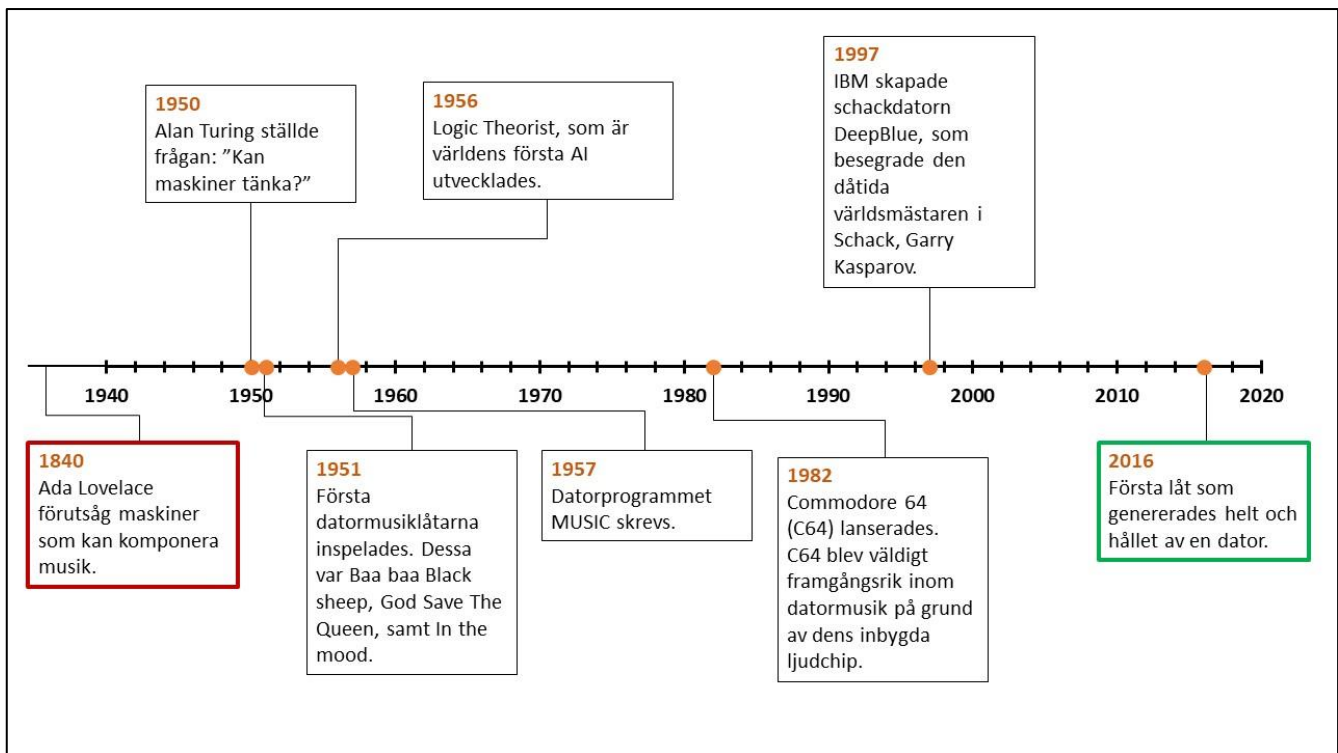
världsmästaren i schack, Garry Kasparov. Vinsten av IBMs dator gjorde att forskare blev allt mera intresserade i utveckling av AI.

2.2 Datormusikens uppkomst samt utveckling

Datormusik (eng. Computer Music) är en form av musik som genereras på dator oavsett om det komponeras av en människa eller om olika algoritmer eller AI används som hjälpmedel. Datormusik uppfanns redan på 1950-talet, redan då när första datorerna dök upp. De första datormusiklåtarna som finns inspelade [5], spelades i USA år 1951. I bandningen hör man låtarna Baa baa black sheep, God Save The Queen, samt In the Mood. Det mest revolutionära året för datormusik anses ändå ofta vara 1957, när Max Matthews skrev datorprogrammet MUSIC [6]. Med detta lyckades han spela in ett instrument till en dator för första gången. Nästa stora steg i datormusikens historia var när datorn Commodore 64 (C64) lanserades år 1982. C64 har nämligen ett inbyggt ljudchip som ger väldigt intressanta 8-bit aktiga ljud, i form av fyrkantsvågor, triangelvågor och sågtandsvågor. Detta ledde till en ny musikstil, chipmusik, som används mycket bland annat i både nya och gamla arkadspel [7].

Den första låten som genererades av en dator, utan människo hjälp, kom ut år 2016 och skapades av Googles Magenta projekt [8, p. 136].

Douglas Eck skapade Magenta med drömmen att kombinera sina två favorithobbyn, musik och AI. Han fick idén att använda Google Brain samt dess djup inlärnings anläggning för att generera låtar. Hans syn att konst och teknologi alltid har utvecklats i samma takt [8, pp. 137-144], gjorde att han valde använda öppen källkod (eng. open source). Öppen källkods programvarorna som han jobbade med på Magenta, var bland annat GitHub som är en onlinesida för att dela kod samt TensorFlow som används för att bygga och träna maskininlärningsmodeller. Detta lät programmerare samt musiker använda sig av Magentas resurser, vilket på samma gång ökade Magentas databas på låtar som kunde användas för musikgenerering [9].



Figur 1: Tidslinje för hur AI samt datormusik har utvecklats under årens lopp.

2.3 AI-algoritmer som används än i dag

Nuförtiden ser man dagligen AI överallt, på mobiltelefonen, i bilen eller på televisionen. En mobiltelefon använder sig av AI när den läser av en persons fingeravtryck eller känner igen en persons ansikte för att låsa upp mobilen. Det är också möjligt att få relevanta annonser med hjälp av AI-segmentering, vilket kan göra surfandet på nätet mer anpassat till ens intressen.

En stor del av AI-algoritmer använder sig av maskininlärning [10]. Maskininlärning betyder att en dator lär sig att utföra specifika uppgifter utan att den är hårdkodad att göra så. Huvudkategorierna i maskininlärning är övervakad inlärning (eng. supervised learning), oövervakad inlärning (eng. unsupervised learning) samt förstärknings inlärning (eng. reinforcement learning) [1]. Också djupinlärning (eng. deep learning) vilken har blivit bättre de senaste åren, tillhör maskininlärning [10]

I övervakad inlärning så har modellen tillgång till både indata och utdata. Efter det skall modellen försöka skapa en vettig utdata med den givna indata [11]. Denna modell används till exempel ofta för väderprognos där modellen kan försöka gissa vad vädret blir i framtiden till och med när vädret ändras plötsligt.

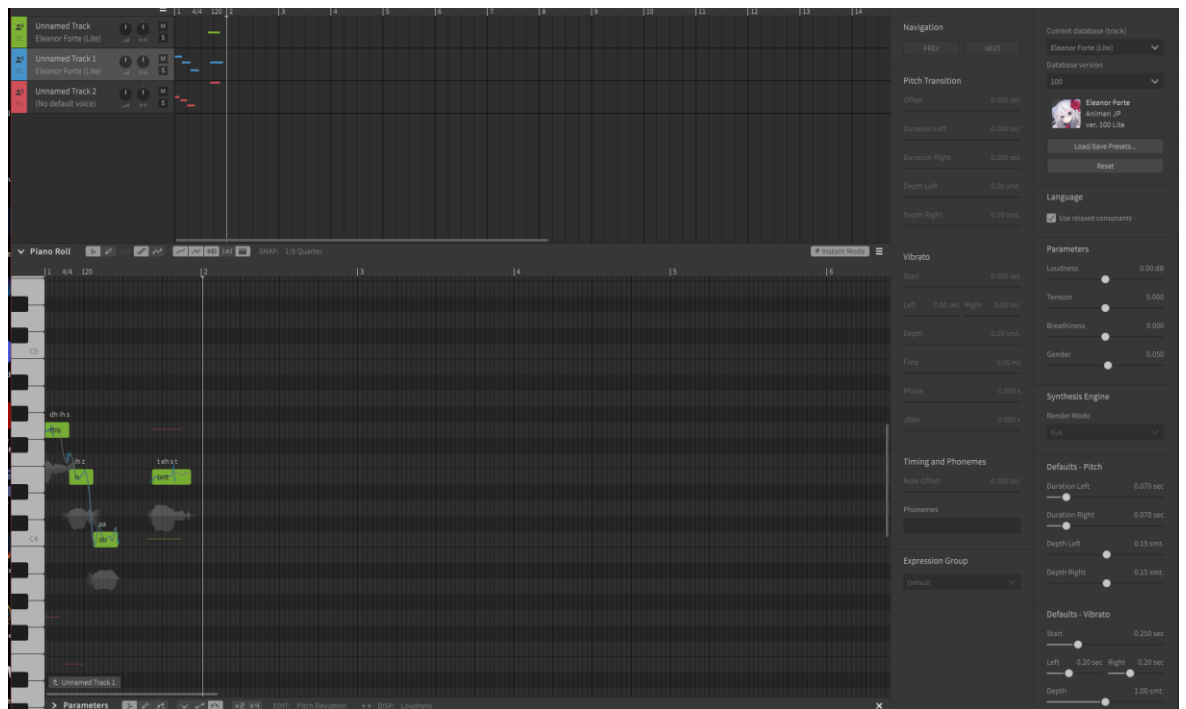
Oövervakad inlärning skiljer sig från övervakad inlärning på så vis, att oövervakad inlärning har tillgång endast till indata. Från det skall modellen försöka skapa vettiga utdata [11]. Modellen har varit väldigt framgångsrik i segmentering, när man vill hitta rätta reklamer åt rätt människa.

Som sista kategori finns det förstärknings inlärning. Detta skiljer sig från övervakad samt oövervakad inlärning, eftersom en modell som använder sig av förstärknings inlärning lär sig genom att man belönar den [12]. Förstärkningsinlärning används bland annat för att träna robotar eller AI motståndare för spel. IBM's DeepBlue använde sig av denna modell [13].

Djupinlärning är hör till maskininlärning och är ofta använd för generering av konst och musik [10]. Djup inlärning använder sig av neurala nätverk med många lager för att bearbeta data. Modellen fungerar på liknande sätt som när Ivan Pavlov utförde sitt psykologiexperiment på hundar, där han matar en hund samtidigt som han spelade någon ljudsignal. Detta ledde till att när hunden hörde en ljudsignal så började hunden producera saliv [14].

3 Musikgenerering med AI

AI-genererad musik är något som länge talats om, men har först under 2010-talet sett en tillväxt inom forskningsområdet. Användningen av olika sorters maskininlärningstekniker samt neurala nätverk i musikgenereringsprogram har lett till en stor mängd musikstycken, med likheter till låtar som är komponerade av människor. Dessa verktyg kan tillämpas på många sätt, allt från att generera en låt helt och hållet med dator, till att använda en applikation såsom en Digital Audio Workstation (DAW) som genererar musik i realtid och ger användaren friheten att justera och redigera musiken vid behov. I figur 2 kan man se en DAW som används för att generera vokaler med hjälp av AI.



Figur 2: En bild på användning av DAW:en Synthesiser V för AI-genererade vokaler. Vokalerna sjungs i ett C durackord.

Som tidigare nämntes så har AI tillämpats för många olika ändamål, men det som ändå inte talas mycket om, är hur det kan användas för musikgenerering. För att en AI skulle kunna generera musik finns det flera krav. För det första behöver modellen tillgång till en stor databas med musik för att lära sig komponera en bra låt [15]. Dessutom behöver modellen vissa regler eller algoritmer för att se till att låten inte bara blir en samling av olika ljud eller artefakter som inte passar ihop eller att musikstycket blir oharmoniskt [8]. I följande kapitel fås en liten inblick i vilka modeller som används för musikgenerering.

3.1 Modeller som används för att generera musik

Vektorkvantiserad variationsautokodare (VQ-VAE) är en modell inom djupinlärning. VQ-VAE består av en vektorkvantiserare (VQ) samt en variationsautokodare (VAE). Variationsautokodaren omvandlar data till latentvektorer som består av siffror. Latent är ett ord som används ofta i sammanhang med maskininlärning. Ordet betyder alltså något som existerar men som inte kan observeras för tillfället. Man kan ändå räkna ut observerbara variabler med hjälp av ett fåtal latent variabler. Latentvektorerna skickas sedan till vektorkvantiseraren. Den har som uppgift att minska datamängden utan att förlora alltför mycket information. Därefter avkodas (eng. decode) eller omvandlas

vektorerna tillbaka till data så noggrant som möjligt. VQ-VAE modellen lär sig med att under träningen och omvandlingsprocessen hitta den de latent variablerna som använts mest och på så vis kan gissa sig fram vilka värden som modellen i framtiden borde använda sig av. Modellen används bland annat av OpenAI:s musikgenereringsprogram Jukebox [16] [17].

TODO:

3.2 Möjligheter för musikgeneration med hjälp av AI

De program eller teorier som nämns i avhandlingen är för det mesta väldigt nya och redan nu märker man framtida möjligheter. Det finns redan nu tillgängligt flera verktyg och program för att generera, reparera samt redigera musikstycken. Verktygen varierar från enkla webbapplikationer till avancerade mjukvaror.

OpenAI som är väldigt känd för sin verksamhet med språkmodeller såsom ChatGPT, har också skapat ett musikgenererings program, Jukebox. Den använder sig av VQ-VAE modellen och tar som input, genre samt artistnamn. Med det försöker den sedan skapa ord samt vokaler baserat på artisten och en melodi baserat på genren [17].

TODO:

4 Utvärdering

Med dagens teknologi kan AI generera musik som är imponerande, men det finns ändå mycket som fattas från dessa musikstycken. Kreativitet och känslor är något som AI har problem med i musikgenerering [18].

TODO:

4.1 Musikala korrektheten

TODO:

4.2 Hur bra musiken egentligen är

TODO:

5 Sammanfattning

TODO:

Litteraturförteckning

- [1] M. A. Boden, *AI*, Oxford University Press, 2016.
- [2] A. M and Turing, "COMPUTING MACHINERY AND INTELLIGENCE," *Mind*, vol. LIX, no. 236, pp. 433-460, 1950.
- [3] L. Gugerty, "Newell and Simon\textquotesingles Logic Theorist: Historical Background and Impact on Cognitive Modeling," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, p. 880–884, October 2006.
- [4] S. Fouse, S. Cross och Z. Lapin, "DARPA's Impact on Artificial Intelligence," *AI Magazine*, vol. 41, p. 3–8, June 2020.
- [5] S. Michaels, "First Recorded computer music unveiled," *The Guardian*, 2008.
- [6] C. Roads och M. Mathews, "Interview with Max Mathews," *Computer Music Journal*, vol. 4, p. 15, 1980.
- [7] S.-Y. Su, C.-K. Chiu, L. Su och Y.-H. Yang, "Automatic conversion of Pop music into chiptunes for 8-bit pixel art," i *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.
- [8] A. I. Miller, *Artist in the Machine The World of AI-Powered Creativity*, MIT Press, 2019, p. 432.
- [9] A. McNamara, "Google's Magenta creates its first AI song," *BBC Science Focus*, June 2016.
- [10] P. P. Shinde och S. Shah, "A Review of Machine Learning and Deep Learning Applications," i *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018.
- [11] T. D. Bui, D. K. Nguyen och T. D. Ngo, "Supervising an Unsupervised Neural Network," i *2009 First Asian Conference on Intelligent Information and Database Systems*, 2009.
- [12] L. Lyu, Y. Shen och S. Zhang, "The Advance of Reinforcement Learning and Deep Reinforcement Learning," i *2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, 2022.

- [13] M. Campbell, A. J. Hoane och F.-h. Hsu, "Deep Blue," *Artificial Intelligence*, vol. 134, p. 57–83, January 2002.
- [14] J. Pan, A. Yoshikawa och M. Yamamura, "Overfitting-based Bias-generating Neural-network : Simulation of Pavlov's Dog," i *2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM)*, 2021.
- [15] A. Agostinelli, T. I. Denk, Z. Borsos, J. Engel, M. Verzetti, A. Caillon, Q. Huang, A. Jansen, A. Roberts, M. Tagliasacchi, M. Sharifi, N. Zeghidour och C. Frank, *MusicLM: Generating Music From Text*, arXiv, 2023.
- [16] A. v. d. Oord, O. Vinyals och K. Kavukcuoglu, *Neural Discrete Representation Learning*, arXiv, 2017.
- [17] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford och I. Sutskever, *Jukebox: A Generative Model for Music*, arXiv, 2020.
- [18] C.-F. Huang och C.-Y. Huang, "Emotion-based AI Music Generation System with CVAE-GAN," i *2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*, 2020.
- [19] M. Civit, J. Civit-Masot, F. Cuadrado och M. J. Escalona, "A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends," *Expert Systems with Applications*, vol. 209, p. 118190, December 2022.