|  |
| --- |
| Rapport |

Är det verkligen du?

*Rapport om deepfakes och hur man upptäcker dem.*

|  |
| --- |
| *Författare:* Oskar Andersson  *Termin:* HT 22  *Kursnamn:* Teknisk kommunikation  *Kurskod:* 1ZT010 |



Sammandrag

Skriv in abstrakt (ca 200 ord)

Nyckelord

Skriv in nyckelord

Innehåll

1 Inledning 3

1.1 Syfte och frågeställningar 3

1.2 Metod och material 3

2 Resultat 4

2.1 Underrubrik 4

2.2 Underrubrik 4

3 Diskussion och slutsatser 5

Referenser 6

# Inledning

Termen ”Deepfake” hänvisar till filmklipp eller bilder som blivit förfalskade eller på annat vis förändrade med hjälp av ”generative neural networks”, en form av artificiell intelligens (AI), där en persons ansikte och röst ersatts av en annan persons likhet. Denna teknologi har flera applikationer i kreativa och utbildande sammanhang, till exempel realistisk dubbning av videor från ett språk till ett annat [1], eller virtuell klädprovning under online-handel [2]. Deepfakes kan även användas i mer skadliga syften. År 2017 kom rapporter om att teknologin användes för att byta ut ansiktet på kvinnor i pornografiska videor med ansiktet på kända filmstjärnor [3].

Deepfakes kan ha positiva användningsområden inom underhållning av olika slag men har också negativa aspekter som skapar frågor om bedrägeri och olika mediers trovärdighet och äkthet. Teknologin har använts för att sprida falsk information i politiska syften. År 2018 släppte ett politiskt parti i Belgien en video där USA:s president Donald Trump uttryckte sin önskan att Belgien drar sig ut ur klimatavtalet från Paris. Trump gav aldrig detta tal, utan det handlade om en deepfake [4].

Med utvecklingen av deepfakes uppkommer ett behov av att lättare hitta dem. Teknologin måste oftast utvecklas för förebyggande åtgärder ska kunna skapas, vilket i sin tur skapar nya problem. I nuläget finns det flera sätt att undersöka om en video är en deepfake då man letar efter onaturliga rörelser, skillnader i ljussättning och dålig läppsynk, samt algoritmer som kan bestämma en videos autenticitet. I takt med att teknologin gör att genererad media blir svårare att skilja från verkligheten blir det ännu viktigare för företag och regeringar att utveckla preventiva åtgärder så att samhället kan hålla koll på riktig och falsk information.

## Syfte och frågeställningar

Syftet med denna rapport är att ge en grundläggande översikt över teorin som deepfakes baseras på samt undersöka några av de effekter som användningen av

deepfakes har haft på samhället. Följande frågor besvaras i

rapporten:

* Vad är den grundläggande teoretiska och teknologiska bakgrunden för deepfakes?
* Vilka försvarmetoder existerar för att upptäcka deepfakes och hur ser utvecklingen ut?

## Metod och material

Denna rapport är en litteraturstudie. Materialet utgörs av vetenskapliga artiklar som systematiskt och metodisk granskats och värderats. Materialet har avgränsats för att ge en övergripande beskrivning av ämnets vetenskapliga bakgrund, samt att ge konkreta exempel på vilka metoder för upptäckt av deepfakes som existerar. Metoden är relevant då rapportens syfte är att ge en grundläggande bakgrund för deepfakes.

# Resultat

I första delen av resultatet utgörs av en grundläggande beskrivning av deepfakes teoretiska bakgrund. Vi beskriva AI, maskininlärning och neuronnät, samt definitioner av olika sorter av deepfakes. I andra delen beskrivs tillvägagångssätt för att upptäcka deepfakes, samt en kort beskrivning av hur utvecklingen av dessa metoder ser ut.

## Vad är en Deepfake?

2.1.1 AI och maskininlärning

Termen deepfakes är ett teleskopord av ”deep learning” och ”fake”. Deep learning är en del av maskininlärning (ML), som har sin grund i *artificiell intelligens* (AI). Inom forskning om AI behandlas frågor om hur datorer kan imitera mänskliga beteenden. Särskilt intressanta är frågor om programmering av datorer för att lösa komplexa problem, gärna oberoende eller med minimalt mänsklig inflytande Med teknologiutveckling ökar behovet för AI-system att förändras och lära utifrån informationen som bearbetas. Detta gav upphov till *maskininlärning*, där koncept från statistik, filosofi, informationsteori och biologi används för lösa olika problem. Maskininlärning delas in i tre olika kategorier, övervakad inlärning, oövervakad inlärning, och förstärkt inlärning [5]. Av dessa är övervakad inlärning den vanligaste.

En begränsning för maskininlärning är behandling av rå data. Här kommer *representationsinlärning* in. Konceptet omfattas av en mängd metoder som tillåter programmen att automatiskt hitta korrekt representation av den data som bearbetas [6].

De metoder som används i deep learning är *representationsinlärning* som används i flera nivåer, där den råa informationen(data) bearbetas upp till en högre representationsnivå. Dessa transformationer kombineras tillsammans med andra och kan på så sätt skapa mer komplexa funktioner.

2.1.2 Neuronnät

Neuronnät (*neural networks*) ett uttryck som går hand i hand med deep learning. Neuronnnät kan beskrivas som en samling av modeller som är uppbyggda av neuroner kopplade med synapser. [7].

2.1.3 Vad kännetecknar en deepfake?

Deepfakes kan delas in i olika kategorier:

* Återskapande deepfakes
* Ersättande deepfakes
* Redigerade deepfakes

## Försvarsmetoder -

Skriv dina resultat i förhållande till frågeställning 2 osv.

# Diskussion och slutsatser

# Referenser

1. J. Roettgers, “*How AI Tech Is Changing Dubbing, Making Stars Like David Beckham Multilingual”*, Variety.com, [https://variety.com/2019/biz/news/ai-dubbing- david-beckham-multilingual-1203309213/](https://variety.com/2019/biz/news/ai-dubbing-%20%20%20%20david-beckham-multilingual-1203309213/) (Tillgänglig 5 Dec. 2022).
2. J. Dietmar, ” *GANs And Deepfakes Could Revolutionize The Fashion Industry”*, Forbes.com, <https://www.forbes.com/sites/forbestechcouncil/2019/05/21/gans-and-deepfakes-could-revolutionize-the-fashion-industry/?sh=30c2d1fc3d17> (Tillgänglig 5 Dec. 2022).
3. ] S. Cole, “*AI-Assisted Fake Porn Is Here and We’re All Fucked*”, Vice.com <https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn> (Tillgänglig 5 Dec. 2022).

H. v/d Burchard, “*Belgian socialist party circulates ‘deep fake’ Donald Trump video*”, Politico.com, <https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video/> (Tillgänglig 5 Dec. 2022).

1. ] C. Janiesch, P. Zschech & K. Heinrich, “Machine learning and deep learning”, *Electron Markets*, vol. 31, no. 3, pp 685–695, Sep. 2022, tillgänglig 5 Dec. 2022, doi: <https://doi.org/10.1007/s12525-021-00475-2> [Online]. Tillgänglig: [https://link.springer.com/article/10.1007/s12525-021-00475-2#citeas](https://link.springer.com/article/10.1007/s12525-021-00475-2%23citeas)
2. Y. LeCun, Y. Bengio & G. Hinton, “Deep learning”, *Nature*, vol. 521, no. 7552, pp 436-444, Maj 2015, tillgänglig 5 Dec. 2022, doi: <https://doi.org/10.1038/nature14539> [Online]. Tillgänglig: <https://www.nature.com/articles/nature14539>
3. Y. Mirsky & W. Lee, “The Creation and Detection of Deepfakes: A Survey”, *ACM Computing Surveys*, vol. 54, no.1, pp 1-41, Jan. 2022, tillgänglig 5 Dec. 2022, doi: <https://doi.org/10.1145/3425780> [Online]. Tillgänglig: <https://dl.acm.org/toc/csur/2022/54/1>