



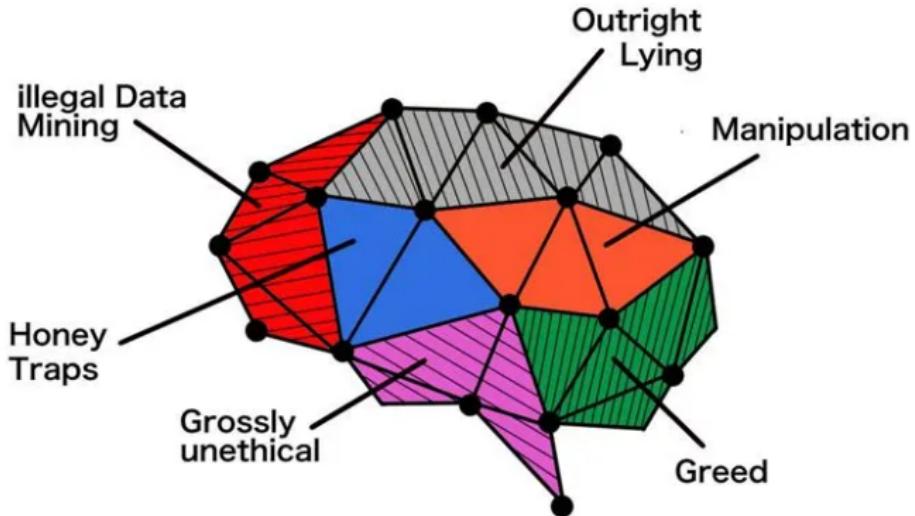
## Human hacking & deepfake: manipolazione tramite social network e IA generativa

# Outline

- ① Motivazione
- ② Definizioni
- ③ Breve storia
- ④ Fondamenti teorici/computazionali
- ⑤ Difesa: come rilevarle ?
- ⑥ Aspetti legali

# Motivazione

# Cambridge Analytica [1]



# Cambridge Analytica

# Cambridge Analytica [2]

## How was Facebook users' data misused?

- 1** In 2014 a Facebook quiz invited users to find out their personality type
  - 2** The app collected the data of those taking the quiz, but also recorded the public data of their friends
  - 3** About 305,000 people installed the app, but it gathered information on up to 87 million people, according to Facebook
  - 4** It is claimed the data was sold to Cambridge Analytica (CA), which used it to psychologically profile voters in the US
  - 5** CA denies it broke any laws and says it did not use the data in the US presidential election
  - 6** Facebook sends notices to users telling them whether their data was breached
- CA denies any wrongdoing. Facebook has apologised to users and says a "breach of trust" has occurred.

BBC

# Cambridge Analytica [3]

≡ MENU     CERCA

**LA STAMPA**

IL QUOTIDIANO  ABBONATI [ACCEDI](#)

## Economia

Lavoro Agricoltura TuttoSoldi Finanza Borsa Italiana Fondi Obbligazioni

# Meta paga 725 milioni di dollari per risolvere class action Cambridge Analytica

TELEBORSA

Pubblicato il 23/12/2022  
Ultima modifica il 23/12/2022 alle ore 13:03

cerca un titolo



Meta Platforms (società che controlla Facebook, Instagram e WhatsApp) ha accettato di pagare 725 milioni di dollari per risolvere una class action che accusava il colosso dei social media di consentire a terzi, tra cui Cambridge Analytica, di accedere alle informazioni personali degli utenti. La società fondata da Mark Zuckerberg non ha ammesso alcun illecito come parte dell'accordo, che è soggetto all'approvazione di un giudice federale di San Francisco, ma ha sottolineato che l'accordo è "nel migliore interesse della nostra comunità e degli azionisti".

**LEGGI ANCHE**

04/04/2024   
New York: risultato positivo per Meta Platforms

19/04/2024

# Cambridge Analytica [4]

ANSA.it › Tecnologia › Internet & Social ›

**Garante Privacy, multa 1 mln a Facebook per Cambridge Analytica**

## Garante Privacy, multa 1 mln a Facebook per Cambridge Analytica

Società attraverso app ha avuto accesso a dati 87 milioni di utenti

**Redazione ANSA**

28 giugno 2019

16:54

ANALISI

 Suggerisci

 Facebook

 Twitter

 Altri



 Stampa

 Scrivi alla redazione



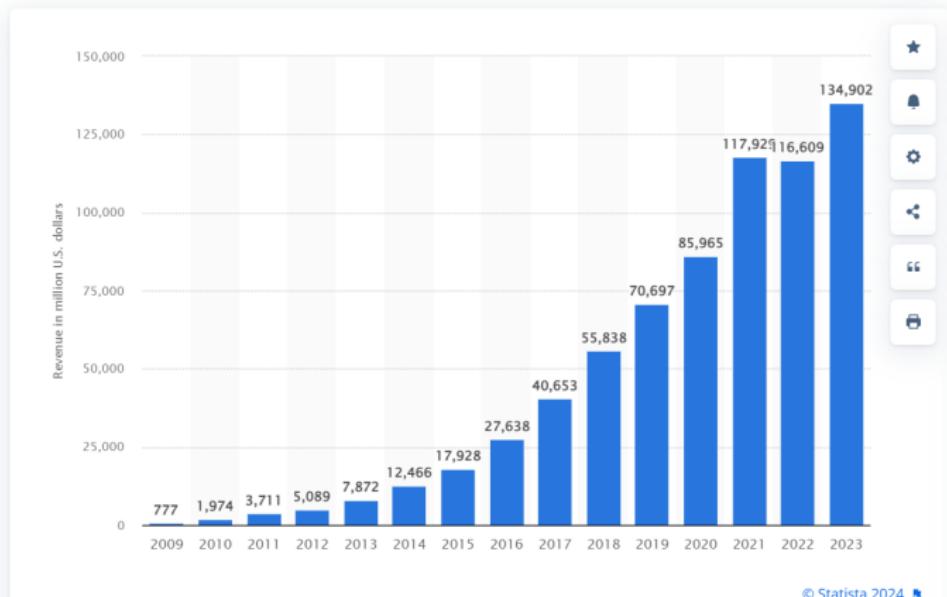
Garante Privacy, multa 1 mln a Facebook per Cambridge Analytica © ANSA/AP

CLICCA PER INGRANDIRE 

# Cambridge Analytica [5]

Internet > Social Media & User-Generated Content

## Annual revenue generated by Meta Platforms from 2009 to 2023 (in million U.S. dollars)



# Mia Ash



**Mia Ash**  
Photographer at Mia's Photography  
London, Greater London, United Kingdom | Photography

500+ connections

Current: Mia's Photography  
Previous: Loft Studios, Clapham Studios  
Education: Goldsmiths, University of London

# Mia Ash [6]



**Timothy Stokes**  
Recruitment Consultant at Teledyne Technologies Incorporated  
Newport Beach, California | Electrical/Electronic Manufacturing  
500+ connections

Current: Teledyne Technologies Incorporated  
Previous: Exxomobil  
Education: University of California, San Diego

Join LinkedIn and access Timothy's full profile. It's free!

As a LinkedIn member, you'll join 300 million other professionals who are sharing connections, ideas, and opportunities.

- See who you know in common
- Get introduced
- Contact Timothy directly

[View Timothy's Full Profile](#)

---

**Summary**

I assist in selecting the best-qualified candidates during open hiring. I am involved with screening applications, interviewing candidates and checking references. Our contracts involve the recruitment and management of temporary employees, project management, and the provision of a resource or international basis to support major engineering, construction, installation and ongoing operations activities. Teledyne Technologies Inc. owns a globally focused operation, active in over 40 locations, driven by a professional and talented team of people, dedicated to achieving excellence.

---

**Experience**

**Recruitment Consultant**  
Teledyne Technologies Incorporated  
March 2012 – Present (3 years 5 months) | Thousand Oaks

- Remarkable experience in Recruitment Consultancy
- Project based recruitment, candidate screening & referral networking for both local and emerging
- Ability to identify and successfully qualify candidates
- Familiarity with contract management and compliance relevant to contractors
- Good understanding of Consultants contracts and terms and conditions
- Amazing ability to manage independently





**Steve Highsmith**  
IT Operations Manager at Doosan  
United Kingdom | Machinery  
5 connections

Join LinkedIn and access Steve's full profile. It's free!

As a LinkedIn member, you'll join 300 million other professionals who are sharing connections, ideas, and opportunities.

- See who you know in common
- Get introduced
- Contact Steve directly

[View Steve's Full Profile](#)

---

**Experience**

**IT Operations Manager**  
Doosan  
May 2009 – Present (5 years 3 months)



# Mia Ash [6]



**Timothy Stokes**  
Recruitment Consultant at Teledyne Technologies Incorporated  
Newport Beach, California | Electrical/Electronic Manufacturing  
Current: Teledyne Technologies Incorporated  
Previous: Exxomobil  
Education: University of California, San Diego

500+ connections

Join LinkedIn and access Timothy's full profile. It's free!

As a LinkedIn member, you'll join 300 million other professionals who are sharing connections, ideas, and opportunities.

- See who you know in common
- Get introduced
- Contact Timothy directly

[View Timothy's Full Profile](#)

---

**Summary**

I assist in selecting the best-qualified candidates during open hiring. I am involved with screening applications, interviewing candidates and checking references. Our contracts involve the recruitment and management of temporary employees, project management, and the provision of a resource or international basis to support major engineering, construction, installation and ongoing operations activities. Teledyne Technologies Inc. owns a globally focused operation, active in over 40 locations, driven by a professional and talented team of people, dedicated to achieving excellence.

---

**Experience**

**Recruitment Consultant**  
Teledyne Technologies Incorporated  
March 2012 – Present (3 years 5 months) | Thousand Oaks

- Remarkable experience in Recruitment Consultancy
- Project based recruitment, candidate screening & referral networking for both local and emerging
- Ability to identify and successfully qualify candidates
- Familiarity with contract management and compliance relevant to contractors
- Good understanding of Consultants contracts and terms and conditions
- Amazing ability to manage independently





**Steve Highsmith**  
IT Operations Manager at Doosan  
United Kingdom | Machinery

5 connections

Join LinkedIn and access Steve's full profile. It's free!

As a LinkedIn member, you'll join 300 million other professionals who are sharing connections, ideas, and opportunities.

- See who you know in common
- Get introduced
- Contact Steve directly

[View Steve's Full Profile](#)

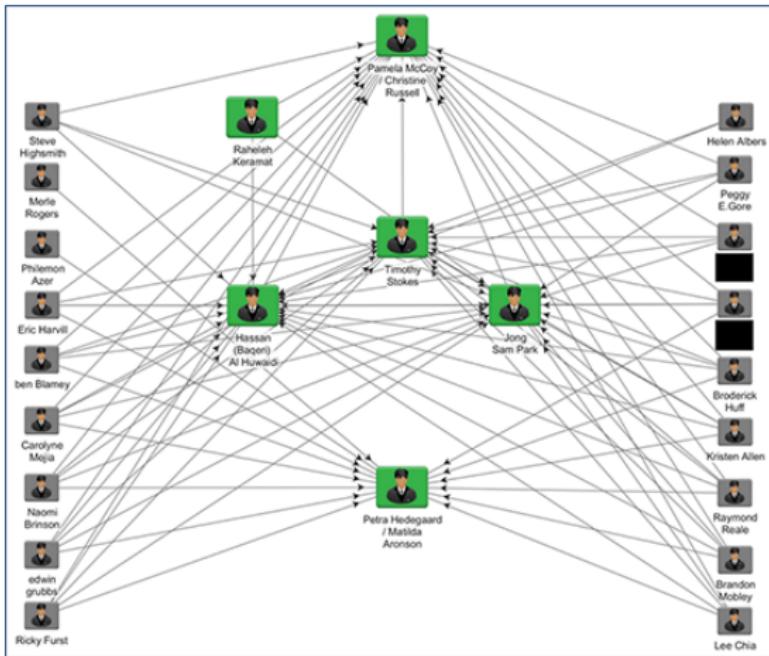
---

**Experience**

**IT Operations Manager**  
Doosan  
May 2009 – Present (5 years 3 months)



# Mia Ash [6]



Exclusive news

 **REUTERS®**

World ▾ Business ▾ Markets ▾ Sustainability ▾ Legal ▾ More ▾

---

Technology

## Iranian hackers used female 'honey pot' to lure targets: researchers

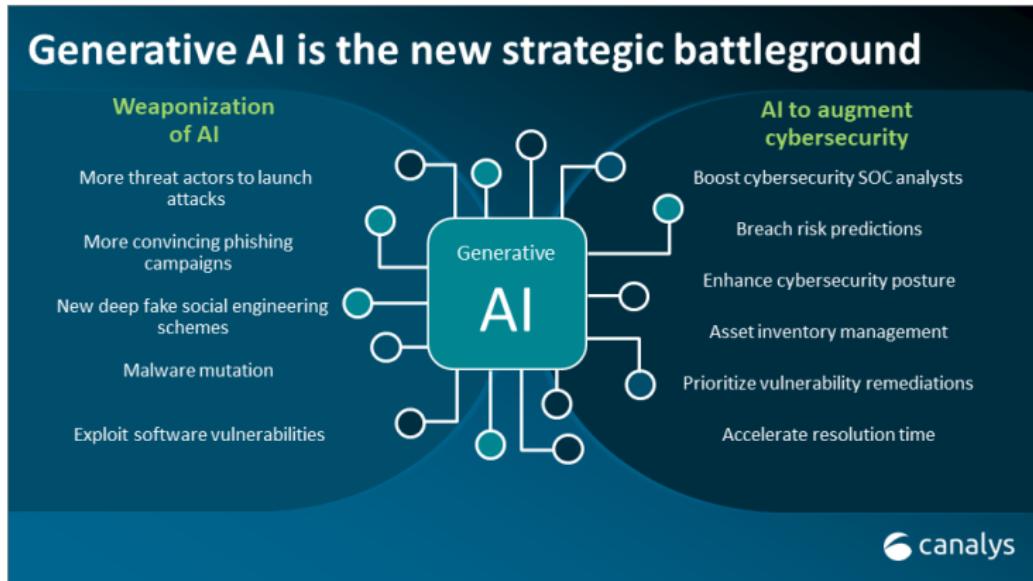
By Dustin Volz

July 27, 2017 4:32 PM GMT+2 · Updated 7 years ago

Aa 



# IA generativa per il phishing [8]



# IA generativa per il phishing [9]

Print subscriptions | Sign in | Search jobs | Search | Europe edition ▾

## Support the Guardian

Fund independent journalism with €10 per month

Support us →

# The Guardian

News    Opinion    Sport    Culture    Lifestyle    More ▾

World UK Climate crisis Ukraine Environment Science Global development Football Tech Business Obituaries

Artificial intelligence (AI)

This article is more than 3 months old

### AI will make scam emails look genuine, UK cybersecurity agency warns

NCSC says generative AI tools will soon allow amateur cybercriminals to launch sophisticated phishing attacks

Dan Milmo  
Global technology

Advertisement



# Click Farm [10]



# Click Farm

The screenshot shows a website for buying Instagram followers. At the top, there's a navigation bar with links for "Faq", "Contact", and "My Account". Below the navigation, a large banner features the text "Special Offer: Buy Instagram Followers at Reduced Prices using AI!" and displays a sample Instagram profile for the user "@why\_not\_you" with 811 posts, 296.7k followers, and 984k likes. A callout box below the profile states "100% Real Active Followers 100%". To the right of the profile, there's a search bar labeled "Your Instagram username" with the placeholder "@". Below the search bar is a section titled "PREMIUM FOLLOWERS" featuring several price options:

Followers	Price	Original Price	Discount
+500	1€	2€	-50%
+1 000	1.8€	3.6€	-50%
+2 500	3.6€	9€	-60%
+5 000	6.8€	17€	-60%
+10 000	12€	35€	-60%
+20 000	20€	50€	-60%
+35 000	32€	85€	-60%
+50 000	45€	110€	-70%
Custom			-50%

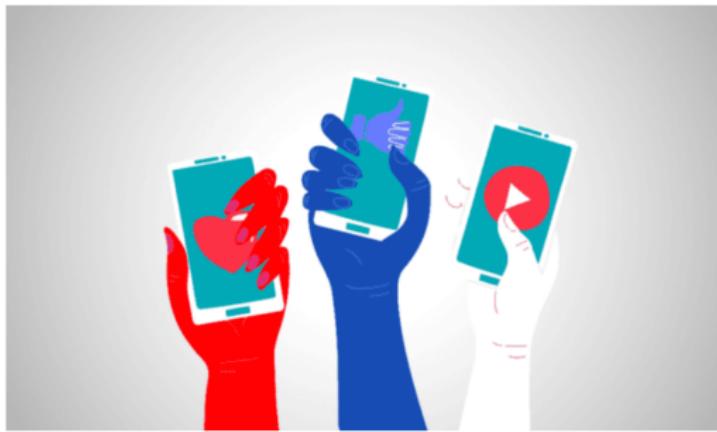
A prominent yellow button at the bottom right says "Order Now!".

# Click Farm

The screenshot shows a web browser window for [piulike.com/categoria-instagram/](https://piulike.com/categoria-instagram/). The page header includes links for Instagram, YouTube, Facebook, Spotify, TikTok, Servizi Italiani, Altri Servizi, and Contattaci. The main content features a large black banner with white text: "Follower, Like e Views su Instagram" and "Diventa popolare su Instagram: aumenta la visibilità!". Below the banner, a section titled "DIVENTA POPOLARE SU INSTAGRAM" encourages users to increase their Instagram followers, receive likes, comments, and views. A search bar asks "Cosa stai cercando?". At the bottom, there are four buttons: "FOLLOWERS", "FOLLOWERS TARGET", "LIKES", and "VIEWS". A live chat bubble from a user named "Ciao" asks for a discount to try the services, and a response says "Si, grazie! No, grazie.".

# Click Farm [11]

## Social Media Influencers and the 2020 U.S. Election: Paying ‘Regular People’ for Digital Campaign Communication



# Manipolazione dei sistemi democratici

## Politica e nano-influencer [11]

"Coordinated networks of social media influencers, especially small-scale influencers with fewer than 10,000 followers, are now a powerful asset for political campaigns, PACs, and special interest groups. Partisan organizations are leveraging these “authentic” accounts in bids to sway political discourse and decision-making in the run up to the 2020 U.S. elections. Political marketers tell us that they see influencers, particularly those with more intimate followings, as regarded as more trustworthy by their followers and therefore better positioned to change their behavior"

# Definizioni

# Deepfake e Intelligenza Artificiale - definizioni

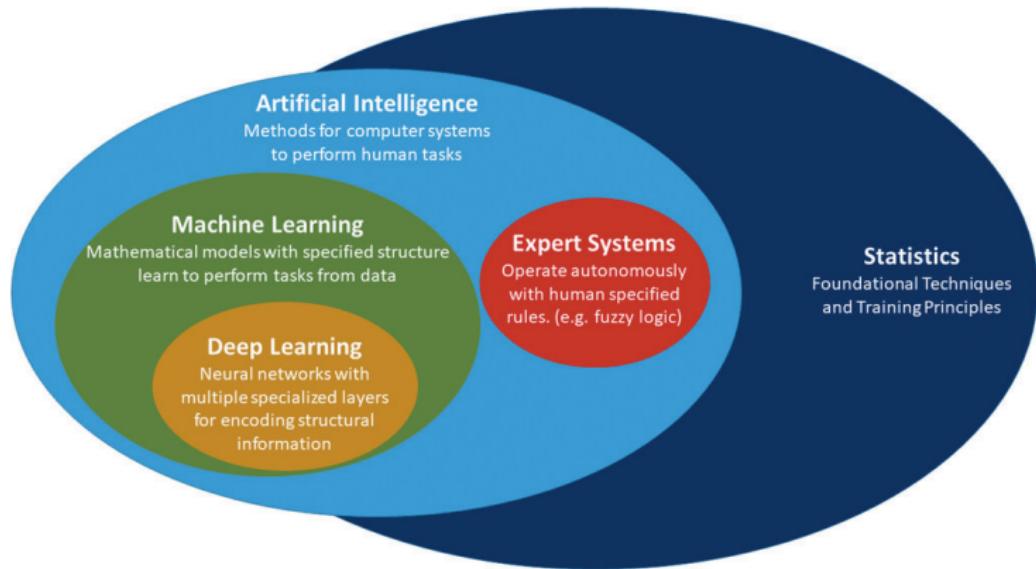
## Deepfake (NSA) [12]

Contenuto multimediale creato sinteticamente o manipolato utilizzando una qualche forma di tecnologia meccanica o di deep learning.

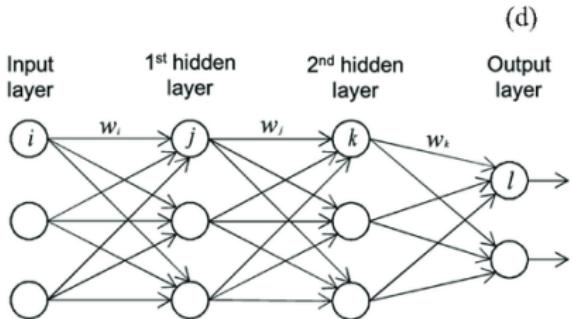
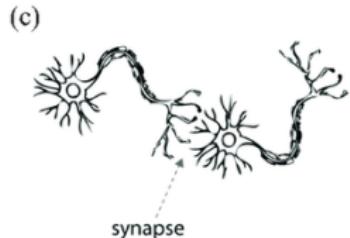
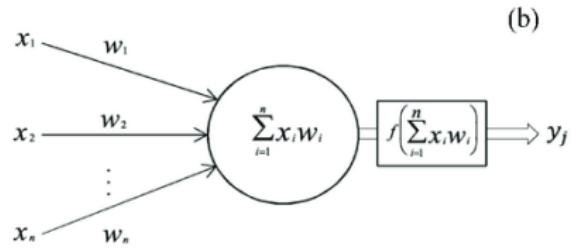
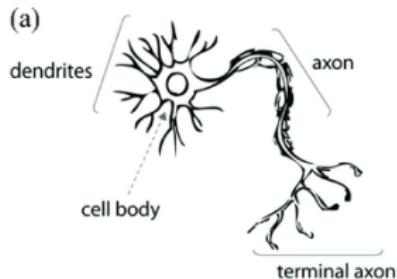
## Intelligenza artificiale (NIST) [13]

Una branca dell'informatica dedicata allo sviluppo di sistemi di elaborazione dati che svolgono funzioni normalmente associate all'intelligenza umana, come il ragionamento, l'apprendimento e l'auto-miglioramento

# Deepfake e Intelligenza Artificiale - tassonomia [14]



# Deepfake e Intelligenza Artificiale - reti neurali [15]



# Breve storia

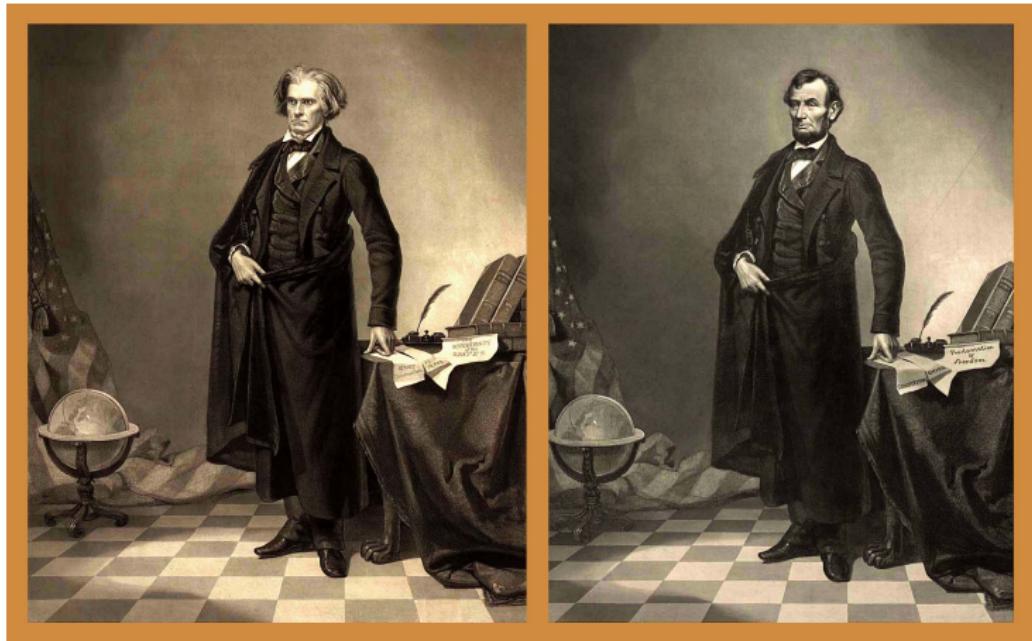
# Storia dei deepfake

“Never trust  
anything you  
see on the  
internet”

Abraham Lincoln



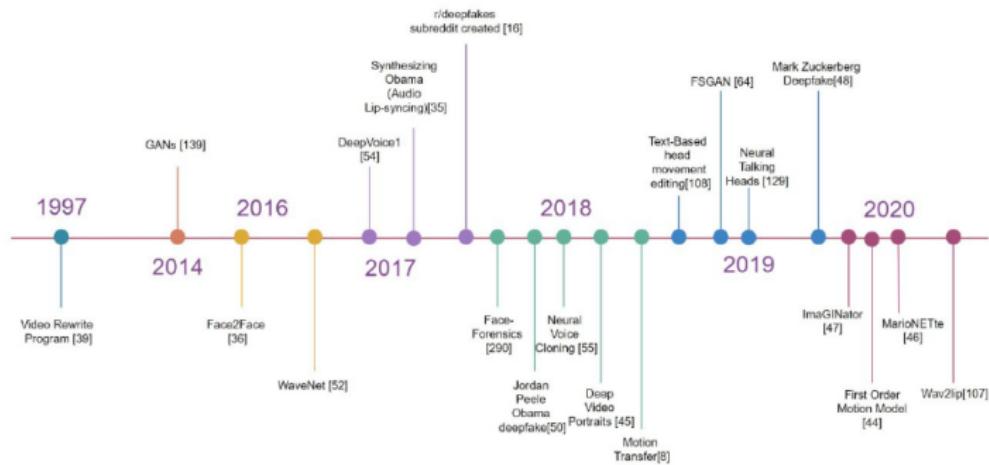
# Storia dei deepfake: 1860 Calhoun/Lincon [16]



# Storia dei deepfake: 1997 Video Rewrite (GAN) [17]



# Deepfake e Intelligenza Artificiale - tassonomia [27]



# Storia dei deepfake: 2017 B.Obama [18]

Print subscriptions Sign in Search jobs Search Europe edition

## Support the Guardian

Fund independent journalism with €10 per month

Support us →

News Opinion Sport Culture Lifestyle More ▾

Technology

This article is more than 6 years old

### The future of fake news: don't believe everything you read, see or hear

A new breed of video and audio manipulation tools allow for the creation of realistic looking news footage, like the now infamous fake Obama speech

Olivia Solon in San Francisco  
Wed 26 Jul 2017 07.00 CEST

Share

Synthesizing Obama: Learning Lip Sync from Audio

Copia link

Without Re-timing With Re-timing (Our Result)

Guarda su YouTube

The University of Washington's [Synthesizing Obama](#) project took audio from one of Obama's speeches and used it to animate his face in an entirely different video



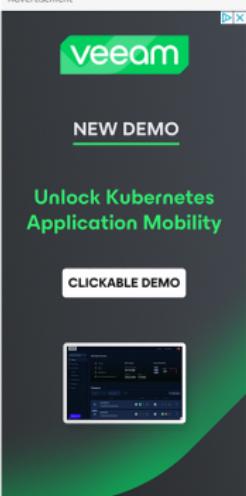
Advertisement

veeam

NEW DEMO

Unlock Kubernetes Application Mobility

CLICKABLE DEMO



# Storia dei deepfake: 2017 Imitazione della voce [19]

BANKRUPTCY CENTRAL BANKING CYBERSECURITY PRIVATE EQUITY SUSTAINABLE BUSINESS VENTURE CAPITAL

WSJ PRO CYBERSECURITY

Home News Research Archive Newsletters Events

WSJ PRO

## Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case

Scams using artificial intelligence are a new challenge for companies

By Catherine Stipp Updated Aug. 30, 2019 12:52 pm ET | WSJ PRO

Share AA Resize



PHOTO: SIMON DAWSON/BLOOMBERG NEWS

Criminals used artificial intelligence-based software to impersonate a chief executive's voice and demand a fraudulent transfer of €220,000 (\$243,000) in March in what cybercrime experts described as an unusual case of artificial intelligence being used in hacking.

The CEO of a U.K.-based energy firm thought he was speaking on the phone with his boss, the chief executive of the firm's German parent company, who asked him



Draußen bleiben,  
trocken bleiben  
Wasserdichter Schutz,  
hergestellt ohne PFCs.

Regenbekleidung entdecken

MUST READS FROM CYBERSECURITY

- Port of Rotterdam Tests Quantum Network to Defend Against Hacks
- Otta Hack Update

# Storia dei deepfake: 2020 N.Pelosi [19]



World ▾ Business ▾ Markets ▾ Sustainability ▾ Legal ▾ Breakingviews ▾ More ▾

World

## Fact check: “Drunk” Nancy Pelosi video is manipulated

By **Reuters**

August 3, 2020 7:23 PM GMT+2 · Updated 4 years ago



U.S. Speaker of the House Nancy Pelosi, joined by Senate Minority Leader Chuck Schumer, speaks to reporters in the U.S. Capitol in Washington, U.S. July 29, 2020. REUTERS/Erin Scott [Purchase Licensing Rights](#)

A video circulating on social media shows House Speaker Nancy Pelosi speaking in a slurred and awkward manner. One popular post boasts 91,000 shares on Facebook and bears a caption reading: “This is unbelievable, she is blown out of her mind, I bet this gets taken down!” The video, however, has been manipulated to make Pelosi appear drunk and incoherent.

A viral example of the video is visible [here](#).

# Storia dei deepfake: Software disponibili [27]

Tool	Type	Reference/Developer	Technique
<b>Cheat fakes</b>			
Adobe Premiere	Commercial Desktop Software	Adobe	Audio Video Editing, AI-powered video refacing
Corel VideoStudio	Commercial Desktop Software	Corel	Proprietary AI
<b>Lip-sync</b>			
dynalips	Commercial Web App	<a href="http://www.dynalips.com">www.dynalips.com</a>	Proprietary
cryctalk	Commercial Web App	<a href="http://www.realmon.com/cryctalk/">www.realmon.com/cryctalk/</a>	Proprietary
Wav2Lip	Open source implementation	<a href="https://github.com/Rudisha/Wav2Lip">github.com/Rudisha/Wav2Lip</a>	GAN with pre-trained discriminator network and visual quality loss function
<b>Facial Attribute Manipulation</b>			
FaceApp	Mobile App	FaceApp Inc	Deep generative CNNs
Adobe	Commercial Desktop Software	Adobe	DNNs + filters
Rosebud	Commercial Web App	<a href="http://www.rosebud.ai/">www.rosebud.ai/</a>	Proprietary AI
<b>Face Swap</b>			
ZAO	Mobile app	Momo Inc	Proprietary
REFACE	Mobile app	Neocortext, Inc	Proprietary
Reflect	Mobile app	Neocortext, Inc	Proprietary
Impressions	Mobile app	Synthesized Media, Inc.	Proprietary
FakeApp	Desktop App	<a href="http://www.malavida.com/en/soft/fakeapp/">www.malavida.com/en/soft/fakeapp/</a>	GAN
FaceSwap	Open source implementation	<a href="http://faceswapweb.com/">faceswapweb.com/</a>	Employed two pairs of encoder-decoder. Shared encoder parameters.
DFaker	Open source implementation	<a href="https://github.com/dfaker/df">github.com/dfaker/df</a>	For face reconstruction DSSIM loss function [44] is utilized. Keras library-based implementation.
DeepFaceLab	Open source implementation	<a href="https://github.com/iperon/DeepFaceLab">github.com/iperon/DeepFaceLab</a>	- provide several face extraction methods, e.g. dlib, MTCNN, SIFT etc. - Extend different Face swap model i.e. H64, H128, LIAEF128, SAE [33]
FaceSwapGAN	Open source implementation	<a href="https://github.com/taosunlu/faceswap-GAN">github.com/taosunlu/faceswap-GAN</a>	Uses two loss functions namely adversarial loss and perceptual loss to the auto-encoder.
DeepFake-Tf	Open source implementation	<a href="https://github.com/StromWine/DeepFake-Tf">github.com/StromWine/DeepFake-Tf</a>	Same as DFaker however, used tensor-flow for implementation.
Facesswapweb	Commercial Web App	<a href="http://facesswapweb.com/">facesswapweb.com/</a>	GAN
<b>Face Recreament</b>			
Face2Face	Open source implementation	<a href="http://web.stanford.edu/~zhifeng/papers/CVPR2016_Face2Face/page.html">web.stanford.edu/~zhifeng/papers/CVPR2016_Face2Face/page.html</a>	Uses 3DMM and ML technique
Dynamixyz	Commercial Desktop Software	<a href="http://www.dynamixyz.com/">www.dynamixyz.com/</a>	Machine-learning
FaceiT3	Open source implementation	<a href="https://github.com/idev3/faceti3_live3">github.com/idev3/faceti3_live3</a>	GAN
<b>Face Generation</b>			
Generated Photos	Commercial Web App	generated photos/	StyleGAN
<b>Video Synthesis</b>			
Overdub	Commercial Web App	<a href="http://www.descript.com/overdub">www.descript.com/overdub</a>	Proprietary (AI based)
Reespecter	Commercial Web App	<a href="http://www.reespecter.com/">www.reespecter.com/</a>	Combined traditional digital signal processing algorithms with proprietary deep generative modeling techniques
SV2TTS	Open source implementation	<a href="https://github.com/Continuum/Real-Time-Voice-Cloning">github.com/Continuum/Real-Time-Voice-Cloning</a>	LSTM with Generalized end-to-end loss
ResembleAI	Commercial Web App	<a href="http://www.resemble.ai/">www.resemble.ai/</a>	Proprietary (AI based)
Voxery	Commercial Web App	<a href="http://www.voxery.com/">www.voxery.com/</a>	Proprietary AI and deep learning
VoiceApp	Mobile app	Zoeei AB	Proprietary (AI-based)

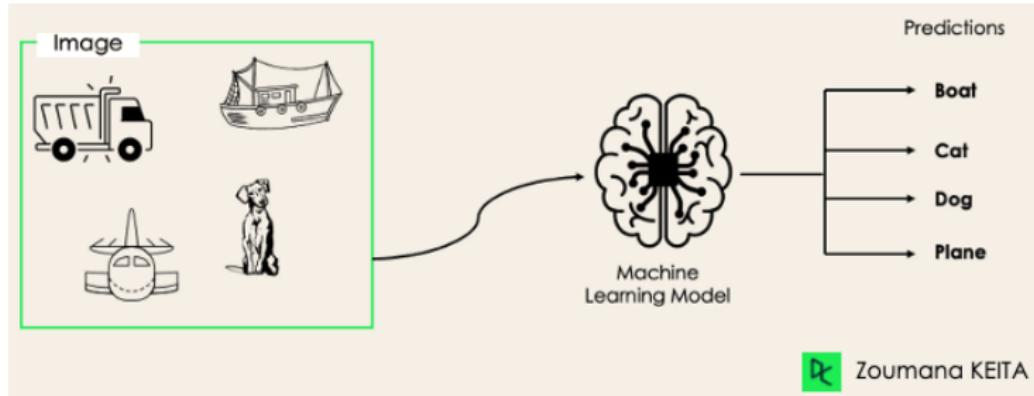
# Fondamenti teorici/computazionali

# Machine learning [29, 28, 20]

## Distribuzione di probabilità condizionata - idea

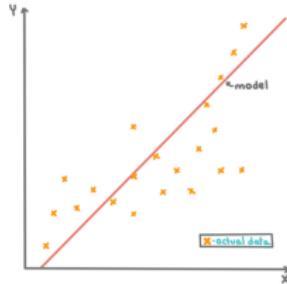
Abbiamo una serie di record che contengono delle variabili input (dimensione del fiore, colore etc..) e una variabile di output (caso semplice, può essere generalizzato). Non sappiamo a priori come input e output siano legati (altrimenti avremmo risolto il problema !), ma siamo interessati a trovare una funzione di probabilità che associa degli input a degli output.  $p(y = c|\mathbf{x}; \theta) = f_c(\mathbf{x}; \theta)$

# Machine learning [29, 28, 20]



# Machine learning [29, 28, 21, 22]

## LINEAR REGRESSION



### HOW IT WORKS

establishes a relationship between X and Y.

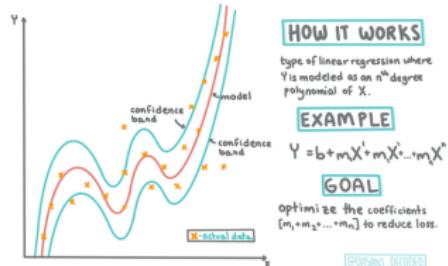
### EXAMPLE

$$Y = b + mX$$

### GOAL

optimize the slope ( $m$ ) to reduce loss

## POLYNOMIAL REGRESSION



### HOW IT WORKS

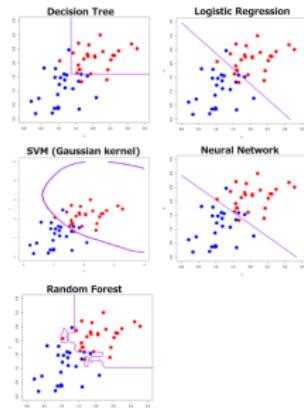
type of linear regression where  $Y$  is modeled as an  $n^{\text{th}}$ -degree polynomial of  $X$ .

### EXAMPLE

$$Y = b + m_1X + m_2X^2 + \dots + m_nX^n$$

### GOAL

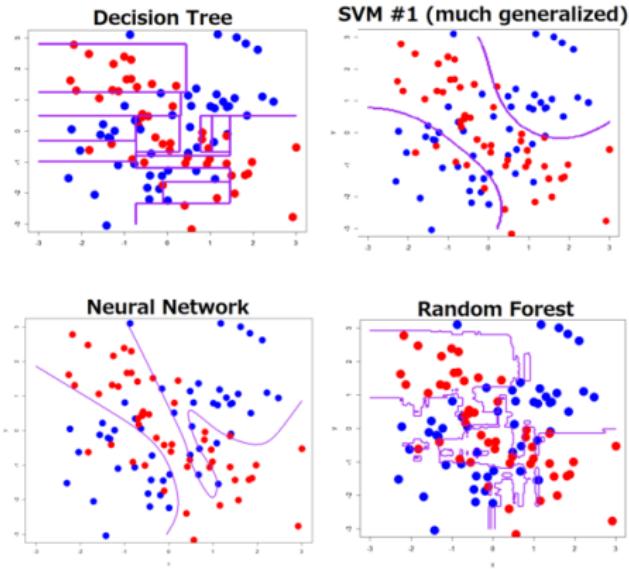
optimize the coefficients  $[m_1, m_2, \dots, m_n]$  to reduce loss.



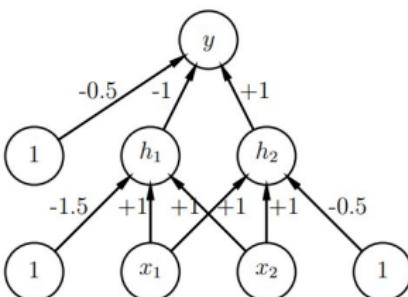
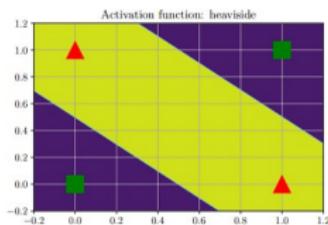
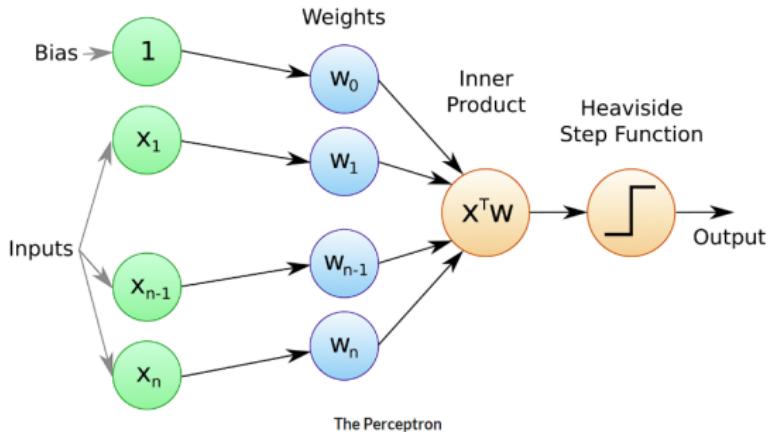
## No free lunch theorem

Non esiste un modello che abbia performance migliori in tutti i casi possibili. Ciò è dovuto al fatto che ogni modello fa delle assunzioni.

# Machine learning [29, 28, 21, 22]



# Reti neurali [29, 28]

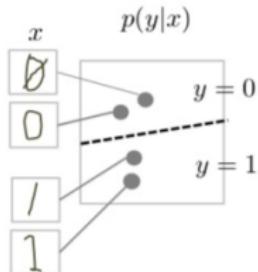


# Modelli generativi [23, 29, 28]

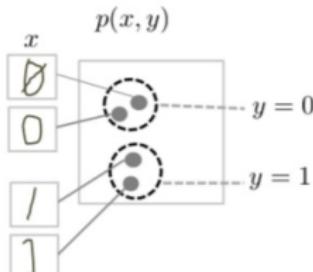
## Modello generativo

Un modello generativo descrive come viene generato un set di dati utilizzando un modello probabilistico ( $p(x) \quad x \in X$ ). Campionando da questo modello, siamo in grado di generare nuovi dati.

- Discriminative Model



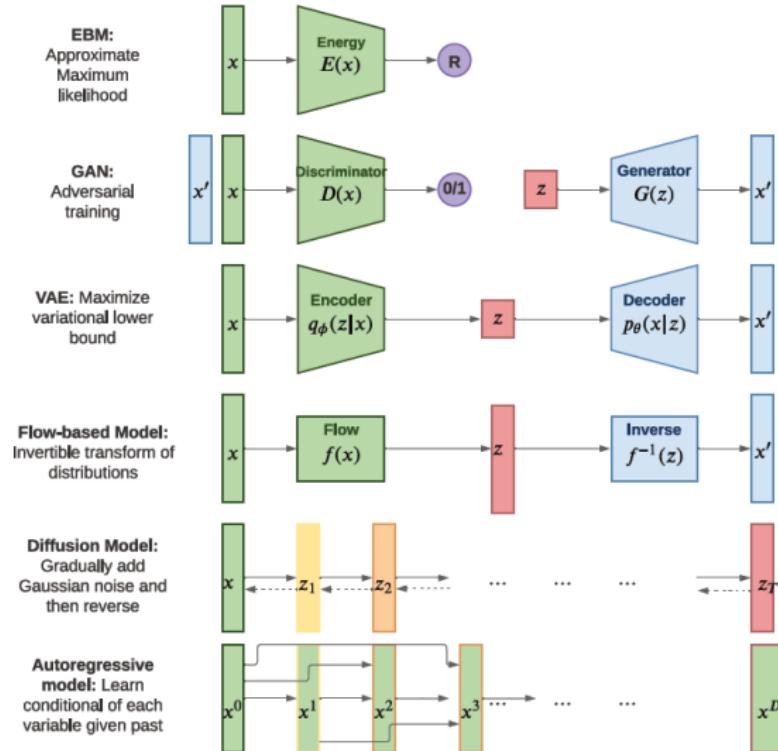
- Generative Model



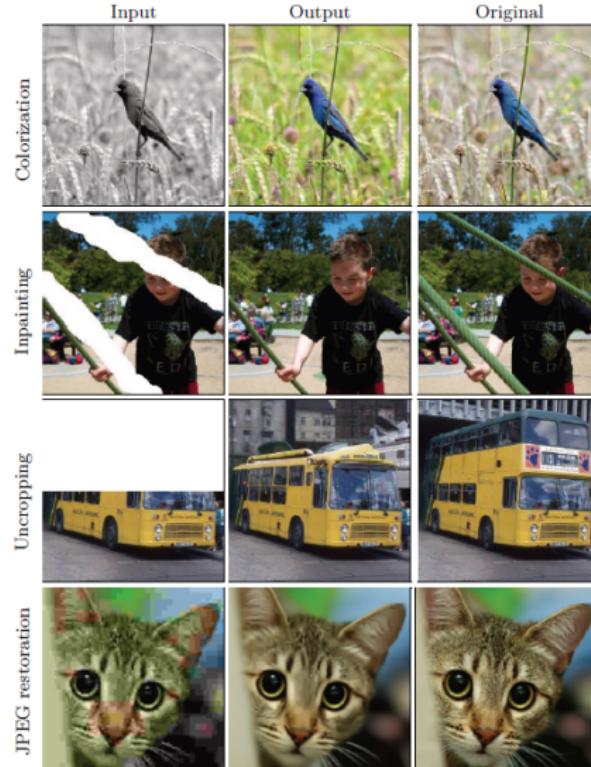
## Tipologia di modelli generativi

- Deep generative models (DGM): viene usata una deep neural network. Comprende i Variational Autoencoder (VAE) e le Generative adversarial network oltre che, ad esempio, i Diffusion models e gli Energy based models (EBM)
- Probabilistic graphical models: viene usato un grafo causale

# Modelli generativi [29, 28]



# Modelli generativi: esempi [29, 28]



# Autoencoders [29, 28]

## Autoencoder

Encoder ( $f_e$ ) + Decoder  $f_d$

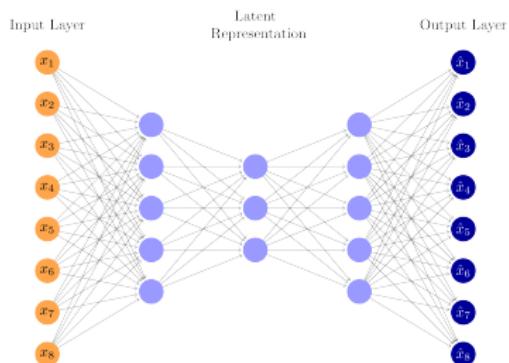
$$f_e : x \rightarrow z \quad f_d : z \rightarrow x$$

$r(x) = f_d(f_e(x))$  (rec. function)

$\mathcal{L}(\theta) = \|r(x) - x\|^2$  (loss function)

## Funzionamento

L'unità in mezzo agisce come collo di bottiglia tra l'input e la sua ricostruzione, in modo da applicare una compressione.



# Autoencoders [29]



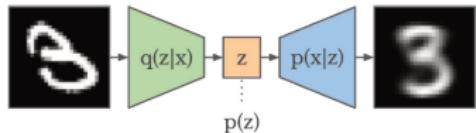
# Variational autoencoders [29, 28]

## Modello generativo

Un modello generativo descrive come viene generato un set di dati utilizzando un modello probabilistico ( $p(x) \quad x \in X$ ). Campionando da questo modello, siamo in grado di generare nuovi dati.

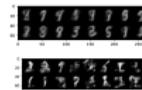
## Differenza rispetto agli AE

Rispetto agli autoencoder i variational autoencoder possono essere visti come una versione probabilistica di un autoencoder deterministico. In questo modo si può avere una IA generativa

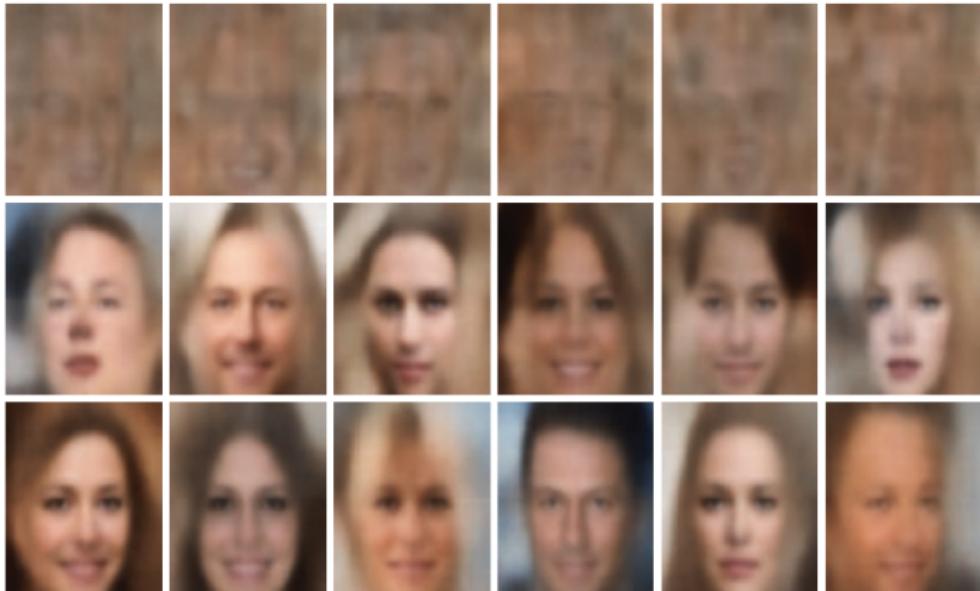


## Remark

La VAE, per come è costruita, è in grado di convertire punti casuali in output, mentre il decoder di AE (deterministico) funziona solo per un punto che è già presente nel training.



# VAR generation [29, 28]

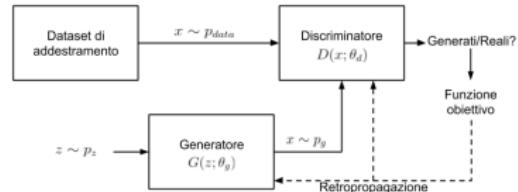


# Generative adversarial networks [29, 28]

## Idea

Ho due reti neurali

- Generatore: creare nuovi dati
- Discriminatore: distinguere i dati nuovi da quelli veri



Algorithm 26.2: GAN training algorithm

```
1 Initialize  $\phi, \theta$ 
2 for each training iteration do
3   for  $K$  steps do
4     Sample minibatch of  $M$  noise vectors  $z_m \sim q(z)$ 
5     Sample minibatch of  $M$  examples  $x_m \sim p(x)$ 
6     Update the discriminator by performing stochastic gradient descent using this gradient:

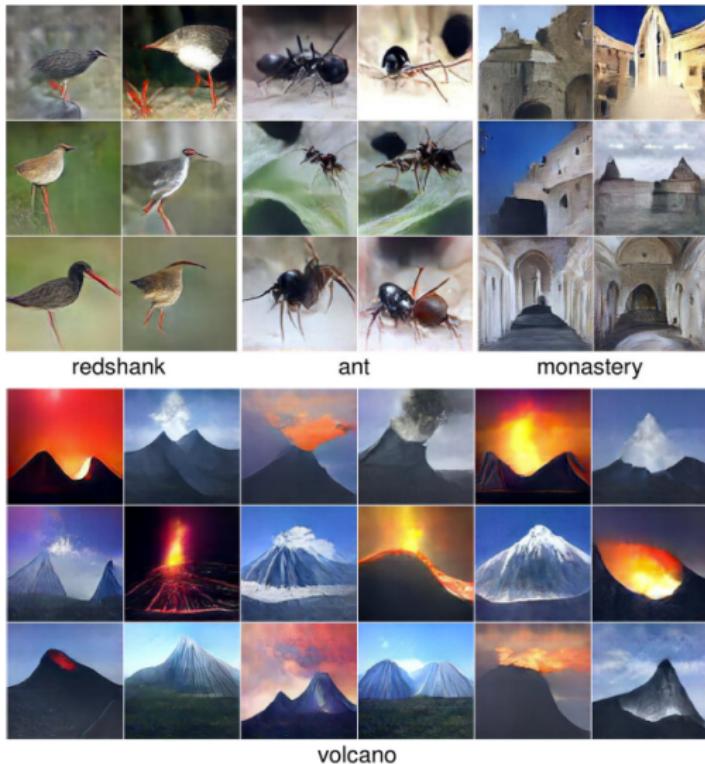
$$\nabla_{\phi} \frac{1}{M} \sum_{m=1}^M [q(D_{\phi}(x_m)) + \nabla_A h(D_{\phi}(G_{\theta}(z_m)))]$$

7     Sample minibatch of  $M$  noise vectors  $z_m \sim q(z)$ 
8     Update the generator by performing stochastic gradient descent using this gradient:

$$\nabla_{\theta} \frac{1}{M} \sum_{m=1}^M l(D_{\phi}(G_{\theta}(z_m)))
9 \text{Return } \phi, \theta$$

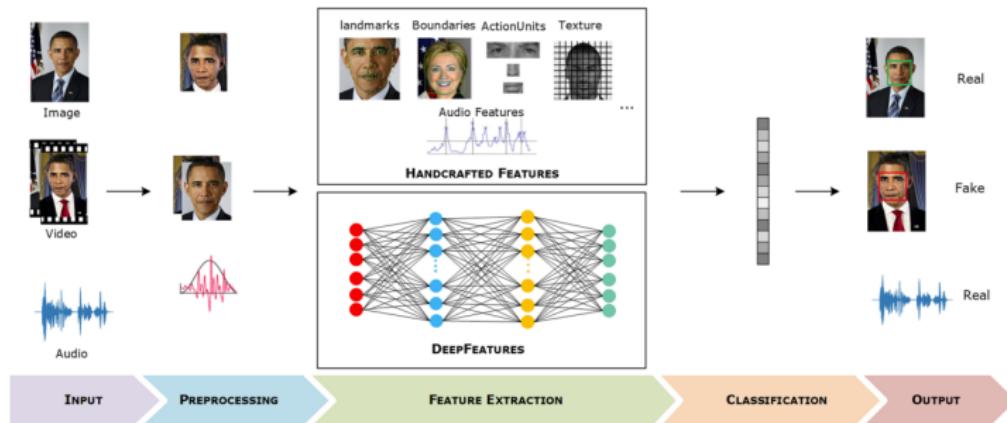
```

# VAR generation [30]

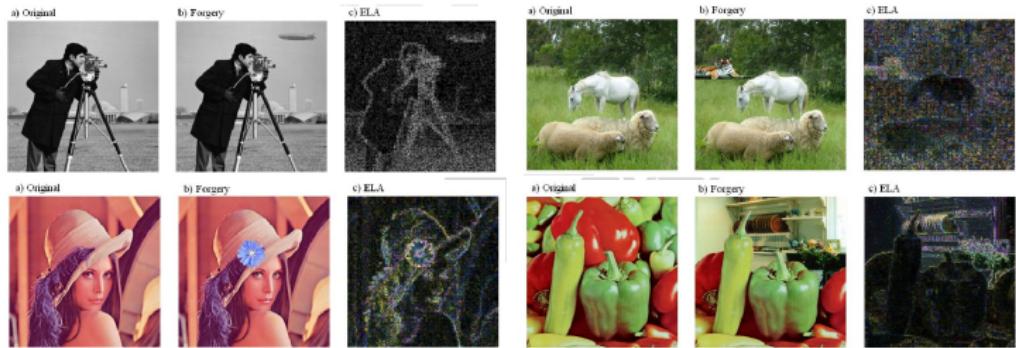


# Difesa: come rilevarle ?

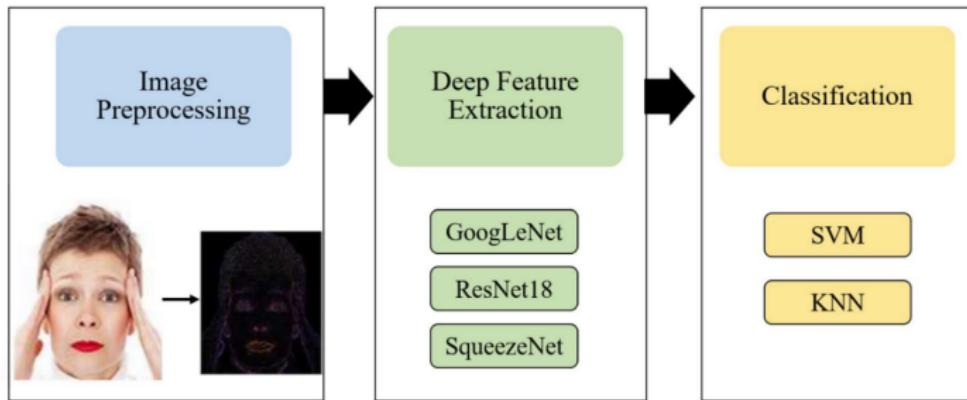
# Metodi [27]



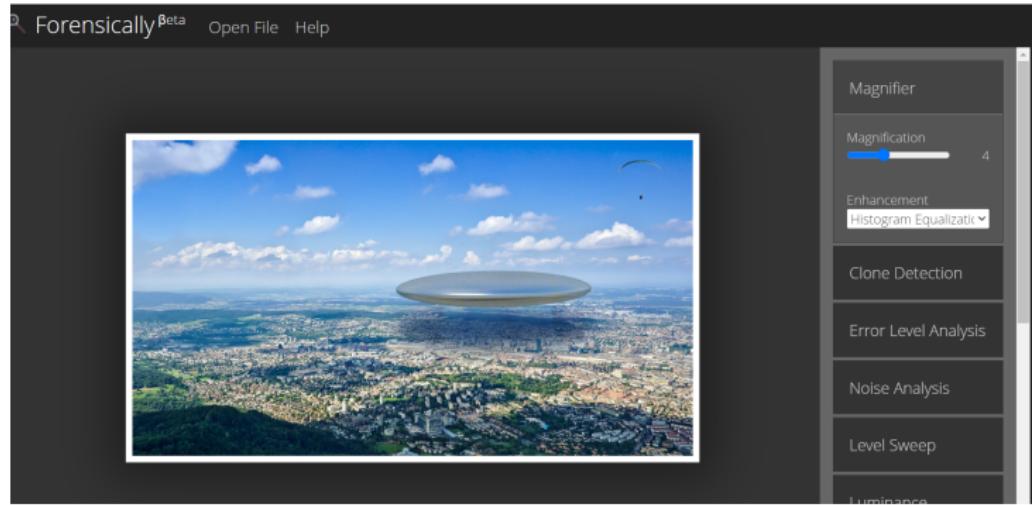
# Error level analysis [26]



# Error level analysis + Deep Learning [31]



# Forensically [24]



# Aspetti legali

70a) A variety of AI systems can generate large quantities of synthetic content that becomes increasingly hard for humans to distinguish from human-generated and authentic content. The wide availability and increasing capabilities of those systems have a significant impact on the integrity and trust in the information ecosystem, raising new risks of misinformation and manipulation at scale, fraud, impersonation and consumer deception. **In the light of those impacts, the fast technological pace and the need for new methods and techniques to trace origin of information, it is appropriate to require providers of those systems to embed technical solutions that enable marking in a machine readable format and detection that the output has been generated or manipulated by an AI system and not a human.** Such techniques and methods should be sufficiently reliable, interoperable, effective and robust as far as this is technically feasible, taking into account available techniques or a combination of such techniques, such as watermarks, metadata identifications, cryptographic methods for proving provenance and authenticity of content, logging methods, fingerprints or other techniques, as may be appropriate

# Conseguenze di un uso improprio dei deepfake [25]

## Uso improprio dei deepfake

- Diffamazione, attraverso la creazione di contenuti che denigrano o danneggiano la reputazione di un individuo
- Furto di identità, che rileva nel caso della creazione di video o audio in cui un soggetto viene rappresentato come un'altra persona
- Violazione della privacy, attraverso la condivisione non consensuale di informazioni relative a una persona

# Bibliography I

- [1] [https://www.ansa.it/sito/notizie/tecnologia/internet\\_social/2019/06/28/garante-privacy-multa-1-mln-a-facebook-per-cambridge-analytica\\_ec9234f5-caae-4485-9beb-e092137a4d67.html](https://www.ansa.it/sito/notizie/tecnologia/internet_social/2019/06/28/garante-privacy-multa-1-mln-a-facebook-per-cambridge-analytica_ec9234f5-caae-4485-9beb-e092137a4d67.html).
- [2] <https://www.bbc.com/news/business-43983958>.
- [3] <https://finanza.lastampa.it/News/2022/12/23/meta-paga-725-milioni-di-dollari-per-risolvere-class-action-cambridge-analytica/NzVfMjAyMi0xMi0yM19UTEI>.
- [4] [https://www.ansa.it/sito/notizie/tecnologia/internet\\_social/2019/06/28/garante-privacy-multa-1-mln-a-facebook-per-cambridge-analytica\\_ec9234f5-caae-4485-9beb-e092137a4d67.html](https://www.ansa.it/sito/notizie/tecnologia/internet_social/2019/06/28/garante-privacy-multa-1-mln-a-facebook-per-cambridge-analytica_ec9234f5-caae-4485-9beb-e092137a4d67.html).
- [5] <https://www.statista.com/statistics/268604/annual-revenue-of-facebook/>.

## Bibliography II

- [6] <https://www.secureworks.com/research/suspected-iran-based-hacker-group-creates-network-of-fake-linkedin-profiles>.
- [7] <https://www.reuters.com/article/us-cyber-conference-iran/iranian-hackers-used-female-honey-pot-to-lure-targets-researchers-idUSKBN1AC28L/>.
- [8] <https://www.theguardian.com/technology/2024/jan/24/ai-scam-emails-uk-cybersecurity-agency-phishing>.
- [9] <https://www.theguardian.com/technology/2024/jan/24/ai-scam-emails-uk-cybersecurity-agency-phishing>.
- [10] .
- [11] <https://mediaengagement.org/research/social-media-influencers-and-the-2020-election/>.

## Bibliography III

- [12] [https://www.nsa.gov/Press-Room/Press-Releases-Statements/Press-Release-View/Article/3523329/nsa-us-federal-agencies-advise-on-deepfake-threats/.](https://www.nsa.gov/Press-Room/Press-Releases-Statements/Press-Release-View/Article/3523329/nsa-us-federal-agencies-advise-on-deepfake-threats/)
- [13] [https://csrc.nist.gov/topics/technologies/artificial-intelligence.](https://csrc.nist.gov/topics/technologies/artificial-intelligence)
- [14] [https://journals.ametsoc.org/view/journals/bams/103/5/BAMS-D-20-0234.1.xml.](https://journals.ametsoc.org/view/journals/bams/103/5/BAMS-D-20-0234.1.xml)
- [15] [https://www.researchgate.net/figure/A-biological-neuron-in-comparison-to-an-artificial-neural-network-a-human-neuron-b\\_fig2\\_339446790.](https://www.researchgate.net/figure/A-biological-neuron-in-comparison-to-an-artificial-neural-network-a-human-neuron-b_fig2_339446790)
- [16] [https://www.nationalgeographic.com/photography/article/digitally-manipulated-ai-altered-photo-images.](https://www.nationalgeographic.com/photography/article/digitally-manipulated-ai-altered-photo-images)

## Bibliography IV

- [17] [https://www.historyofinformation.com/detail.php?id=4792.](https://www.historyofinformation.com/detail.php?id=4792)
- [18] [https://www.reuters.com/article/idUSKCN24Z2B1/.](https://www.reuters.com/article/idUSKCN24Z2B1/)
- [19] [https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402.](https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402)
- [20] [https://www.datacamp.com/blog/classification-machine-learning.](https://www.datacamp.com/blog/classification-machine-learning)
- [21] [https://tjo-en.hatenablog.com/entry/2014/01/06/234155.](https://tjo-en.hatenablog.com/entry/2014/01/06/234155)
- [22] [https://github.com/microsoft/ML-For-Beginners/blob/main/2-Regression/3-Linear/solution/R/lesson\\_3-R.ipynb.](https://github.com/microsoft/ML-For-Beginners/blob/main/2-Regression/3-Linear/solution/R/lesson_3-R.ipynb)

## Bibliography V

- [23] <https://developers.google.com/machine-learning/gan/generative?hl=it>.
- [24] <https://29a.ch/photo-forensics/#help>.
- [25] <https://www.smartius.it/data-it-law/reato-usare-deepfake/>.
- [26] Daniel Cavalcanti Jeronymo, Yuri Cassio Campbell Borges e Leandro dos Santos Coelho. "Image forgery detection by semi-automatic wavelet soft-thresholding with error level analysis". In: *Expert Systems with Applications* 85 (2017), pp. 348–356.
- [27] Momina Masood et al. "Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward". In: *Applied intelligence* 53.4 (2023), pp. 3974–4026.

## Bibliography VI

- [28] Kevin P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023. URL: <http://probml.github.io/book2>.
- [29] Kevin P. Murphy. *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. URL: [probml.ai](http://probml.ai).
- [30] Anh Nguyen et al. “Plug & play generative networks: Conditional iterative generation of images in latent space”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4467–4477.
- [31] Rimsha Rafique et al. “Deep fake detection and classification using error-level analysis and deep learning”. In: *Scientific Reports* 13.1 (2023), p. 7422.