

Fake news, deepfake e allucinazioni dell'AI

Capire, verificare, non farsi ingannare

Marinella Petrocchi

IIT-CNR, Pisa • AI Consapevole – AIxIA 2026

Perugia, 5 ottobre 2026

Perché ne parlo io

**Marinella
Petrocchi**

Ricercatrice

IIT-CNR
Pisa

- Cybersecurity e sicurezza dell'informazione
- Falsi account, disinformazione e fonti poco affidabili
- Ricerca su intelligenza artificiale e comportamento online

Dal 2012 studio come l'informazione può essere manipolata e diffusa online.

Fake news: non solo errori

DEFINIZIONE

Una notizia inventata, che imita il contenuto della vera nella forma, ma senza un controllo editoriale rigoroso

BUONA FEDE

Disseminazione involontaria di notizia falsa, non con intenzioni malevoli:
MISINFORMAZIONE

MANIPOLAZIONE

Contenuto costruito o modificato per ingannare, influenzare o ottenere attenzione: DISINFORMAZIONE

Weapons of Mass Distraction: Foreign State-Sponsored Disinformation in the Digital Age
Park Advisors, 2019

Foto vera, didascalia errata



- Manifestazione contro il riscaldamento globale
- London's Hyde Park, 2029
- Mumbai, India
- Sempre Hyde Park, dopo un festival di musica

Quando i social non erano onnipresenti..

2008: UN GIORNALISTA IRLANDESE S'INVENTA DI SANA PIANTA LESISTENZA DI UN GIOVANE CALCIATORE, SEDICENNE, DALLA MOLDAVIA, UN FENOMENO

NOME, FOOTBALL TEAM, MOLDAVIAN NEWSPAPER: TUTTO INVENTATO

TRA I 50 GIOCATORI PIU' PROMETTENTI NELLA CLASSIFICA DEL TIMES
L'ARSENAL ESPRIME INTERESSE

Image source: indosport.com



Perché si diffondono così bene?

EMOZIONI

rabbia, paura, sorpresa

VELOCITÀ

un clic e il contenuto circola

CONFERMA

ci piace ciò che conferma le nostre idee

VISIBILITÀ

le piattaforme premiano attenzione e interazioni

Una condivisione non è una verifica.

L'AI cambia la scala del problema

Prima: falsificare richiedeva tempo, competenze e risorse.

Oggi: strumenti generativi possono produrre rapidamente contenuti molto credibili.

T

Testo



Immagini



Voci



Video

Il problema non è che l'AI crea: è che può creare in grande quantità, a basso costo e con grande realismo.

Riconoscete queste immagini?



Deepfake: quando vedere non basta

Un deepfake è un contenuto audio o video sintetico o manipolato che sembra autentico.

- Voce clonata
- Volto o labiale alterati
- Video di una persona che non ha mai pronunciato quelle parole

Possibili conseguenze

- Truffe
- Danno reputazionale
- Propaganda e confusione

“Vedo e sento” non equivale più automaticamente a “è accaduto davvero”.

Prima di credere a audio/video/foto: 3 controlli

1

CHI LO PUBBLICA?

È un account identificabile? È una fonte affidabile?

2

DOV'È L'ORIGINALE?

Cerca il video completo, non solo uno spezzone.

3

CHI LO CONFERMA?

Controlla se fonti indipendenti e affidabili riportano lo stesso fatto.

Un chatbot (es: ChatGPT) non è una fonte

Un chatbot può essere utile per:

- Spiegare un concetto
- Riorganizzare idee
- Fare una prima bozza
- Suggestire parole chiave per cercare

Ma non è automaticamente affidabile per:

- Notizie attuali
- Dati, date e statistiche
- Citazioni e bibliografie
- Informazioni su persone reali

Usalo come assistente: non come autorità finale.

Come genera una risposta?

In modo molto semplificato: prevede quale parola o pezzo di parola è più probabile dopo il testo precedente.



Fluidità linguistica ≠ verifica dei fatti

Può produrre una frase ben scritta anche se l'informazione non è corretta.

Allucinazioni dell'AI

Risposte plausibili e sicure... ma non corrispondenti ai fatti.

- Fatti o date inventati
- Citazioni mai pronunciate
- Libri, articoli o link inesistenti

Non è “una bugia” nel senso umano.

Il sistema genera una risposta plausibile anche quando non ha informazioni sufficienti o verificate.

Una risposta con un tono molto sicuro non è necessariamente una risposta sicura.

Allucinazioni dell'AI

Un mercato nell'antica Roma, anno 50 DC



Tre regole da portarsi a casa

1

FERMATI

Non condividere di impulso.

2

VERIFICA

Cerca fonte, data, contesto e conferme.

3

CONFRONTA

Non usare una sola fonte, né umana né artificiale.

L'AI non rende tutto falso. Rende la verifica ancora più importante.

L'AI può aiutarti a pensare. Non deve sostituirsi al tuo giudizio.

Grazie — domande?