

**L'Intelligenza Artificiale nei giornali**

**Sfide e Opportunità per i Publisher**

**Romolo Velati - CTO MONRIF**

## “Move fast and break things – Motto di Facebook fino al 2014.



Il fenomeno dell'AI è travolgente e difficile da arginare e controllare, simile a quanto è avvenuto con l'avvento delle altre piattaforme web.

Ogni giorno ci sono novità, siamo in un momento instabile e non si è ancora capito quale sarà il nuovo punto di equilibrio dopo questa trasformazione

**Siamo solo all'inizio di questa grande trasformazione, che minaccia i già fragili equilibri delle news digitali.**

# Cosa può fare l'AI Gen nelle news

---

Mentre l'**AI è già utilizzata** in diverse piattaforme digitali a supporto del giornalismo online ( es: monetizzazione adv e adaptive paywall ) l'arrivo dell'**intelligenza generativa** ha posto il giornalismo ( e tutti i lavori di tipo intellettuale ) di fronte ad una minaccia / opportunità . Chi ha utilizzato una qualsiasi AI ha ben chiaro che essa potrebbe:

- > creare articoli di senso compiuto
- > citare fonti e riferimenti
- > riassumere il testo
- > creare immagini
- > creare titoli, metadati per il SEO



Si, potrebbe fare quasi tutto questo **ma ...**

# Le allucinazioni nell'AI

---

**Le allucinazioni nell'IA si verificano quando un sistema di intelligenza artificiale produce risultati non veritieri o non coerenti con la realtà.**

In parole semplici, l'IA "immagina" cose che non esistono o fornisce informazioni errate.

**Le cause possono essere diverse:**

- **Modelli troppo complessi:** Reti neurali troppo elaborate possono generare dati non realistici durante l'addestramento.
- **Dati di input di scarsa qualità:** Se l'IA viene alimentata con dati errati o incompleti, può produrre risultati inaffidabili.
- **Mancanza di contesto:** L'IA potrebbe non avere abbastanza informazioni per comprendere correttamente la richiesta o la situazione, portando a risposte illogiche.



**Anche se il tasso di allucinazione sperimentato è circa il 3%, i possibili errori che potrebbe generare, non permettono l'auto pilot, ma solo il **copilot**, ovvero la necessità di un giornalista che controlli e validi l'output dell'AI**

# Esempi di allucinazione

**Gannett AI sbaglia gli articoli sportivi delle scuole superiori**  
Nell'agosto del 2023, il gruppo editoriale Gannett ha annunciato di voler sospendere l'uso di uno strumento di intelligenza artificiale chiamato LedeAI, dopo che diversi contenuti delle sue testate scritti dall'intelligenza artificiale erano stati virali perché erano mal scritti, ripetitivi, e privi di dettagli-chiave.

**LIA per il recruitment di iTutor Group che rifiuta i candidati a seconda della loro età**  
Nell'agosto del 2023, la società di tutoring iTutor Group ha accettato di pagare 5.000 dollari [in inglese] per risolvere una causa [in inglese] intentata dalla US EEOC.

**Sports Illustrated potrebbe aver pubblicato articoli di falsi scrittori generati dall'IA**  
Nel novembre del 2023, la rivista online Futurism ha scoperto che Sports Illustrated stava pubblicando articoli di scrittori non in carne e ossa ma generati dall'intelligenza artificiale [in inglese].

## Hallucination Leaderboard [↗](#)

Public LLM leaderboard computed using Vectara's Hallucination Evaluation Model. This evaluates how often an LLM introduces hallucinations when summarizing a document. We plan to update this regularly as our model and the LLMs get updated over time.

Last updated on November 1st, 2023

Model	Accuracy	Hallucination Rate	Answer Rate	Average Summary Length (Words)
GPT 4	97.0 %	3.0 %	100.0 %	81.1
GPT 3.5	96.5 %	3.5 %	99.6 %	84.1
Llama 2 70B	94.9 %	5.1 %	99.9 %	84.9
Llama 2 7B	94.4 %	5.6 %	99.6 %	119.9
Llama 2 13B	94.1 %	5.9 %	99.8 %	82.1
Cohere-Chat	92.5 %	7.5 %	98.0 %	74.4
Cohere	91.5 %	8.5 %	99.8 %	59.8
Anthropic Claude 2	91.5 %	8.5 %	99.3 %	87.5
Mistral 7B	90.6 %	9.4 %	98.7 %	96.1
Google Palm	87.9 %	12.1 %	92.4 %	36.2

# Il giornalismo di qualità non è sostituibile

---

The  
New York  
Times

Il direttore del New York Times: «Nessun bot o intelligenza artificiale sostituirà un giornalista che bussava a una porta, pone domande e trova notizie»

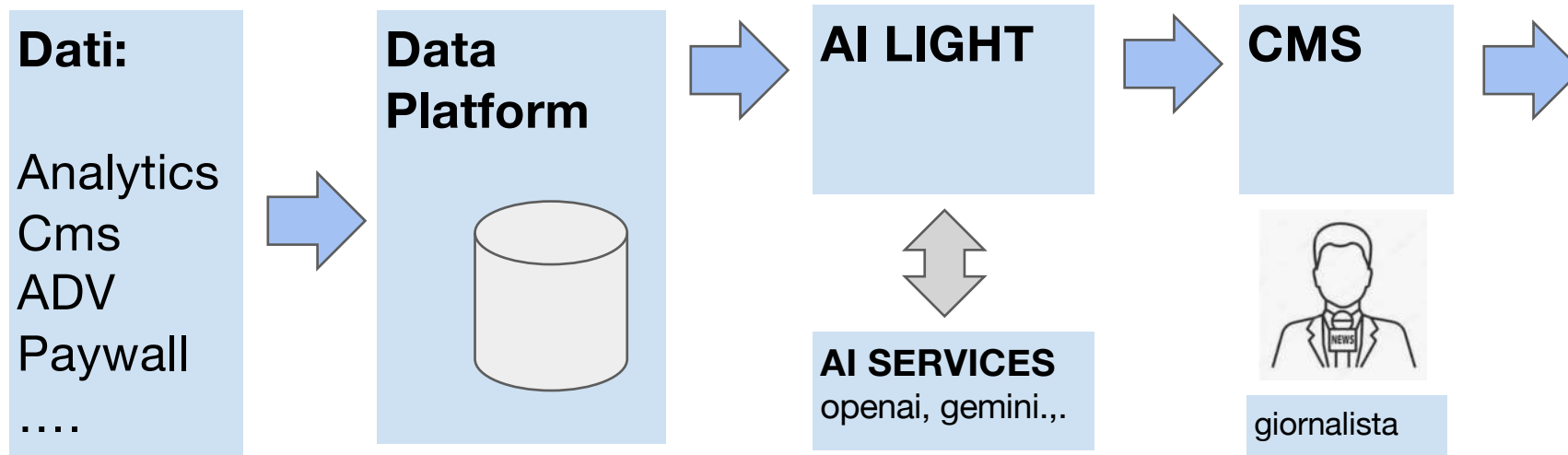


Questo perché il giornalismo di qualità richiede un livello di **creatività, intuito, giudizio etico e capacità di indagine** che l'AI attualmente non può replicare.

L'AI può assistere nei compiti ripetitivi e nell'elaborazione dei dati, **ma non** può comprendere il contesto complesso, le sfumature culturali e l'empatia necessaria per raccontare storie umane in modo significativo e responsabile.

# Monrif: Come impieghiamo l'AI in modo sicuro

I 2 punti principali di governance dell'AI in Monrif sono la Data platform e il sottosistema Ai-Light con il quale veicoliamo al giornalista tramite il CMS i suggerimenti dell'AI



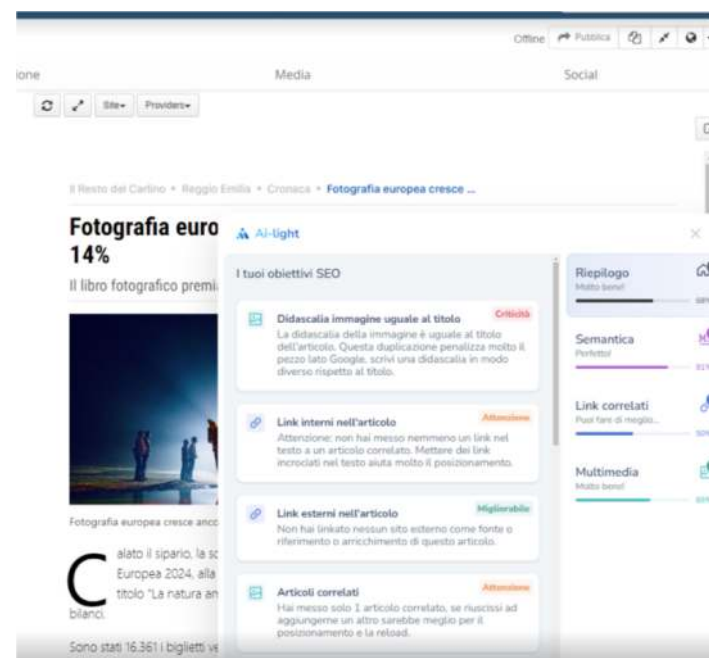
# Ai Light - Il giornalista al centro.



The screenshot shows a CMS editor interface. At the top, there are tabs for 'Generale' and 'Info'. Below, the article title is 'Fotografia europea cresce ancora. Aumentati i biglietti del 14%'. An 'Ai-light' panel is open, displaying three suggestions for the article title:

- Fotografia Europea 2024 chiude con successo: oltre 75mila visitatori e nuovi record
- Bilancio positivo per Fotografia Europea 2024: crescita del 14% nei biglietti venduti
- Dal Chiostro di San Pietro a Villa Zironi: Fotografia Europea 2024 celebra l'arte viva

The article text below the suggestions reads: 'Fotografia europea cresce ancora. Aumentati i biglietti del 14%'. Sono stati 16.361 i biglietti venduti (+14% rispetto al 2023) per 35.732 presenze contate ai Chiostri di San Pietro, a Palazzo de' Medici, a Palazzo Mediceo e a Villa Zironi, sedi in cui si



The screenshot shows a CMS editor interface with an 'Ai-light' panel displaying SEO analysis for the article 'Fotografia europea cresce ancora. Aumentati i biglietti del 14%'. The analysis includes:

- Didascalia immagine uguale al titolo** (Critica): La didascalia della immagine è uguale al titolo dell'articolo. Questa duplicazione penalizza molto il prezzo lato Google: scrivi una didascalia in modo diverso rispetto al titolo.
- Link interni nell'articolo** (Attenzione): Attenzione: non hai messo nemmeno un link nel testo a un articolo correlato. Mettere dei link incrociati nel testo aiuta molto il posizionamento.
- Link esterni nell'articolo** (Migliorabile): Non hai linkato nessun sito esterno come fonte o riferimento o arricchimento di questo articolo.
- Articoli correlati** (Attenzione): Hai messo solo 1 articolo correlato, se riuscisci ad aggiungere un altro sarebbe meglio per il posizionamento e la reload.

On the right side, there are progress bars for 'Riepilogo' (82%), 'Semantica' (81%), 'Link correlati' (50%), and 'Multimedia' (55%).



**Il giornalista, puo' decidere cosa scegliere e verificare se l'articolo ha tutti i contenuti necessari per il web. Il prossimo punto sarà utilizzare un LLM trainato sui nostri contenuti.**



# Dove siamo con la maturità del modello ?



may you live in interesting times ...

## FUTURO: WASHINGTON POST WANT TO BE AN “AI FIRST COMPANY”

---

### The Washington Post

Il Washington Post è sicuramente uno dei Publisher più avanzati tecnologicamente, il loro CTO ( Vineet Khosla ) ex di Apple e Uber, ha una chiara idea dell'utilizzo dell'AI, non solo nel settore giornalistico, dove ha già creato un LLM specifico per il suo giornale, trainato con gli articoli in modo da avere lo stesso TOV, ma ha coniato il termine, **AI First Company**, ovvero impiegare in ogni settore, es: legal, finance, hr .. tutti gli strumenti dell'AI disponibili.



**Questo è forse l'approccio piu' pragmatico, rispetto al NYT di cui evidentemente si rispettano le premesse rispetto al giornalismo di qualità, ma allo stesso tempo abbraccia l'innovazione per non perdere la sfida complessiva che l'AI pone.**

# FUTURO: WASHINGTON POST WANT TO BE AN “AI FIRST COMPANY”



**BACKUP**

## Sarà un futuro AI FIRST per le aziende ?

When poll is active  
respond at

**PollEv.com**  
**/romolosamuelevelati560**



**Credete che l'intelligenza artificiale (AI) sia destinata a diventare un elemento fondamentale per tutte le aziende, indipendentemente dal loro settore, portando a un modello di business "AI First"?**

# NYT vs OPENAI: Le accuse sono dimostrabili ?

---

## 1. Analisi dei Modelli Linguistici

Output di GPT-3 o GPT-4: Il NYT potrebbe aver analizzato le risposte generate dai modelli di OpenAI e identificato contenuti che sembrano derivare direttamente dagli articoli del NYT. Questo può includere citazioni, strutture di frasi, o informazioni uniche presenti negli articoli originali del NYT.

## 2. Riferimenti Espliciti o Impliciti

Contenuti distintivi: Il NYT potrebbe aver trovato riferimenti a contenuti distintivi o specifici che sono unici ai loro articoli, suggerendo che i modelli sono stati addestrati utilizzando tali materiali.

## 3. Metadati e Tracciabilità

Tracciare l'uso dei dati: Se OpenAI avesse utilizzato articoli del NYT che contengono metadati o informazioni di tracciamento, potrebbe essere possibile per il NYT identificare l'uso non autorizzato attraverso l'analisi dei dati di addestramento di OpenAI.

## 4. Riconoscimento del Pattern

Analisi comparativa: Tecniche di riconoscimento del pattern e di analisi del testo possono essere utilizzate per confrontare grandi volumi di output dei modelli di OpenAI con il corpus degli articoli del NYT, cercando similarità significative.

## 5. Testimonianze e Documentazione Interna

Dichiarazioni di dipendenti: Se ci sono state comunicazioni o documentazioni interne di OpenAI che fanno riferimento all'uso dei contenuti del NYT per l'addestramento dei modelli, queste potrebbero costituire prove importanti.

Testimonianze: Testimonianze di dipendenti o ex-dipendenti di OpenAI che possano confermare l'uso dei contenuti del NYT potrebbero essere utilizzate

**Non ci dovrebbero essere dubbi sull'utilizzo dei dati, da capire che danno ha provocato al NYT e quale possa essere l'impatto economico.**

# Definizione di Fair Use e Fair Dealing

---

## 1. Fair Use (Uso Equo)

Definizione: Il "fair use" è una dottrina legale negli Stati Uniti che permette l'uso limitato di materiale protetto da copyright senza ottenere il permesso del titolare del copyright, per scopi specifici come critica, commento, notizia, insegnamento, studio e ricerca.

Fattori di Fair Use: La legge statunitense considera quattro fattori per determinare se un uso rientra nel fair use:

Scopo e carattere dell'uso: Uso commerciale vs. non profit, trasformativo vs. copiativo.

Natura dell'opera protetta: Opere creative (es. romanzi) vs. opere fattuali (es. articoli di notizie).

Quantità e sostanzialità della parte utilizzata: La porzione del lavoro originale utilizzata.

Effetto dell'uso sul mercato potenziale: Impatto economico sull'opera originale.

## 2. Fair Dealing (Trattamento Equo)

Definizione: Simile al "fair use", il "fair dealing" è una dottrina presente in molte altre giurisdizioni, tra cui il Regno Unito e il Canada. Anche qui, l'uso di materiale protetto da copyright è consentito per determinati scopi come ricerca, studio, critica, recensione, notizia e satira.

Limitazioni: Ogni giurisdizione ha regole e limitazioni specifiche su cosa costituisce "fair dealing".