

## 4. Analisi delle fonti online

*di Paola Napoli e Valentina Russo*

In questa fase iniziale ci siamo concentrati sulle fonti online per estrapolare i principali strumenti utilizzati dalle redazioni con una particolare attenzione per quelle al di fuori dell'Italia. Lo scopo principale è stato quello di individuare lo stato attuale e le tendenze dell'automatizzazione del lavoro nelle redazioni. Nelle testate giornalistiche analizzate convergono sia strumenti di miglioramento del lavoro del giornalista (in termini di tempo e monitoraggio), sia strumenti di intelligenza artificiale che sostituiscono alcune mansioni redazionali. Un'ulteriore traccia utile a capire il cambiamento che stanno affrontando le redazioni è l'inserimento, in alcune testate, di figure professionali a supporto della gestione e dell'utilizzo di questi software all'interno del proprio organigramma.

La ricerca ha evidenziato un panorama di nuovi strumenti a supporto delle testate giornalistiche estremamente disomogeneo. Alcune testate usano tool con una sola funzione, mentre altre si avvalgono di suite composte da più strumenti capaci di svolgere numerose funzioni ed automatizzare compiti che prima richiedevano necessariamente un intervento umano. La ricerca si è inoltre concentrata sul grado di proprietà degli strumenti, ossia se si tratta di software sviluppati autonomamente dalle testate oppure prodotti da società terze che offrono i propri tool e servizi.

Che ruolo hanno o assumeranno in futuro questi software nelle redazioni? Per rispondere a questa domanda abbiamo scelto di suddividere i software analizzati in base alle loro funzioni nelle diverse fasi di produzione delle notizie. Le funzioni prese in considerazione sono presentate di seguito.

### *Analisi e monitoraggio trend*

I giornalisti si avvalgono di strumenti utili non solo ad estrapolare dati ma anche a tenere sotto controllo le pubblicazioni online e sui social di determinate keywords. Il monitoraggio dei trend rappresenta una fase fondamentale che precede la scrittura dell'articolo stesso. I software analizzati, che svolgono funzioni di monitoraggio, richiedono al momento l'intervento umano e non sono totalmente automatizzati.

### *Ricerca dati*

Si tratta di software che effettuano ricerche sulla base di database ed archivi, interni o esterni alle testate. Alcuni software effettuano scansioni sul web e sono spesso usati a supporto del giornalista per procurare dati utili ad articoli di tipo meteorologico, finanziario e sportivo. Questi algoritmi velocizzano per il giornalista la fase meccanica di acquisizione dei dati che precede la scrittura di un articolo. Altri algoritmi più evoluti con l'implementazione di tecnologie per processare il linguaggio naturale (NLP) arrivano ad estrapolare dati attraverso la comprensione dei testi.

### *Scrittura automatizzata*

I software di scrittura automatizzata sono quelli che presentano maggiori tratti di automazione. Questi hanno la capacità di partire da dati e/o trend scansionati dal web ed elaborare testi in linguaggio naturale, leggibile e pubblicabile dalla testata stessa. I software di scrittura automatizzata possono generare in autonomia piccoli articoli e report utilizzando template e algoritmi di Natural Language Generating. Questi software vengono spesso affiancati da strumenti di ricerca trend o dati affinché l'algoritmo sia completo nel processo di ricerca/elaborazione /scrittura. Talvolta il software è in grado di chiudere il ciclo, provvedendo anche alla pubblicazione dell'articolo in maniera del tutto automatizzata.

### *Pubblicazione programmata*

I software di pubblicazione programmata rientrano principalmente nell'area di gestione di social, dove vengono utilizzati algoritmi per pub-

blicare testi in linea con i trend e con le aspettative del pubblico di ogni specifica testata. La scelta dei titoli o la pubblicazione dell'articolo può comunque avvenire manualmente grazie all'intervento umano, ma l'opportuno momento della pubblicazione e l'uso di determinate keyword può essere completamente automatizzata.

### *Moderazione commenti*

I software di moderazione commenti sono utili alle grandi testate per tenere costantemente sotto controllo le community dei propri lettori sui siti delle testate e sui canali social. Il compito di moderazione avviene in modo autonomo, di fatto eliminando la fase di controllo preliminare svolta da operatori umani.

### *Monitoraggio utenti*

Il monitoraggio utenti serve per rilevare il sentiment dei lettori, per analizzare il lettore tipo della specifica testata ed individuarne i gusti e gli interessi. Il monitoraggio è uno strumento utile ai giornalisti per correggere il tiro sugli argomenti di maggiore interesse e sugli orari di maggior navigazione. Questi software forniscono statistiche e grafici sul comportamento degli utenti, dunque sono a supporto degli editori e non sono processi completamente automatizzati.

### *Altre funzioni*

Altri software hanno compiti più specifici o sono ancora in fase di sperimentazione, ma sono stati comunque presi in considerazione come parte dei software nei quali investono le grandi testate. Alcuni di questi si occupano della divulgazione di notizie tramite bot user-friendly. Altri servono ad archiviare i dati sia interni che esterni alle redazioni. Mentre nel primo caso ci troviamo di fronte a software programmati in modo automatizzato, nel secondo caso sono utili strumenti di supporto alle redazioni. Di seguito si riportano i dati raccolti relativi ai software.

**Analisi e monitoraggio - Trend**

Usato/testato da	Descrizione utilizzo	Proprietario	Software	Ai Autonoma
CBC Internazionale	monitoraggio news	no	Ecosec	no
Sole24ore	Monitoraggio stampa online / offline	no	Mimesi	no
Adnkronos	Monitoraggio stampa online / offline	no	Mimesi	no
The Washington Post	Workflow	si	Arc Publishing	no
Forbes	Analizza e gestisce trend	si	Bertie	no
BBC	monitora account social per controllare i trend	no	CrowdTangle	no
Vox	monitora account social per controllare i trend	no	CrowdTangle	no
Buzzfeed	monitora account social per controllare i trend	no	CrowdTangle	no
Mic	monitora account social per controllare i trend	no	CrowdTangle	no
ProPublica	monitora account social per controllare i trend	no	CrowdTangle	no
The Telegraph	monitora i feed per argomento o dominio	no	BuzzSumo	no
BuzzFeed	monitora i feed per argomento o dominio	no	BuzzSumo	no
The Guardian	monitoraggio trends e comportamenti	no	Scorecard Researchi	no

**Ricerca dati pre-pubblicazione**

Usato/testato da	Descrizione utilizzo	Proprietario	Software	Ai Autonoma
Yahoo! Sport	Ricerca dati sportivi	no	Automated Insight	si
AccuWeather	Ricerca dati meteorologici	no	Automated Insight	si
Associated Press	Ricerca dati	no	Automated Insight	si
Bloomberg	Aggregatore di dati finanziari	si	Cyborg	si
Sportplatz Media	Natural Language Understanding, ricerca e comprensione dei dati interno ai database delle testate	no	Retresco	si
The Washington Post	Archivio dati di ogni tipo	no	Nicar	no
New York Times	Fornitore dati sportivi	no	Statsheet (Non più esistente)	si
Giornali del Press Association (PA)	Fornisce storie basate su dati di siti Web e news	si (PA + Urbs Media)	RADAR (Reporters and Data and Robots)	si
	Trasforma un testo scritto in linguaggio naturale in dati (Semantic Web Machine Reading)	no	Fred	no
News Argentine, UNO	Ricerca attendibilità dichiarazioni	no	Chequeabot	no
	Ricerca trend per fact checking	no	Fullfact	no

**Scrittura automatizzata**

Usato/testato da	Descrizione utilizzo	Proprietario	Software	Ai Autonoma
The Washington Post	genera aggiornamenti twitter e blog	si	Heliograf (Arc Publishing)	si
The Guardian	genera report unendo template e dati	si	ReporterMate	si
<i>Testato dal</i> The Guardian	NLP, genera articoli	no	GPT2	si
Los Angeles Times	articoli e report su dati specifici con pubblicazione	si	The Homicide Report	si
Los Angeles Times	articoli e report su dati specifici con pubblicazione	si	Quakebot	si
Forbes	Suggerisce titoli e immagini in base ai trend	si	Bertie	no
New York Times	Natural Language Generation	no	Arria	si
Forbes	Natural Language Generation	no	Arria	si
Sportplatz Media	Natural Language Generation su dati sportivi e meteorologici	no	Retresco	si
Giornalisti di Forbes	NLG	no	Narrative Science	si
Giornali del Press Association (PA)	Robot Writer NLG di local news	si (PA + Urbs Media)	RADAR (Reporters and Data and Robots)	si
DPG Media	Robot Writer fornisce dati relativi a partite di calcio in tempo reale	si (Ricerca Universit + DPG Media)	PASS	si

**Pubblicazione programmata**

Usato/Testato da	Descrizione utilizzo	Proprietario	Software	AiAutonoma
Le Monde	Post social	no	Echobox	si
Vogue	Post Social	no	Echobox	si
Sole24ore	Post Social	no	Echobox	si
Wired	Post Social	no	Echobox	si
The Guardian	Post Social	no	Echobox	si

**Moderazione commenti**

Usato/Testato da	Descrizione utilizzo	Proprietario	Software	AiAutonoma
The Washington Post	Moderazione commenti, newsletter	si	ModBot & Talk	si
New York Times	Moderazione commenti, newsletter	si	ModBot & Talk	si

**Monitoraggio utenti**

Usato/Testato da	Descrizione utilizzo	Proprietario	Software	AiAutonoma
Vox Recode	monitoraggio abitudini	no	Parsely	no
Pointer	monitoraggio abitudini	no	Parsely	no
Wired	monitoraggio abitudini	no	Parsely	no
The Guardian	monitoraggio competitors, principalmente strumento di marketing	no	SimilarWeb	no
The Atlantic	analytics del sito e dei social	no	Chartbeat	no
New York Times	analytics del sito e dei social	no	Chartbeat	no
The Washington Post	analytics del sito e dei social	no	Chartbeat	no
CNN	analytics del sito e dei social	no	Chartbeat	no
Le Monde	analytics del sito e dei social	no	Chartbeat	no
The Telegraph	analytics del sito e dei social	no	Chartbeat	no
The Buffalo News	analytics del sito	no	Metrics for News	no
Le Monde	monitoraggio comportamento	no	Amplitude	no

**Archiviazione**

Usato/Testato da	Descrizione utilizzo	Proprietario	Software	AiAutonoma
New York Times	condivisione ed archiviazione file via Google Docs	si	Library	no
BBC	aggregatore di notizie e tag per archiviazione	si	The Juicer	si

**News Bot**

Usato/Testato da	Descrizione utilizzo	Proprietario	Software	AiAutonoma
New York Times	Botification of news, chat user friendly per chiedere news	si	Sam Manchester	si

**AD**

Usato/Testato da	Descrizione utilizzo	Proprietario	Software	AiAutonoma
Channel 4	ai che analizza i contenuti per contestualizzare le pubblicità	si	Contextual Moments	si

I software analizzati sono 35 e abbracciano una o più funzioni. Oltre alla suddivisione per funzioni si è tenuto conto se i software in questione siano stati sviluppati dalle testate stesse o da compagnie esterne. Il 58,8 % dei software non sono proprietari, spesso le testate si avvalgono di professionisti esterni o si impegnano nello sviluppo di nuovi strumenti in collaborazione con Università. Molti strumenti sono attualmente ancora in fase embrionale, ma le tendenze di ricerca suggeriscono una maggiore automatizzazione delle mansioni più meccaniche, come la ricerca dei dati o il monitoraggio.

Le testate utilizzano spesso strumenti di recupero dati sul web o interni al proprio database per abbattere le tempistiche di pubblicazione. Infatti, le pubblicazioni in cui maggiormente ci si avvale di strumenti automatici coincidono con pubblicazioni a cadenza regolare, basate principalmente su dati esterni, come aggiornamenti meteo, finanziari o sportivi. I software analizzati in tabella sono stati considerati come “AI Autonome” quando sono capaci di svolgere autonomamente un’intera mansione del giornalista o di altri membri della redazione (come per la moderazione dei commenti online), oppure di estrapolare dai dati un prodotto finito e pronto alla pubblicazione (come nel caso dei report finanziari).

Il 64,7 % delle AI Autonome sono software proprietari e si occupano per lo più di scrittura automatizzata. I software si avvalgono di NLP e NLG creando testi leggibili capaci di essere anche immediatamente pubblicati. Si evince dunque una tendenza da parte delle grandi testate di automatizzare la scrittura dedicata alla stesura di news di aggiornamento.

I software più autonomi (che escludono quasi completamente l'intervento umano) svolgono funzioni di scrittura automatizzata, pubblicazione programmata e moderazione commenti.

Dai dati raccolti in tabella e dall'analisi delle fonti online emerge che fra le testate che guidano il fenomeno di automatizzazione delle redazioni c'è il New York Times, che non solo utilizza diversi software per svariate funzioni (si è potuto confermare che automatizza almeno 6 delle 9 funzioni analizzate), ma ha sviluppato grazie a team interni 3 dei 6 software presi in analisi per la testata.

Inoltre il New York Times ha dichiarato apertamente di avere un piano per lo sviluppo delle AI all'interno della propria redazione. Facendo riferimento a uno studio recentemente pubblicato *"New powers, new responsibilities. A global survey of journalism and artificial intelligence"* di Charlie Beckett, si evince che una delle maggiori cause che portano le testate a rimanere un passo indietro nel campo delle AI è proprio la mancanza di una strategia per la gestione e lo sviluppo delle tecnologie di automatizzazione. Analizzando i dati raccolti, è possibile effettuare una comparazione fra le testate emerse durante la ricerca. Nella seguente tabella sono riportati il numero di funzioni automatizzate dalle principali testate scelte come campione e la percentuale di proprietà dei software utilizzati.

Testata	% funzioni automatizzate	% di software proprietari
New York Times	67%	33%
The Washington Post	56%	33%
The Guardian	44%	11%
Le Monde	22%	0%

L'altro aspetto da tenere in considerazione per quanto riguarda il cambiamento delle redazioni è il personale impiegato in questo processo di automatizzazione. Confrontando le posizioni lavorative aperte in alcune testate con i dati del sopracitato studio *"New powers, new responsibilities. A global survey of journalism and artificial intelligence"*, si evince inoltre che l'uso delle AI nelle redazioni sta portando alla creazione di nuovi ruoli all'interno delle testate, piuttosto che alla sostituzione di quelli già esistenti. Prendendo ad esempio il The Guardian, è possibile osservare che sono richiesti profili come Data Analyst, Data Architect, Data Engineer e Data Scientist. Lo stesso New York Times ha posizioni aperte per ruoli come Software Engineer (Machine Learning), Web Engineer, Backend Software Engineer. Sta quindi emergendo la necessità di team specializzati all'interno delle testate per la creazione di tool proprietari o, nel caso di servizi esterni, di figure capaci di interfacciarsi con le compagnie che forniscono software. Tuttavia essendo un fenomeno ancora emergente, non si è creato un unico ruolo chiave comune a tutte le redazioni.