Individuare sistemi data-intensive nella Pubblica Amministrazione italiana: uno strumento basato su web scraping e machine learning

di *Enrica Amaturo, Biagio Aragona, Suania Acampa, Roberto Artiaco*[*]

Il lavoro presenta uno strumento capace di individuare ed esplorare automaticamente, attraverso il Machine Learning, l'impiego di sistemi data-intensive da parte della Pubblica Amministrazione (PA) italiana.

Nonostante la digitalizzazione della PA implichi l'uso di tecnologie basate sui dati, manca una panoramica di dove e come questi sistemi sono utilizzati e sui rischi che possono generare. Il lavoro addestra un algoritmo di machine learning su documenti e risorse web per individuare i sistemi di decisione automatica nella PA. Sono stati raccolti 15.087 contenuti tramite web scraping da siti ministeriali italiani, etichettati manualmente e usati per addestrare un modello BERT.

Parole chiave: datificazione; digitalizzazione della Pubblica Amministrazione; machine learning; sistemi Data-intensive; BERT; Web Scraping.

Identifying data-intensive systems in the Italian Public Administration: a tool based on web scraping and machine learning

This study presents a tool capable of automatically identifying and exploring the use of data-intensive systems by the Italian Public Administration (PA) through Machine Learning.

Despite the digitalization of the PA involving data-driven technologies, there is no clear overview of where and how these systems are used or the potential risks they may pose. This study trains a machine learning algorithm on documents and web resources to detect automated decision-making systems within the PA. A total of 15,087 contents were collected through web scraping from Italian ministerial websites, manually labeled, and used to train a BERT model.

Keywords: datafication, Public Administration digitalization, machine learning; Data-intensive system; BERT; Web Scraping.

DOI: 10.5281/zenodo.17297505

[*] Università degli Studi di Napoli Federico II. enrica.amaturo@unina.it, biagio.aragona@unina.it, suania.acampa@unina.it, roberto.artiaco@unina.it.

Ricerca finanziata da PRIN: Prot. 2022H2CFLZ – "Datafication and Citizenship: the case of Public Administrations in Italy"

Introduzione

Il lavoro presenta uno strumento capace di individuare ed esplorare in maniera automatica, attraverso il Machine Learning, l'impiego di sistemi data-intensive da parte della pubblica amministrazione italiana (PA)[1]. L'adozione di sistemi ad alto impiego di dati sta cambiando il rapporto tra istituzioni e cittadini già da molto tempo, con alcuni passaggi chiave come il Piano d'azione per l'e-government del 2000, lo sviluppo della piattaforma PagoPA e, più di recente, il Programma Strategico Intelligenza Artificiale 2022-2024, che pone tra i suoi obiettivi anche l'impiego responsabile e trasparente dell'IA nella pubblica amministrazione, in linea con il principio di *AI a misura d'uomo* delineato nel documento ufficiale (PCM, 2022). Nonostante queste iniziative abbiano promesso significativi progressi nella gestione e nell'erogazione dei servizi pubblici, restano aperte questioni legate alla qualità dei dati, alla trasparenza dei modelli algoritmici e alla protezione dei diritti dei cittadini. Se da un lato i sistemi data-intensive mirano a migliorare l'efficacia nell'allocazione delle risorse, personalizzare i servizi e aumentare l'efficienza, dall'altro emergono criticità legate alla possibilità che i bias presenti nei dati di addestramento si riflettano nelle decisioni pubbliche, amplificando pregiudizi o generando discriminazioni (Buolamwini, Gebru, 2018; Mehrabi *et al.,* 2023). Per questo motivo, quando strumenti come sistemi decisionali automatizzati, algoritmi di raccomandazione o sistemi di ranking vengono impiegati, è fondamentale che le PA ne comunichino l'uso in modo trasparente, garantendo ai cittadini un accesso chiaro alle informazioni relative al loro funzionamento. Tuttavia, nonostante il crescente utilizzo, questi sistemi rimangono spesso opachi. L'accesso alle informazioni sulla loro adozione risulta complesso, malgrado le disposizioni normative come l'art. 22 del Codice dell'Amministrazione Digitale (D.lgs. 82/2005), che impone trasparenza e conoscibilità delle decisioni automatizzate nella PA. La mancanza di trasparenza e la scarsa comunicazione su questi strumenti genera un divario informativo che ostacola la conoscenza e la valutazione del loro impatto sulle politiche pubbliche e sulla cittadinanza (Veale *et al*., 2018; Pasquale, 2015; Wirtz *et al.,* 2022). In questo scenario, il presente studio si inserisce nel dibattito più ampio del processo di digitalizzazione dei contesti istituzionali, proponendo uno strumento volto a migliorare la trasparenza e la comprensione dei sistemi data-intensive impiegati nella PA.

[1] Il lavoro è parte del progetto PRIN "*Datafication and Citizenship: the case of Public Administrations in Italy"* (Prot. 2022H2CFLZ), volto a esplorare come la datificazione della PA stia trasformando il concetto di cittadinanza in Italia.

Negli ultimi anni, l'adozione di tecnologie basate su machine learning e big data da parte delle pubbliche amministrazioni ha alimentato un ampio dibattito sulla cosiddetta governance algoritmica, ovvero l'impiego di sistemi computazionali per supportare, indirizzare o automatizzare decisioni pubbliche, spesso in assenza di forme strutturate di trasparenza e accountability (Kitchin, 2014). Al centro di questi processi si colloca la datificazione, cioè la trasformazione di fenomeni sociali, interazioni e pratiche amministrative in dati strutturati, interpretabili da modelli predittivi o classificatori (van Dijck, 2014). Questa automazione tende a ridurre la complessità sociale a mere correlazioni statistiche, con il rischio di produrre decisioni discriminatorie (Couldry, Mejias, 2019). In questo contesto, alcuni autori propongono di interpretare l'algoritmo come una nuova forma di razionalità regolativa, in continuità o in rottura con la burocrazia moderna. Chiara Visentin (2018), ad esempio, analizza questa ambivalenza in profondità, mostrando come l'algoritmo, al pari della burocrazia weberiana, tenda a standardizzare le procedure, a inibire la discrezionalità e ad estendere il controllo attraverso regole formalizzate. Sulla base di un ampio confronto teorico, l'autrice individua tre principali interpretazioni: una visione continuista, secondo cui l'algoritmo rappresenta una burocrazia digitale più efficiente e automatizzata; una prospettiva critica, che ne denuncia l'opacità e l'invisibilità sociale, con effetti di depoliticizzazione; e una lettura discontinua, che introduce concetti come *algocrazia* o *infocrazia*, per descrivere forme di potere computazionale che agiscono modellando direttamente l'ambiente decisionale. Visentin sottolinea inoltre che l'algoritmo incarna una razionalità performativa e non discorsiva, che tende a espellere la dimensione valoriale e deliberativa dell'azione pubblica, con il rischio di compromettere le basi simboliche e democratiche della sfera pubblica. Questa riflessione si rivela particolarmente utile per comprendere l'adozione crescente di sistemi data-intensive nella pubblica amministrazione, i quali non solo razionalizzano i processi decisionali, ma trasformano profondamente i presupposti normativi e comunicativi del rapporto tra Stato e cittadini.

Inserito in questa cornice, il lavoro si ispira a iniziative internazionali consolidate, come *Algorithm Tips* (http://algorithmtips.org/), un database statunitense che analizza l'uso degli algoritmi nei processi decisionali governativi. Questo progetto fornisce informazioni sugli algoritmi impiegati a livello federale, statale e locale, illustrandone il funzionamento, l'ambito di applicazione e le implicazioni per i cittadini (Stark, Greene, 2020). Il progetto è stato rilevante nel dibattito sulla trasparenza algoritmica, evidenziando sia le opportunità di questi sistemi, sia i rischi legati alla loro opacità e ai bias sistemici. Allo stesso modo, il nostro lavoro mira a sviluppare un

database automatizzato per la PA italiana, offrendo un repertorio strutturato sui sistemi data-intensive in uso con l'obiettivo finale di fornire una piattaforma accessibile a diversi attori, raccogliendo documenti sulle applicazioni data-intensive nel settore pubblico. Questo strumento aiuterà a colmare il divario informativo derivante dalla frammentazione delle piattaforme web, dalla mancanza di politiche di trasparenza e accountability (Stankovich *et al.,* 2023) e dal rumore generato dalla dispersione degli open data, spesso privi di metadati adeguati (Sadiq, Indulska, 2017). I contenuti sono stati selezionati attraverso parole chiave e acquisiti tramite web scraping dai principali siti ministeriali. Per garantire una consultazione efficiente, è stato sviluppato un modello di classificazione basato su BERT (Devlin *et al.,* 2019), capace di categorizzare autonomamente i documenti e renderli successivamente accessibili tramite un'interfaccia utente che fungerà da motore di ricerca.

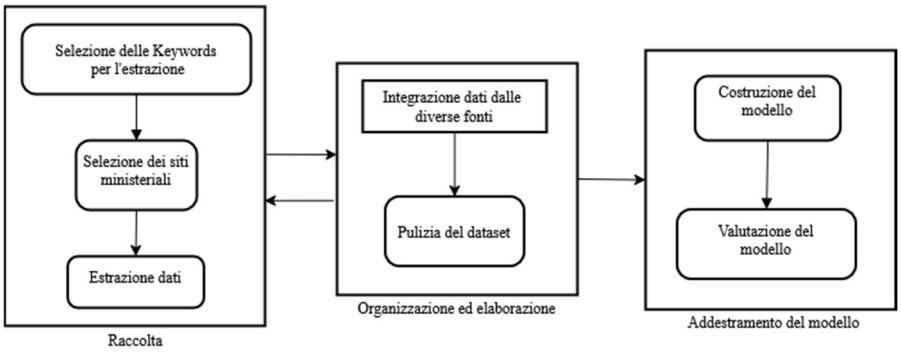
1. La creazione dello strumento

Il lavoro muove da due obiettivi: il primo è costruire un database interamente consultabile in cui sono presenti i contenuti dei siti ministeriali italiani riguardanti l'adozione di sistemi *data-intensive,* ossia sistemi che gestiscono ed elaborano grandi quantità di dati (terabyte e petabyte) generalmente complessi e distribuiti (Bonner *et al.,* 2017) – come open data e big data – e dedicano gran parte del loro tempo di elaborazione nelle fasi di input/output e manipolazione dati (Middleton, 2010). Il secondo obiettivo è addestrare un modello di Machine Learning (ML) che permetta, oltre ad una consultazione specifica per singole tematiche, anche la possibilità di inserire contenuti nuovi che saranno automaticamente riconosciuti dallo strumento in base alla loro rilevanza. Il fine ultimo è quindi di fornire uno strumento fruibile da cittadini ed esperti, attraverso il quale è possibile reperire informazioni aggiornate riguardo le applicazioni di sistemi data-intensive nella PA italiana. L'impiego del ML nel campo delle scienze sociali sta diventando un approccio consolidato per analizzare fenomeni complessi (Molina, Garip, 2019); in particolare, le reti neurali e i modelli di apprendimento supervisionato hanno dimostrato un grande potenziale nell'elaborazione di testi (Young *et al.,* 2018) e nella classificazione semantica. Nel nostro caso, il modello di ML è stato costruito seguendo tre fasi (figura 1):

1. *Raccolta:* che prevede la selezione delle keyword e la verifica della fattibilità dell'estrazione dai siti ministeriali.

2. *Organizzazione ed elaborazione*: dopo l'estrazione, il dataset è stato ripulito rimuovendo caratteri speciali e verificando la coerenza tematica con le keyword, integrando progressivamente i dati da diverse fonti.
3. *Addestramento del modello*: Il dataset è stato suddiviso in training e test set; quindi, il modello è stato valutato tramite metriche di accuratezza del ML.

Fig. 1 - Diagramma di costruzione dello strumento in fasi.



1.1. Fase 1 – Raccolta dati

Per raccogliere i dati dai siti ministeriali, il primo passo è stato definire una lista di parole chiave attinenti al tema, basata sul progetto Algorithm Tips, che fornisce una raccolta di keywords per individuare sistemi algoritmici negli USA. Questa lista è stata adattata e integrata per rispondere al contesto italiano. Dopo la selezione delle keywords (tab. 2), è stata analizzata la struttura dei siti ministeriali con dominio.gov.it e di altri siti istituzionali, come beniculturali.it, nei quali dovrebbero essere contenute le informazioni rilevanti sui sistemi data-intensive adottati dalla PA. Esaminata la struttura dei siti web individuati, sono emerse alcune problematiche comuni e condivise tra diversi siti. Per affrontare le difficoltà riscontrate, è utile descrivere la tecnica di estrazione dati adottata: la raccolta dei contenuti è avvenuta tramite web scraping, ovvero l'estrazione e il salvataggio strutturato delle informazioni visualizzate su pagine web (Lotfi *et al.,* 2021); in questa ricerca, è stato implementato un web scraping automatizzato in Python, utilizzando Selenium. L'estrazione ha riguardato i titoli e gli URL dei contenuti HTML,

evitando testo superfluo per ridurre il rumore (Agarwal *et al.,* 2007) e il rischio di errore. Selenium ha permesso di interagire direttamente con i siti, aggirando restrizioni come l'impossibilità di copiare URL specifici o l'assenza di operatori booleani.

Tuttavia, alcuni ostacoli hanno limitato l'efficacia del processo, tra cui elementi HTML nascosti, blocchi dei server, incoerenza nei risultati rispetto alle keyword e carenza di contenuti pertinenti. Queste problematiche, intrinseche alla struttura dei siti web ministeriali, hanno ridotto il numero di portali da 15 a 9. I siti sui quali è stato possibile raccogliere i dati sono:

- Ministero dei beni culturali;
- Ministero dell'interno;
- Ministero del lavoro;
- Ministero dell'agricoltura, della sovranità alimentare e delle foreste;
- Ministero dell'economia e delle finanze;
- Ministero delle imprese e del made in Italy;
- Ministero dell'istruzione e del merito;
- Ministero dell'Università e della Ricerca;
- Ministero del turismo.

1.2. Fase 2 – Organizzazione ed elaborazione

Attraverso il web scraping sui siti selezionati, l'uso delle keywords ha restituito 15.087 contenuti. Per ridurre la ridondanza, è stata applicata la lemmatizzazione delle keywords, considerando la limitata capacità di alcune piattaforme di riconoscerle solo nella loro forma base. Alcune di queste keywords presentano una scarsa autonomia semantica (Marradi, 1995: 34), risultando ambigue e poco valide per l'individuazione di fonti sui sistemi data-intensive. Per migliorare la pertinenza, è stata effettuata una selezione filtrando i circa 15.000 contenuti iniziali. In particolare, alle keywords "calcolo", "punteggio", "valutazione", "monitoraggio", "gradazione", "numeric" e "software" è stata associata almeno un'altra keyword tra quelle individuate, per raffinare i risultati. Le tabelle 1 e 2 mostrano il numero totale di contenuti inizialmente estratti e quelli rimanenti dopo la pulizia.

Tab. 1 – Frequenza dei documenti estratti prima e dopo la pulizia.

Ministero	*Frequenza (estrazione)*	*Frequenza (pulizia)*
Beni culturali	21	9
Interno	628	62
Lavoro	12568	1966

Masaf	250	10
Mef	71	37
Mimit	1158	183
Miur	203	122
Mur	176	63
Turismo	12	6
Totale	**15087**	**2458**

Tab. 2 – Frequenza delle Keywords prima e dopo la pulizia.

Keyword	*Frequenza (estrazione)*	*Frequenza (pulizia)*
algoritm	171	161
apache	17	17
automa	242	239
big data	53	51
biometric	17	17
calcolo	2815	82
classificazione	1210	36
cloud	228	227
data access	21	12
data capture	2	2
data governance	10	7
data lake	2	2
data mining	10	10
data privacy	11	9
data processing	16	16
data science	186	186
data source	5	2
data storage	4	4
data system	11	10
data warehouse	49	49
deep learning	4	3
ETL	6	6
GPT	3	3
graduazione	178	0
hadoop	2	2
intelligenza artificiale	173	173
internet of things	36	32
machine learning	23	19
metadat	87	84
monitoraggio	1747	26
neural network	7	7
numeric	56	54
open data	194	190
predittiv	109	102
punteggio	1190	29
rango	428	418
software	1290	42
statistica	45	45

valutazione	4429	84
Totale	**15087**	**2458**

È utile fare alcune considerazioni sulla struttura del dataset: la forte riduzione dei contenuti tra la prima e la seconda estrazione evidenzia la scarsa presenza di queste tematiche sui siti ministeriali, anche per keyword diffuse come "intelligenza artificiale", "algoritm*" e "big data". Quasi l'80% dei contenuti proviene dal Ministero del Lavoro. Le keyword inizialmente ritenute poco pertinenti hanno confermato la loro irrilevanza, diminuendo significativamente di frequenza. A ogni contenuto è stato assegnato un punteggio di 0 (non rilevante) o 1 (rilevante), creando una variabile per valutare la pertinenza dei temi. I contenuti classificati come rilevanti seguono i criteri indicati nelle tabelle 1 e 2. Successivamente, è stato effettuato un ulteriore controllo manuale, che ha portato a una lieve riduzione del numero complessivo di contenuti.

1.3. Fase 3 – Addestramento del modello

Per l'addestramento è stato scelto il modello BERT (Bidirectional Encoder Representation for Transformer) (Devlin *et al.,* 2019), che offre vantaggi significativi rispetto ai modelli NLP tradizionali. Grazie alla sua architettura basata su una rete neurale profonda, BERT elabora l'intero testo in input simultaneamente, migliorando l'identificazione delle relazioni tra parole e frasi (Bolasco, 2021). Per la classificazione del testo, BERT utilizza un transformer multistrato bidirezionale, che rappresenta il testo in un iperspazio a N dimensioni considerando l'intero contesto di ogni parola. Inoltre, il suo stato di preaddestramento su una vasta gamma di documenti gli consente di apprendere strutture linguistiche complesse e di essere ottimizzato per task specifici, come la classificazione binaria del testo. Dal dataset totale, composto da 15.087 osservazioni, sono state selezionate casualmente 2.458 rilevazioni non rilevanti, da affiancare alle 2.458 classificate come rilevanti, in modo da ottenere un campione bilanciato al 50% ed evitare problemi di oversampling. Il dataset è stato suddiviso in training e test set con un rapporto 80/20 (Gholamy *et al.,* 2018); l'addestramento, eseguito con la tecnica "1cycle" (Smith, Topin, 2018), è durato due epoche, sufficienti a garantire un buon livello di accuratezza senza rischi di overfitting.

Tab. 3 – Parametri di perdita e accuratezza del modello.

Epoche	*Loss*	*Accuracy*	*Val_loss*	*Val_accuracy*
1	0,37	0,83	0,19	0,95

2	0,1	0,96	0,08	0,97

Com'è possibile notare dalla tabella 3 infatti, il modello, alla fine della seconda epoca, presenta un'accuratezza di 0,96 nel caso del training set, e di 0,97 nel test set. Anche i valori di perdita risultano bassi (0,1 e 0,08). Per fornire un ulteriore *layer* di validazione al modello, può essere utile fare riferimento ad alcune misure più specifiche, ossia la *precision*[3], il *recall*[4] e la media armonica dei due, definita f*1- score.*

Tab. 4 – Parametri di validazione per rilevanza.

Rilevanza	*Precision*	*Recall*	*F1-score*
Non rilevante (0)	0,98	0,97	0,98
Rilvente (1)	0,97	0,98	0,98

Anche dalla tabella 4 possiamo notare come le misure scelte restituiscano valori molto alti, suggerendo una robustezza del modello elevata.

Conclusioni

Il lavoro contribuisce a colmare il divario informativo sull'uso dei sistemi data-intensive nella PA italiana, fornendo uno strumento innovativo per raccogliere e classificare documenti pubblici. L'efficacia del modello BERT è confermata dall'elevata accuratezza nella classificazione dei contenuti rilevanti. Tuttavia, alcune limitazioni rimangono, in particolare in riferimento alla ridotta quantità di dati disponibili nei siti ministeriali e la dipendenza dell'algoritmo dalla qualità dei documenti indicizzati. Non tutti i siti governativi sono adatti al web scraping; la bassa mole di contenuti disponibili sui siti ministeriali ha imposto una riduzione della lista iniziale di keywords per una maggiore focalizzazione su tematiche specifiche. Per migliorare il sistema, sarebbe utile ampliare il database con fonti aggiuntive, come Web of Science, e affinare la classificazione per maggiore precisione. Al tempo stesso, la selezione di keyword più mirate e la pulizia dei dati si sono rivelate fondamentali per garantire una raccolta più efficace. Nonostante questi limiti, lo strumento presenta importanti potenzialità applicative, sia in ambito accademico che istituzionale.

In prospettiva, questo strumento contribuisce anche a una più ampia riflessione sul ruolo che le tecnologie algoritmiche stanno assumendo nei processi di governance. La crescente datificazione delle pratiche amministrative e l'adozione di modelli predittivi e classificatori pongono infatti interrogativi

profondi sul piano della trasparenza, dell'equità e della trasformazione delle logiche decisionali pubbliche (Couldry, Mejias, 2019; Ananny, Crawford, 2018) considerando che i sistemi algoritmici non si limitano a razionalizzare i processi amministrativi, ma introducono una forma di razionalità performativa che tende a marginalizzare la dimensione discorsiva e deliberativa dell'azione pubblica (Visentin, 2018). Inoltre, in assenza di riflessioni critiche sui dati di partenza e sulle logiche di classificazione, l'adozione di strumenti di machine learning da parte della PA rischia di produrre effetti autoconfermativi: le istituzioni tendono a considerare legittimo e prioritario ciò che risulta nei dati – già distorti da condizioni strutturali diseguali – trascurando esigenze e fenomeni che sfuggono alla rappresentazione computazionale; rischiando così di rafforzare una visione parziale e diseguale della realtà sociale. In questo modo, gli algoritmi si configurano anche come prolungamento di alcune caratteristiche della burocrazia tradizionale, intensificando il controllo e inibendo ulteriormente la privacy e la partecipazione del cittadino (Visentin, 2018). È dunque fondamentale affiancare all'adozione tecnica degli algoritmi una riflessione critica su come vengono definiti e costruiti i criteri che regolano il funzionamento di questi strumenti automatizzati affinché non riproducano – o amplifichino – le disuguaglianze già presenti nel tessuto sociale.

Riferimenti bibliografici

Agarwal S., Godbole S., Punjani D., Roy S. (2007). How much noise is too much: A study in automatic text classification. *Seventh IEEE International Conference on Data Mining* (ICDM 2007) (pp. 3-12). New York: IEEE.

Bolasco S. (2021). *L'analisi automatica dei testi. Fare ricerca con il text mining*. Roma: Carocci.

Bonner S., Kureshi I., Brennan J., Theodoropoulos G. (2017). Exploring the evolution of big data technologies. *Software architecture for big data and the cloud* (pp. 253-283). Cambridge (MA): Morgan Kaufmann.

Buolamwini J., Gebru T. (2018). Gender Shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (77-91). New York: PMLR.

Couldry N., Mejias U.A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press.

Decreto Legislativo 7 marzo 2005, n. 82. *Codice dell'Amministrazione Digitale*. Art. 22, commi 1 e 2.

Devlin J., Ming-Wei C., Lee K., Toutanova K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv*. https://arxiv.org/abs/1810.04805
.

Gholamy A., Kreinovich V., Kosheleva O. (2018). Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. *Departmental Technical Reports (CS)*, University of Texas at El Paso.

Kitchin R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: SAGE.

Lotfi C., Srinivasan S., Ertz M., Latrous I., Manjushree S. (2021). Web scraping techniques and applications: A literature review. *SCRS Conference Proceedings on Intelligent Systems* (pp. 381-394). New Delhi: SCRS.

Marradi A. (1995). *Metodologia delle scienze sociali*. Bologna: il Mulino.

Mehrabi N., Morstatter F., Saxena N., Lerman K., Galstyan A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR),* 54(6), 1-35. https://doi.org/10.1145/3457607.

Middleton A.M. (2010). Data-intensive technologies for cloud computing. *Handbook of Cloud Computing* (pp. 83-136). Boston (MA): Springer.

Molina M., Garip F. (2019). Machine learning for sociology. *Annual Review of Sociology*, 45(1), 27-45.

Presidenza del Consiglio dei Ministri (2022). *Programma Strategico per l'Intelligenza Artificiale 2022-2024*. Roma.

Sadiq S., Indulska M. (2017). Open data: Quality over quantity. *International Journal of Information Management*, 37(3), 150-154.

Smith L.N., Topin N. (2018). Super-convergence: Very fast training of neural networks using large learning rates. *arXiv.* https://arxiv.org/abs/1708.07120.

Stankovich M., Behrens E., Burchell J. (2023). Toward meaningful transparency and accountability of AI algorithms in public service delivery. *Government Information Quarterly*, early access.

Stark B., Stegmann D., Magin M., Jürgens P. (2020). Are algorithms a threat to democracy? The rise of intermediaries: A challenge for public discourse. *Algorithm Watch Report*, 26.

van Dijck J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197-208.

Veale M., Van Kleek M., Binns R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-14). New York: ACM.

Visentin C. (2018). Il potere razionale degli algoritmi tra burocrazie e nuovi idealtipi. *The Lab's Quarterly*, 20(3), 47-72.

Young T., Hazarika D., Poria S., Cambria E. (2018). Recent trends in deep learning-based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3), 55-75.