

Legiferare con l'Intelligenza Artificiale

Manola Cherubini*, Francesco Romano**¹

* Ricercatore Istituto di Informatica Giuridica e Sistemi
Giudiziari del CNR
manola.cherubini@igsg.cnr.it

** Ricercatore Istituto di Informatica Giuridica e Sistemi
Giudiziari del CNR
francesco.romano@igsg.cnr.it

Abstract: The crisis of the legislative assemblies is a phenomenon observed for several years and has various causes. Since the 1970s, it was already referred to expert systems and more generally to information technologies to support the activity of the legislative offices and decision-makers. Nowadays, the tools provided by Artificial Intelligence can help the legislator and, in any case, the public decision-maker in enacting more effective laws and measures and therefore regulate society and provide for increasingly changing and complex needs. This paper highlights the possibilities currently offered by the evolution of linguistic models for processing natural language and outlines some particularly significant case studies in the legislative procedure while stressing critical issues.

Keywords: *procedimento legislativo, informatica giuridica, intelligenza artificiale, legimatica*

¹ Il paragrafo 1 è da attribuire a Francesco Romano, il paragrafo 2 a Manola Cherubini. I paragrafi 3 e 4 sono frutto di una riflessione comune dei due autori, che quindi sono da considerarsi co-autori alla pari.

1. Contesto della ricerca e quadro normativo

La crisi della democrazia è uno degli argomenti ricorrenti nel dibattito pubblico, scientifico e giornalistico. Sul tema la riflessione è molto ampia. Alcuni autori, ad esempio, lo declinano dal punto di vista della perdita di controllo da parte dello Stato o di altre autorità regolatrici, nei confronti di un “mercato aggressivo” che starebbe mettendo a rischio diritti quali quelli “alla privacy, alla conoscenza e al suo uso”, attraverso un “progetto commerciale famelico e completamente nuovo” definito “capitalismo della sorveglianza”². Saremmo alla presenza di un fenomeno molto pericoloso per la democrazia e la verità, il cui esito potrebbe condurre a sistemi politici in cui “l’autorità democratica” sarebbe “scavalcata da un’autorità computazionale esercitata dal capitale privato della sorveglianza”³.

Anche le piattaforme online se da un lato sono spazi aperti per la discussione pubblica e privata, dall’altro sarebbero “strumenti per la disinformazione, l’interferenza elettorale e la propaganda”⁴.

In realtà il duello tra tecnica e democrazia è aperto da molto tempo se già Aldous Huxley nel 1958 riteneva che fosse difficile sperare “che fiorisca la democrazia nelle società in cui il potere economico si concentra e si centralizza sempre di più. Ma il progresso della tecnologia ha portato, e sta portando proprio a questa centralizzazione del potere”⁵. Allo stesso modo Anders (già nella prima edizione del suo celebre saggio del 1956 e poi di nuovo nel 1980), parlando di rivoluzione della tecnica, teorizzava che essa non si sviluppasse nella direzione della libertà dell’uomo, bensì nella direzione del totalitarismo degli apparecchi⁶.

Ma quando si parla di crisi della democrazia si può anche fare riferimento al fenomeno che sta coinvolgendo le assemblee legislative e, in particolare, lo stesso strumento legislativo, che starebbe progressivamente perdendo di centralità, in un processo di sostanziale “arretramento della fonte legislativa” che sarebbe dovuto a vari fattori, quali l’incessante sviluppo del “processo di integrazione europeo”, l’incedere inesorabile di “fonti normative non previste dalla Costituzione”, la sempre maggiore rilevanza delle “fonti di soft-law”⁷.

² Zuboff 2019, p. 17. Sul problema del rapporto tra tecnologie della libertà e tecnologie del controllo si veda anche Rodotà 2021, p. 127.

³ Zuboff 2021, p. 37-38.

⁴ Kaye 2021, p. 30.

⁵ Huxley 1997, p. 250-251.

⁶ Anders 1992, p. 98.

⁷ Caretti 2010, p. 3.

In realtà il fenomeno di grave crisi dei parlamenti è osservato dalla prima metà del secolo scorso. Già nel 1946 si parlava di “spostamento di sovranità” dai parlamenti agli “uffici” quali “corpi sovrani dello Stato «illimitato» della società dei tecnici”⁸.

Una crisi che si manifesterebbe pure nel lessico, se è vero che parlare di governance e non di indirizzo politico denota, a livello linguistico, la “progressiva decadenza delle assemblee elettive”, messe in crisi dalla mancanza di “conoscenza e approfondimento per poter elaborare i pensieri e condurre le azioni conseguenti”⁹.

Tale crisi può essere ricavata anche da alcuni numeri. Da un lato si evidenzia come negli ultimi 10 anni “il Parlamento ha approvato tre volte meno leggi di quelle che approvava alla fine del secolo scorso”¹⁰; dal marzo 2018, mese di inizio della diciottesima legislatura “i 98 decreti legge varati hanno richiesto la bellezza di 1.178 decreti attuativi”, solo in parte emanati, mancandone 675 e dunque di fatto “il 57,3% delle norme approvate dal Parlamento è rimasto scritto solo sulla carta”¹¹. Ma la crisi del Parlamento e della legge non è da ricollegarsi solo a una questione di numeri. La pandemia da Covid-19 ha, infatti, rafforzato l’iniziativa legislativa governativa e quindi l’incidenza dei decreti-legge e la loro crescente multisettorialità. Si è manifestato, inoltre, un “diritto dell’emergenza” che si caratterizza per la “proliferazione di regimi speciali e derogatori rispetto alla normativa generale”¹².

La qualità della produzione normativa, poi, non può non risentire di ciò che una recente definizione ha colto come “occasionalismo normativo”, cioè un “nascere e morire di norme senza unitaria ragione e senza durata” che avrebbe come effetto quello di portare “all’anomia tucididea, a un provvedere da sé soli alla propria salvezza”¹³.

Ci troveremmo, infatti, di fronte a un processo di produzione normativa che “sciolto da qualsiasi ratio, trova causa in se stesso, in un’inesauribile produzione di regole – arbitrarie nel contenuto, oscure nel linguaggio, tremolanti nella durata – a cui l’individuo non è in grado di prestare ascolto e obbedienza”¹⁴.

Neppure lo stato di eccezione causato dalla pandemia, che avrebbe richiesto “proprietà e sobrietà di linguaggio”, è riuscito ad attenuare la nota situa-

⁸ Burnham 1947, p. 166, 169.

⁹ Aquilanti 2021, p. 63.

¹⁰ Cassese 2021a.

¹¹ Giannini 2021, p. 15. Anche l’autorevole *Rapporto sulla legislazione 2021* conferma che nel primo triennio della XVIII legislatura le leggi sono diminuite di quasi 30 unità rispetto al medesimo periodo della XVII legislatura. Vedi Osservatorio sulla legislazione 2021, p. 47.

¹² Osservatorio sulla legislazione 2021, p. 9.

¹³ Irti 2021, p. 26.

¹⁴ Ivi, p. 27.

zione di inflazione normativa e oscurità linguistica¹⁵, che si è manifestata ancora una volta in una “oscura prosa del caos legislativo”. Basti pensare al decreto, ormai passato alla storia, di ben 123.000 parole – 13 volte la Costituzione – fatto di rinvii e deroghe ad altre leggi (ben 131)¹⁶.

Tuttavia, il fenomeno della crisi dei Parlamenti è pure da ricercare in un ricorso improprio allo strumento legislativo, dovuto al fatto che, per vari motivi, gli uffici della pubblica amministrazione non riescono a regolare materie che potrebbero invece trovare il loro “sbocco naturale in una accurata decisione amministrativa”, anziché affiorare a livello parlamentare. Si ricorre dunque alla legge per interventi anche in settori marginali, alimentando così il c.d. “circolo vizioso della legislazione”¹⁷.

Scrivere leggi migliori dal punto di vista sostanziale, e non solo formale, può sicuramente contribuire ad attenuare tale crisi. Ogni legge dovrebbe essere in grado di “perseguire gli obiettivi che si propone”. A tal fine ogni norma dovrà rispettare le disposizioni della Costituzione e di tutte le altre norme in materia di produzione normativa, coinvolgere gli stakeholder nella progettazione e nella fase di istruttoria normativa, poter ricevere attuazione effettiva, armonizzarsi nell’ordinamento e, in definitiva, essere adeguata a “dare risposte alle domande che hanno costituito il presupposto dell’attività normativa”¹⁸.

Le tecnologie informatiche possono di certo aiutare il legislatore nel potenziare le proprie attività, anche al fine di “convogliare” all’interno dell’iter parlamentare le domande di partecipazione e le forme di dialogo e consultazione con la società civile¹⁹.

Il Libro bianco sull’Intelligenza Artificiale al servizio del cittadino, nella sua versione del marzo 2018, propone questa tecnologia anche per il “decisore pubblico”. Nel documento si prevede che l’Intelligenza Artificiale (IA) possa essere usata per “processare grandi moli di dati” e quindi permettere ai legislatori, di “ottenere una visione più chiara delle scelte che devono o che vogliono compiere”²⁰. In pratica l’IA potrebbe costituire “uno strumento

¹⁵ Su questi temi non si può prescindere da Ainis 1997. Ma si vedano, ad esempio, anche Recchia, Dickmann (a cura di) 2002; Carli 2007; Costanzo (a cura di) 2011; Picchi 2012.

¹⁶ Irti 2021, p. 63.

¹⁷ Cassese 2021b, p. 28.

¹⁸ Tafani 2017, p. 142-143. Laura Tafani ha ribadito che è assolutamente necessario aumentare la trasparenza del processo legislativo, introducendo pratiche di consultazione efficaci che coinvolgano cittadini e stakeholder, al fine di migliorare la base informativa del processo normativo. Vedi il suo intervento dal titolo “*Enhancing the quality of legislation: the Italian experience*” al convegno *Talking law in the EU: clear language, rule of law and legitimacy in the European legal space*, online 21-22 gennaio 2021.

¹⁹ Vedi Osservatorio sulla legislazione 2021, p. 17.

²⁰ Task force sull’Intelligenza Artificiale dell’Agenzia per l’Italia Digitale (a cura di) 2018, p. 30.

valido per valutare (anche preventivamente) il livello di fattibilità delle leggi e degli atti normativi, migliorando così la qualità delle decisioni”²¹.

Recentemente gli uffici del Parlamento italiano stanno dedicando specifici studi al tema.

In occasione della Conferenza interparlamentare “Legiferare meglio da una prospettiva digitale”, il Servizio studi del Senato e l’Ufficio rapporti con l’UE della Camera hanno predisposto un dossier per le Commissioni riunioni interparlamentari. La conferenza aveva lo scopo di “discutere il ruolo della digitalizzazione nel processo legislativo e di come garantire la conformità della nuova legislazione ai requisiti moderni”²². Nel dossier sono riportati i casi di tre Paesi: Francia, Estonia e Danimarca. Per la Francia si segnalano, tra le altre cose, dei “progetti faro” destinati all’uso di chatbot per facilitare l’accesso alle normative di alcuni settori²³. Per quanto riguarda il caso estone è messa in risalto l’importanza dell’infrastruttura dati *X-road* e dell’*ID* digitale nazionale obbligatorio; in particolare la piattaforma di interoperabilità *X-road* permetterà lo scambio dati tra vari attori del settore pubblico e privato²⁴. Per la Danimarca, infine, il dossier evidenzia il vasto programma di digitalizzazione teso a creare un ecosistema digitale in grado di sviluppare benefici effetti, anche per le attività del legislatore. Tra le varie linee guida, che si stanno perseguendo a tale fine, si segnala, ad esempio, la necessità di avere legislazioni semplici e distinte, per “contribuire a un’amministrazione più uniforme e digitale”, oppure legislazioni in grado di “sostenere la comunicazione digitale con cittadini e aziende”, così come di “consentire un’elaborazione digitale automatizzata dei casi”²⁵.

Altre proposte di utilizzo dell’IA per facilitare le attività delle assemblee e dei loro uffici possono essere segnalate.

Una Taskforce del governo dello stato australiano del Queensland ha prodotto un dossier, nel quale, fra le altre ipotesi, si propone che l’Intelligenza Artificiale venga utilizzata per assistere nel processo decisionale e come “assistente virtuale” in grado di rispondere alle domande più comuni degli utenti²⁶.

Così in un ulteriore report del 2017 si evidenziava, fra l’altro, il caso pilota del Ministero dell’economia giapponese, che sta sperimentando un sistema

²¹ D’Aloia 2019, p. 23.

²² Servizio studi del Senato e Ufficio rapporti con l’UE della Camera (a cura di) 2020, p. 9.

²³ Ivi, p. 16.

²⁴ Ivi, p. 20.

²⁵ Ivi, p. 26.

²⁶ Queensland Cabinet, Better Regulation Taskforce (eds.) 2018, p. 37, ove si afferma che la AI “is being used by regulators via algorithms to assist in decision making and as “virtual assistants” capable of answering common questions”.

di IA per aiutare gli uffici parlamentari a rispondere alle domande dei cittadini²⁷.

Posto che l'uso dell'Intelligenza Artificiale come strumento a disposizione del legislatore è argomento ancora poco trattato²⁸, il presente contributo evidenzia le possibilità attualmente offerte dall'evoluzione dei modelli linguistici per l'elaborazione del linguaggio naturale e, pur sottolineandone le criticità, ne prospetta l'utilizzo a supporto del legislatore, sia per soddisfare le sue necessità informative nella fase di progettazione e istruzione della legge in formazione, sia nella fase di verifica dell'impatto della norma rispetto alle altre norme dell'ordinamento, ma anche e soprattutto rispetto alla realtà da regolare, una realtà sempre più connessa, digitalizzata e quindi sempre più computabile²⁹.

2. Strumenti di Intelligenza Artificiale per l'elaborazione del linguaggio

Attualmente l'Intelligenza Artificiale ha fatto notevoli passi avanti soprattutto nell'elaborazione del linguaggio naturale (*Natural Language Processing - NLP*). Le tecniche di *Language Modeling (LM)* hanno raggiunto, infatti, risultati sorprendenti, peraltro in continua e veloce evoluzione, e sono utilizzate per attività quali le traduzioni automatiche, il riconoscimento vocale, l'information retrieval, l'etichettatura di porzioni di testo (labeling).

I risultati degli applicativi che usano tecniche di *Language Modeling* sono misurabili in base a specifiche attività di *NLP*, come ad esempio *Question Answering*, *Reading Comprehension*, *Natural Language Interface*, *Natural Language Inference*, *Paraphrasing*, *Natural Language Generation*, e testati su strumenti di benchmark³⁰, fra i quali possono essere presi come riferimento *GLUE (General Language Understanding Evaluation)*³¹, *SuperGLUE*³², *RACE*

²⁷ Mehr 2017, p. 8, ove si legge che il Ministero “is piloting a system to help parliament member offices respond to citizen questions by drafting answers using AI”.

²⁸ Stradella 2020, p. 3345.

²⁹ Carli 2021, p. 26. Nell'articolo si evidenzia come il mercato dell'Internet of Things (IoT) abbia conosciuto un'ascesa ininterrotta dal 2014 attestandosi al valore di mercato di 6 miliardi di euro nel 2020. In base ad alcuni studi già dal 2020 “il volume dei dati trattati a livello planetario sarà di 44 zettabyte”, in pratica si tratterebbe di “un multiplo di quasi 8 miliardi di volte tutte le informazioni conservate nella biblioteca del Congresso di Washington”. Vedi Casonato 2019, p. 713.

³⁰ Con il termine benchmark si intendono indicatori chiave di prestazione, parametri di riferimento. In ambito informatico gli strumenti di benchmark sono software che misurano la capacità di un altro software nell'eseguire attività determinate.

³¹ Wang, Singh, Michael, Hill, Levy, Bowman 2018, p. 353–355.

³² Wang, Pruksachatkun, Nangia, Singh, Michael, Hill, Levy, Bowman 2019.

(*ReAding Comprehension dataset from Examinations*)³³, *SQuAD (Stanford Question Answering Dataset)*³⁴.

In generale, i modelli linguistici si basano sul valore di probabilità assegnato ad una data sequenza di parole. Assunto di base è che il significato di una parola dipende dal numero n di parole che la precedono e che ne rappresentano il “contesto”.

A partire dai dati presi in input, ovvero i testi in linguaggio naturale forniti, i modelli linguistici generano per ogni unità elementare del testo (token) un vettore di probabilità³⁵, utilizzato per prevedere il token successivo, che a sua volta fungerà da input³⁶.

Si tratta, quindi, di modelli statistici³⁷, poi potenziati grazie all’uso di reti neurali³⁸.

È evidente che, a parità di qualità dei dati, i risultati ottenibili sono tanto più significativi quanto è maggiore la mole di dati che è possibile far gestire agli applicativi.

Le prestazioni dei modelli NLP hanno avuto, nel tempo, miglioramenti sempre più significativi e su porzioni di testo di lunghezza sempre maggiore. Negli ultimi anni, in particolare dal 2018, la natura sequenziale è stata superata dalla parallelizzazione. Le sequenze di input sono diventate, infatti, gestibili dalla macchina nello stesso tempo, considerando intere porzioni di testo contemporaneamente. Ciò ha permesso di accelerare l’addestramento dei sistemi nella loro capacità previsionale e di addestrare l’applicativo con set di dati sempre più numerosi³⁹.

³³ Lai, Xie, Liu, Yang, Hovy 2017.

³⁴ Rajpurkar, Zhang, Lopyrev, Liang 2016, p. 2383–2392.

³⁵ Vettori di probabilità sono vettori di numeri reali in grado di codificare il significato di parole o frasi e quelle con significato semantico simile sono rappresentate con vettori “vicini”.

³⁶ Fino a qualche anno fa, alcune delle architetture più avanzate per l’elaborazione del linguaggio naturale erano basate principalmente sul *Recurrent Neural Networks* e successivamente sull’architettura *Long Short-Term Memory*. Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin 2017; e Hochreiter, Schmidhuber 1997.

³⁷ I modelli statistici sono basati su *N-grams*, dove un *N-gram* è una sequenza di un numero n di token. Alcuni modelli sono stati poi costruiti secondo il modello di Markov, *Hidden Markov Models (HMM)*, che distingue eventi e stati. Nella catena di stati di Markov ogni stato con una certa probabilità produce un evento, ma solo gli eventi sono osservabili mentre gli stati sono nascosti. Blunsom 2004; e Morwal, Jahan, Chopra 2012.

³⁸ Goldberg 2017.

³⁹ Uno degli elementi che ha contribuito al notevole miglioramento delle performance dei modelli di elaborazione del linguaggio naturale è stato, infatti, il *Transformer*. Tramite meccanismi di auto-attenzione (*self-attention mechanism*), la rilevanza del token corrente è considerata rispetto a tutti gli stati precedenti, anche quelli più distanti e non necessariamente sequenziali, permettendo così la valutazione delle relazioni tra tutte le parole. Pathak 2021, p. 165-166; e Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin 2017.

Funzionale al miglioramento delle prestazioni dei modelli NLP, inoltre, è il sistema del doppio training.

Una prima fase, computazionalmente più onerosa, prevede il pre-addestramento su obiettivi non supervisionati, utilizzando corposi set di dati, ovvero testi non etichettati, tratti ad esempio da Wikipedia, da siti Web o da piattaforme social. Questi dataset permettono al modello di apprendere le abilità linguistiche fondamentali e la semantica generale nei vari domini. In questa fase la comprensione dei testi può essere migliorata grazie a sistemi quali quelli che mascherano casualmente alcuni token in una frase, con l'obiettivo di prevedere il loro valore originale, basandosi solo sul contesto delle altre parole presenti⁴⁰, oppure addestrano alla previsione della frase successiva, migliorando le prestazioni del sistema nelle attività di interfaccia del linguaggio naturale e di risposta alle domande⁴¹.

In una seconda fase, invece, i risultati vengono ulteriormente perfezionati in base ad altri set di dati supervisionati per specifiche attività definite a valle (fine-tuning)⁴². Possono essere, dunque, set di dati meno corposi, ma classificati e relativi ad un dominio più specifico.

Un primo modello linguistico per NLP è stato *BERT* (*Bidirectional Encoder Representations from Transformers*)⁴³, cui sono seguiti diversi nuovi modelli, che ne hanno ulteriormente migliorato le prestazioni nelle varie attività. Tra questi si citano *Transformer-XL*⁴⁴, *RoBERTa*⁴⁵, *ALBERT* (*A Lite BERT*)⁴⁶,

⁴⁰ *Masked Language Modeling (MLM)*. Devlin, Chang, Lee, Toutanova 2018.

⁴¹ *Next Sentence Prediction (NSP)*. Devlin, Chang, Lee, Toutanova 2018 e Shi, Demberg 2019.

⁴² Pathak 2021, p. 165-166.

⁴³ BERT ha permesso di pre-addestrare rappresentazioni bidirezionali profonde da testo non etichettato, considerando congiuntamente il contesto sinistro e destro in tutti i livelli. Devlin, Chang, Lee, Toutanova 2018.

⁴⁴ *Transformer-XL* ha migliorato la capacità di apprendere le dipendenze relazionali delle parole nel contesto, aumentando la lunghezza delle sequenze di testo considerate e la velocità di prestazione. Pathak 2021, p. 166.

⁴⁵ *RoBERTa* ha ripreso l'approccio di *BERT*, rendendolo però più robusto e ottimizzato, anche grazie al pre-addestramento su una base di dati molto più ampia. Liu, Ott, Goyal, Du, Joshi, Chen, Levy, Lewis, Zettlemoyer, Stoyanov 2019.

⁴⁶ *ALBERT* ha introdotto nella fase di pre-addestramento la *Sentence order prediction (SOP)* per la previsione del corretto ordine di occorrenza di frasi consecutive, al posto della *Next Sentence Prediction*. Lan, Chen, Goodman, Gimpel, Sharma, Soricut 2019.

*T5 (Text-to-Text Transfer Transformer)*⁴⁷, *PEGASUS*⁴⁸, *ELECTRA*⁴⁹ e la serie di modelli *GPT (Generative Pre-trained Transformer)*⁵⁰.

In particolare GPT-3 rappresenta un modello linguistico di terza generazione, che necessita di un numero di esempi e istruzioni molto più limitato nella fase di addestramento (few-shot learning) e si dimostra applicabile su qualsiasi attività di NLP, raggiungendo prestazioni simili a quelle umane nelle attività a valle della generazione del linguaggio naturale.

Se è vero che, a livello di sistemi generalisti, i modelli linguistici di ultima generazione riescono a raggiungere risultati eccellenti, utilizzando combinazioni di set di dati di grandi dimensioni, è altrettanto vero che, a livello di sistemi ristretti, con addestramento su domini specifici, le abilità del sistema possono arrivare a risultati molto più stabili.

Date le enormi possibilità applicative, è importante che anche il diritto, con le sue note specificità di linguaggio altamente tecnico e problematicità, riesca a cogliere la sfida offerta da questi modelli, ad oggi ancora poco sperimentati.

Tuttavia, l'applicazione pratica degli strumenti di Intelligenza Artificiale, sopra illustrati, deve necessariamente tener conto di almeno 3 aspetti tecnici.

- In primo luogo, i modelli linguistici, al di là del diverso livello delle loro performance nel trattamento del linguaggio naturale, sono addestrati su enormi moli di dati e necessitano quindi di una grande potenza di calcolo⁵¹.
- Una seconda considerazione va fatta sulla lingua utilizzata, in quanto questi modelli sono pre-addestrati, per la maggior parte o esclusivamente, in lingua inglese e non con un addestramento multilingua⁵².

⁴⁷ *T5* ha introdotto il *framework text-to-text* unificato, con cui diverse attività NLP sono risolte da un singolo modello pre-addestrato e non da modelli specifici perfezionati per attività a valle. Raffel, Shazeer, Roberts, Lee, Narang, Matena, Zhou, Li, Liu 2020.

⁴⁸ *PEGASUS* ha un modello di pre-addestramento basato sul *Gap Sentences Generation (GSG)*, ovvero genera riepiloghi astratti del testo da un input di frasi mascherate. Questa attività "auto-supervisionata" si avvicina molto all'addestramento specifico a valle, riducendo la mole di dati necessaria per la messa a punto e velocizzando tutto il sistema. Zhang, Zhao, Saleh, Liu 2020.

⁴⁹ *ELECTRA* presenta un modello di codificatore di testo a due trasformatori (generatore e discriminatore), pre-addestrato, distinguendo i veri token nella frase originale; non utilizza il *Masked Language Modeling* ma rileva il token sostituito (*RTD – Replaced Token Detection*). Clark, Luong, Le, Manning 2020.

⁵⁰ Brown, Mann, Ryder, Subbiah, Kaplan, Dhariwal, Neelakantan, Shyam, Sastry, Askell, et al. 2020. E vedi anche Gomez-Perez, Denaux, Garcia-Silva 2020, p. 20-21.

⁵¹ Possibili rimedi a questa criticità di natura tecnica possono essere forniti dall'uso di librerie, quali *PyTorch-Transformers* (in linguaggio *Python*), per mezzo delle quali avviare una sperimentazione che risparmi tutta la fase di pre-addestramento. In merito si veda, ad esempio https://pytorch.org/hub/huggingface_pytorch-transformers/.

⁵² De Mattei, Cafagna, Dell'Orletta, Nissim, Guerini 2020, p. 1.

- Un terzo tipo di criticità è connesso al fatto che, nei set di dati per il pre-addestramento vi possono essere stereotipi e pregiudizi umani che si propagano poi nel modello linguistico (ad esempio bias di genere)⁵³.

Per sopperire a queste due ultime criticità andrebbero, dunque, resi disponibili appositi dataset, costituiti da risorse in lingua italiana, non contenute in archivi generalisti reperibili in rete, ma del dominio legislativo o comunque appartenenti al linguaggio di testi giuridici in genere.

La ricerca scientifica, come affermato all'inizio del presente paragrafo, è in continua e rapida evoluzione al fine di superare le diverse criticità riscontrate, come dimostrano alcuni dei più recenti modelli linguistici.

Tra questi, ad esempio, *AfriBERTa*⁵⁴, per la possibilità di utilizzare una quantità molto inferiore di dati per le attività di training, *ByT5*⁵⁵, per non operare tramite token, ma direttamente sul testo (byte o caratteri), e, sempre per migliorare le capacità di apprendimento, FLAN (Finetuned LAnguage Net)⁵⁶ e LEXFIT (Lexical Fine-Tuning)⁵⁷.

3. Possibili utilizzi degli strumenti di IA nel procedimento legislativo italiano

Da molti anni, l'informatica si propone di supportare il procedimento legislativo. Da un lato con strumenti che siano di aiuto alla generazione del testo, dall'altro con strumenti e metodi in grado di gestire anche fasi successive a quella generativa, come quella della valutazione.

Se da un lato le regole di tecnica legislativa hanno costituito il logico e necessario antecedente per ideare e sviluppare strumenti che adottassero gli standard linguistici e formali proposti, dall'altro una serie di progetti hanno costituito l'altrettanto necessario ecosistema digitale che ha reso possibile che alcune tecnologie siano già utilizzate con successo, da anni, nel campo della comunicazione dell'informazione normativa.

Il progetto *Norme In Rete* ha costituito, infatti, la pietra miliare per una serie di standard e applicazioni che hanno permesso di rappresentare i testi normativi nel formato *XML*⁵⁸ e di annotare meta-informazione necessaria grazie ad appositi *DTD* e a strumenti in grado di supportare tale annotazione⁵⁹.

⁵³ I set di dati utilizzano, infatti, come fonti archivi online quali siti web, forum, social ecc.

⁵⁴ Ogueji, Zhu, Lin 2021.

⁵⁵ Xue, Barua, Constant, Al-Rfou, Narang, Kale, Roberts, Raffel 2021.

⁵⁶ Wei, Bosma, Zhao, Guu, Yu, Lester, Du, Dai, Le 2021.

⁵⁷ Vulić, Ponti, Korhonen, Glavaš 2021.

⁵⁸ Biagioli, Francesconi, Spinosa, Taddei 2004.

⁵⁹ Per gli obiettivi e i risultati del *Progetto NIR*, si veda Ciampi, Archi, Biagioli, Cherubini,

Oggi possiamo disporre di banche dati normative che hanno fatto tesoro di queste ricerche e che permettono di consultare il testo vigente e plurivigente delle norme e che, di sicuro, costituiscono applicazioni e basi di conoscenza strutturate di rilevante importanza nel settore dell'informatica legislativa. Si pensi al database di *Normattiva*, ma anche a molti repository di norme regionali che offrono modalità di consultazione semplificata e ipertestuale delle norme, o anche a database tematici che consentono la navigazione fra norme di vario livello ma afferenti al medesimo dominio⁶⁰.

Tali applicazioni, oltre a supportare il legislatore nella fase di comunicazione delle norme, possono essere utilizzate per una valutazione di impatto delle norme stesse sull'ordinamento nel quale vanno ad inserirsi.

Come noto, la valutazione d'impatto normativo non può prescindere dall'analisi dei nessi normativi, data la loro importante funzione per ricostruire la "dimensione intratestuale", che, assieme a "quella extratestuale", è fondamentale per valutare le conseguenze che l'emanazione di un nuovo testo normativo può comportare sull'intero ordinamento⁶¹. Sono state già ipotizzate, infatti, "valutazioni automatiche" che prendessero in considerazione "elementi testuali espliciti che di solito si manifestano in strutture puntualmente definite, sia giuridicamente, sia linguisticamente dalle regole legistiche e da altre regole normative molto vincolanti"⁶².

A tal fine sono stati sperimentati strumenti di *NLP* per riconoscere e marcare in linguaggio *XML* nessi normativi come i riferimenti testuali espliciti e le modifiche testuali esplicite⁶³. In alcuni casi sono stati anche realizzati prototipi che, associando al riconoscimento automatico di tali nessi normativi e alla loro marcatura, la modellazione *UML*, fornivano in via automatica la ricostruzione del testo coordinato e vigente⁶⁴.

Ad oggi gli strumenti di Intelligenza Artificiale per l'elaborazione del linguaggio, potrebbero essere proposti per una sperimentazione che abbia come finalità quella di supportare il legislatore durante la progettazione e la

Fameli, Guidotti, Marinai, Mercatali, Nannucci, Ragona, Serrotti, Socci, Spinosa 2000; e anche Ciampi, Guidotti, Ragona, Serrotti, Socci, Spinosa, Marchetti, Vitali, Roller, Megale, Seta 2001. Inoltre, Pietropaoli 2009, p. 113.

⁶⁰ Ci si riferisce, ad esempio, alla banca dati di normativa regionale presente sul sito della Regione Umbria, ma anche a quella del *Portale PAeSI* che comprende tutta la normativa italiana in materia di immigrazione annotata in XML. Cherubini, Romano 2014, p. 483-498.

⁶¹ Mercatali, Romano 2013, p. 81.

⁶² Ivi, p. 82, dove era stato proposto di valutare la legittimità costituzionale o la coerenza con gli indirizzi degli Statuti regionali, ma anche la conformità all'ordinamento comunitario e agli obblighi internazionali, il coordinamento con la normativa esistente o l'incidenza dei rinvii a future norme di attuazione.

⁶³ Mercatali, Romano, Dini, Bolioli 2002.

⁶⁴ Mercatali, Romano, Boschi, Spinicci 2005.

redazione delle leggi e sono molti gli scenari in cui si può pensare di prefigurare l'intervento di tali applicativi.

In particolare, sono stati presi in considerazione due diversi scenari, che presentano notevoli complessità di gestione per lo stesso legislatore.

Il primo e più immediato scenario è quello relativo alla gestione delle modifiche normative, ovvero alla abrogazione, sostituzione o integrazione, da parte di una norma, del testo di un'altra norma dell'ordinamento o di porzioni di esso.

Un primo esempio può essere quello in cui si ipotizza l'abrogazione di una legge, o di singole sue parti, al fine di verificarne l'impatto sull'intero ordinamento.

Consideriamo, ad esempio, la seguente modifica testuale esplicita introdotta dal Testo unico in materia di spese di giustizia (D.P.R. 115/2002).

Esaminiamo in particolare una porzione dell'articolo 299 recante "Abrogazioni di norme primarie" che recita:

"1. Dalla data di entrata in vigore del presente testo unico sono abrogate le seguenti disposizioni: (...) - del regio decreto 18 dicembre 1941, n. 1368 (disposizioni di attuazione del codice di procedura civile), gli articoli: da 38 a 43; 107; dell'articolo 134, come sostituito dall'articolo 3, della legge 7 febbraio 1979, n. 59, al secondo comma, il n.1, limitatamente alle parole da "o le ricevute" sino a "cancelleria e" e alle parole "diritto di chiamata di causa"; il n. 4, limitatamente alle parole "o ricevute di versamenti sui conti correnti postali"; il quarto comma, limitatamente alle parole "o ricevute di versamento sui conti correnti postali"; il settimo comma; l'articolo 137, come modificato dall'articolo 5 della legge 7 febbraio 1979, n. 59, secondo comma, limitatamente alle parole da "la quale" sino a "la somma dovuta"; l'articolo 137, commi terzo, quarto e sesto (...)."

Come si può vedere, la ricostruzione del quadro giuridico di riferimento è resa complicata da più di un fattore. L'articolo 299 sopra citato, infatti:

- opera una modifica multipla di più norme (*gli articoli: da 38 a 43; 107*);
- modifica una norma non nel suo testo storico, ma nel suo testo modificato (*dell'articolo 134, come sostituito dall'articolo 3, della legge 7 febbraio 1979, n. 59*);
- modifica singole sotto-partizioni di altre norme, ma limitatamente ad alcune parole (*il n. 4, limitatamente alle parole "o ricevute di versamenti sui conti correnti postali"*);
- modifica una partizione di un'altra norma limitatamente ad una porzione di testo non espressa, ma delimitata da alcune parole (*l'articolo 137, come modi-*

*ficato dall'articolo 5 della legge 7 febbraio 1979, n. 59, secondo comma, limitatamente alle parole da "la quale" sino a "la somma dovuta"*⁶⁵.

La questione diventa particolarmente interessante e di grande utilità per il legislatore, anche in relazione alla valutazione dell'impatto degli emendamenti proposti ad un disegno di legge in corso di approvazione.

Come noto, durante l'iter di approvazione di una legge possono essere proposte molte modifiche e spesso l'impatto che possono avere sul testo di legge portato in votazione non risulta così evidente per i deputati o i senatori nel corso del dibattito parlamentare.

Il secondo scenario preso in considerazione riguarda, invece, la fase dell'analisi di fattibilità dei testi normativi a livello regionale. Si tratta di un altro importante settore in cui, da molti anni, si studia la possibilità di utilizzare il supporto delle tecnologie informatiche⁶⁶. In particolare, nell'ambito della valutazione ex post, da sempre si rileva la permanenza di "elementi di criticità nella valutazione delle leggi e delle politiche pubbliche"⁶⁷.

Per fare fronte a queste criticità le regioni si sono dotate di appositi strumenti, anche normativi, che hanno, ad esempio, comportato l'introduzione delle c.d. 'clausole valutative', ovvero "particolari disposizioni, inserite nell'articolo, che indicano i tempi, i modi e gli obiettivi delle successive attività di controllo e di valutazione sull'attuazione della legge"⁶⁸.

La loro introduzione nelle più importanti leggi regionali è stata promossa dal marzo del 2002 con l'avvio del *Progetto CAPIRe*, nato per iniziativa dei Consigli regionali Emilia-Romagna, Lombardia, Piemonte e Toscana, per "offrire una risposta costruttiva ed originale ad un'insoddisfazione diffusa nei confronti delle tradizionali attività conoscitive svolte in seno alle assemblee legislative"⁶⁹. Tra gli altri obiettivi, infatti, il progetto si proponeva di "modificare la prassi legislativa corrente per introdurre nelle leggi specifici riferimenti ad attività di controllo e valutazione delle politiche da parte del Consiglio".

Le clausole valutative sono di solito usate in "quelle leggi che, per la materia trattata, assumono particolare rilevanza e richiedono una costante at-

⁶⁵ Peralto, la complessità di ricostruire il quadro dell'ordinamento normativo su cui va ad impattare la regolamentazione fornita da una nuova prescrizione normativa, ha dato spesso adito alla prassi legislativa della cosiddetta "abrogazione innominata", che si verifica quando il legislatore adotta formule quali, ad esempio, "sono abrogate tutte le norme incompatibili con la presente legge", lasciando all'interprete il compito di applicarne in concreto la casistica. Cfr. Rescigno 1983, p. 113.

⁶⁶ Si veda ad esempio Giachi 1997.

⁶⁷ Libertini, Mercatali, Romano 2013, p. 917.

⁶⁸ *Ibidem*.

⁶⁹ Si veda il sito web del Progetto CAPIRe, nella pagina dedicata alle "Origini" del progetto, di cui al link http://www.capire.org/il_progetto/origini/.

tività di monitoraggio della loro attuazione ad esempio perché impegnano rilevanti risorse finanziarie”⁷⁰.

Di seguito si riporta il testo dell’articolo 8 della legge della Regione Umbria 17 aprile 2014, n. 7, “Disposizioni per la somministrazione ad uso terapeutico dei farmaci cannabinoidi”:

“Art. 8 (Clausola valutativa)

1. A partire dal mese di aprile 2016 e successivamente con cadenza biennale, la Giunta regionale invia all’Assemblea legislativa una relazione dettagliata che renda conto dell’attuazione della legge.

2. In particolare la relazione deve contenere informazioni su:

a) il numero di pazienti trattati con medicinali cannabinoidi in ciascuna azienda sanitaria della regione, distinti per patologia e per tipologia di assistenza;

b) l’ammontare della spesa annua sostenuta per l’acquisto dei farmaci cannabinoidi e la sua incidenza sulla spesa farmaceutica del Servizio sanitario regionale;

c) eventuali criticità emerse nell’applicazione della legge, con particolare riguardo all’acquisto, distribuzione ed erogazione dei farmaci cannabinoidi, ivi comprese le eventuali difficoltà incontrate nel garantire continuità di trattamento al paziente secondo prescrizione medica”.

Seppure, in questo specifico caso, le richieste espresse nel testo legislativo per rendere conto dello stato di attuazione della legge appaiono abbastanza chiare, non è da sottovalutare lo sforzo occorrente per la raccolta, strutturazione e rielaborazione delle relative informazioni, affinché siano poi davvero leggibili da parte del legislatore. A questi fini, è possibile ipotizzare il supporto degli strumenti di IA nell’acquisizione dei dati richiesti per la valutazione e soprattutto nella loro strutturazione ed elaborazione.

Gli sforzi degli studi di informatica giuridica, fino ad oggi, si sono concentrati in particolar modo sulla definizione di standard, strumenti e applicativi software per la identificazione, marcatura, elaborazione e ricerca dell’informazione giuridica.

Attualmente le potenzialità dimostrate dai modelli linguistici sopra citati⁷¹, seppur nella consapevolezza dei limiti altrettanto ampiamente discussi in dottrina⁷², evidenziano l’opportunità di una maggiore sperimentazione anche in campo giuridico.

⁷⁰ Servizio Legislazione e Commissioni del Consiglio regionale Umbria (a cura di), *Rapporto legislazione 2013-2015*, p. 127.

⁷¹ Si veda, ad esempio, relativamente a GPT-3, Brown, Mann, Ryder, Subbiah, Kaplan, Dhariwal, Neelakantan, Shyam, Sastry, Askell, et al. 2020.

⁷² Si veda, ad esempio, sempre relativamente a GPT-3, Floridi, Chiriatti 2020.

Proprio per questo motivo è stata avviata, da parte degli autori, una sperimentazione che, a partire dagli scenari sopra descritti, si articolerà seguendo tre passaggi fondamentali:

- raccogliere un set di dati di esempi di addestramento per il dominio specifico, che dimostrino il comportamento corretto per un'attività desiderata;
- addestrare il sistema a imitare questi comportamenti;
- testare le sue prestazioni su esempi indipendenti e identicamente distribuiti⁷³.

Per ciascuno di questi passaggi, ma in modo particolare per la fase di training, le esperienze maturate e i risultati raggiunti a livello di trattamento dell'informazione giuridica sono stati e continueranno a essere fondamentali per affinare il modello linguistico e ottenere output sempre più efficienti⁷⁴.

Al di là di ogni discussione sulla reale capacità di “comprensione” da parte degli applicativi di Intelligenza Artificiale che si possono realizzare, ovvero sull'effettivo riconoscimento di una loro “intelligenza”, interessano le potenzialità di tali strumenti, non per proporre una sostituzione uomo-macchina, ma per supportare l'operatore del diritto⁷⁵.

4. Conclusioni

Accostare l'uso di tecnologie così potenti e pervasive al processo di formazione delle decisioni delle assemblee rappresentative solleva, naturalmente, perplessità, che devono essere valutate anche da parte dei giuristi.

Innanzitutto, bisognerà valutare se al procedimento legislativo siano applicabili, per analogia, risoluzioni già adottate dalla giustizia amministrativa in materia di decisioni automatizzate nel procedimento amministrativo. Il Tar del Lazio ha, infatti, stabilito che “le procedure informatiche, finanche ove pervengano al loro maggior grado di precisione e addirittura alla perfezione, non possano mai soppiantare (...) l'attività cognitiva, acquisitiva e di giudizio che solo un'istruttoria affidata ad un funzionario persona fisica è in grado di svolgere”⁷⁶.

Inoltre, esiste, come già sopra accennato, un problema posto dalla trasparenza che coinvolge la formazione delle decisioni⁷⁷, o semplicemente relativo a come si reperiscono le informazioni in sistemi informativi basati sull'Intel-

⁷³ Radford, Wu, Child, Luan, Amodei, Sutskever 2018.

⁷⁴ Zhong, Xiao, Tu, Zhang, Liu, Sun 2020, p. 3.

⁷⁵ Ivi, p. 9.

⁷⁶ Simoncini 2019, p. 75.

⁷⁷ È il tema della c.d. “trasparenza algoritmica” che entra in gioco quando i criteri con cui è formato l'algoritmo possono violare diritti fondamentali. Cfr. Alpa 2021.

ligenza Artificiale che siano addestrati con tecniche di *machine learning* e in particolare di *deep learning*⁷⁸.

Le così dette *black box*, non essendo spesso comprensibili nemmeno ai programmatori che le hanno elaborate, rischiano di rendere inutilizzabili tali strumenti. Una rete neurale, ad esempio, può raggiungere risultati molto più accurati di un sistema basato su regole di classificazione, ma la comprensione dei processi con cui tali risultati vengono ottenuti è estremamente complessa. Si può affermare che la spiegazione di un modello di apprendimento automatico è solitamente inversa rispetto alla sua accuratezza di previsione, per cui può essere problematico, se non impossibile, utilizzare nel mondo reale la decisione proposta dall'IA⁷⁹, in settori in cui la trasparenza e la possibilità di motivare la decisione presa sono requisiti irrinunciabili⁸⁰.

I dati hanno assunto una sempre più rilevante importanza nell'organizzazione della conoscenza⁸¹, ma la "accountability" dei dati utilizzati da tali sistemi è un requisito importante per un uso in ambito giuridico e più in generale per il loro uso in una "moderna e tecnologica democrazia"⁸², specie qualora si vadano ad impiegare per produrre decisioni di rilevanza pubblica.

Si tratta, ad esempio, di usare dati che non producano o riproducano i pregiudizi spesso insiti nella decisione umana. Un problema che potrà forse essere risolto sul versante degli "umani", fintanto che la "scelta degli algoritmi con cui equipaggiare la macchina" e la configurazione "degli scenari e degli esempi" con cui addestrarla - e in cui applicarne i risultati - sono "operazioni affidate all'operatore umano"⁸³. Tuttavia esiste il concreto pericolo che il 'pregiudizio' arrivi dalla elaborazione della macchina stessa, a prescindere dall'esattezza e dalla qualità dei dati elaborati. Basti pensare al noto caso del software *COMPAS*, basato sull'apprendimento automatico e utilizzato negli Stati Uniti per valutare la probabilità di recidiva di un imputato⁸⁴. La questione della qualità dei dati e della "non discriminazione", dunque, è senza dubbio centrale per poter pensare di usare l'Intelligenza Artificiale anche solo come sistema di supporto al reperimento delle informazioni da parte del legislatore.

⁷⁸ Bird, Fox-Skelly, Jenner, Larbey, Weitkamp, Winfield 2020, p. 2.

⁷⁹ Xu, Uszkoreit, Du, Fan, Zhao, Zhu 2019.

⁸⁰ Casonato 2019, p. 717. Secondo l'autore l'utilizzo di strumenti di questo tipo – che non rendono possibile motivare una decisione – renderebbe "sostanzialmente incostituzionale ogni sistema di giustizia automatizzata", *ivi* p. 718.

⁸¹ Sisti 2021, p. 13.

⁸² Zeno-Zencovich 2020, p. 24.

⁸³ Casonato 2019, p. 721.

⁸⁴ Bird, Fox-Skelly, Jenner, Larbey, Weitkamp, Winfield 2020, p. 15-16. Per un esempio riferito a GPT-2 anche Wallace, Feng, Kandpal, Gardner, Singh 2019.

Sul punto, l'articolo 22 del GDPR europeo, fatte salve specifiche eccezioni, sancisce il diritto per chiunque di non essere soggetto a decisioni automatizzate basate sulla profilazione, che producano effetti giuridici che lo riguardano o che incidono in modo significativo sulla sua persona⁸⁵.

La proposta di regolamento europeo denominata "Artificial Intelligence Act" (COM(2021) 206 final, dell'aprile 2021), inoltre, al punto 40⁸⁶ considera che alcuni sistemi di IA, destinati all'amministrazione della giustizia e dei processi democratici, dovrebbero essere classificati come "ad alto rischio", considerando il loro impatto potenzialmente significativo sulla democrazia, sullo stato di diritto, sul diritto a un processo equo e in generale sulle libertà individuali. A tali sistemi qualificati "ad alto rischio", dovrebbero applicarsi particolari vincoli in tema di qualità delle serie di dati utilizzati, di documentazione tecnica e di tracciabilità della stessa, di trasparenza di informazioni per gli utenti, di supervisione umana e di sicurezza informatica⁸⁷.

Tale elevata qualità dei dati sarebbe da considerarsi essenziale per le prestazioni di molti sistemi di Intelligenza Artificiale, in particolare quando vengano utilizzate tecniche che comportano l'addestramento di modelli, ciò per garantire che i sistemi di IA qualificati ad "alto rischio" si comportino nel modo previsto e non diventino fonte di discriminazione⁸⁸.

Tuttavia, nonostante le perplessità e criticità che possono essere rilevate, i numerosi repository di dati pubblici (legislazione statale e regionale, nonché atti della PA), che la legislazione degli ultimi anni ha richiesto che siano pubblicati in Rete, la possibilità di reperire la dottrina giuridica, nonché un cospicuo patrimonio di dati giurisprudenziali digitalizzati, possono costituire la necessaria premessa per l'addestramento di strumenti di IA, quali quelli sopra illustrati.

Inoltre, la presenza di numerosi centri di ricerca pubblici e di aziende private, con esperienza pluriennale nel trattamento di tali dati, può fornire il necessario ecosistema digitale in grado di sperimentare questo tipo di tecnologia anche al servizio del legislatore.

D'altra parte il Governo italiano, in un recentissimo documento di programmazione strategica per i prossimi anni, indica alcune linee di sviluppo relative a "la promozione dell'IA all'interno della PA e per la PA". In parti-

⁸⁵ Art. 22 del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati).

⁸⁶ Proposta di regolamento, *Artificial Intelligence Act* (COM(2021) 206 final, dell'aprile 2021), Punto 40, p. 28.

⁸⁷ Ivi, Punto 43, p. 29.

⁸⁸ Ivi, Punto 44, p. 29.

colare si suggeriscono varie azioni, tra le quali sembrano particolarmente utili, ai fini proposti in questo contributo, quelle volte a promuovere interoperabilità e dati aperti per creare modelli di IA oppure a creare un dataset comune di lingua italiana “su cui formare modelli linguistici di intelligenza artificiale”⁸⁹.

Bibliografia

- Ainis, M. 1997. *La legge oscura: come e perché non funziona*. Bari, Laterza.
- Alpa, G. 2021. *Intelligenza artificiale dentro la società*. In *Il Sole 24 Ore*, 5 dicembre 2021.
- Anders, G. 1992. *L'uomo è antiquato: Sulla distruzione della vita nell'epoca della terza rivoluzione industriale*. Torino, Bollati Boringhieri.
- Aquilanti, P. 2021. *Nella decadenza del Parlamento la crisi della democrazia*. In *L'Espresso*, 23 maggio 2021.
- Biagioli, C., Francesconi, E., Spinosa, P., Taddei, M. 2004. *Xml documents within a legal domain: Standards and tools for the italian legislative environment*. In *International Workshop on Document Analysis Systems* (pp. 413-424). Berlin, Heidelberg, Springer.
- Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E., Winfield, A. 2020. *The ethics of artificial intelligence: Issues and initiative. Study Panel for the Future of Science and Technology*. Brussels, European Union.
- Blunsom, P. 2004. *Hidden Markov Models*. University of Melbourne, Department of Computer Science and Software Engineering.
- Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. 2020. *Language models are few-shot learners*. arXiv preprint arXiv:2005.14165.
- Burnham, J. 1947. *La rivoluzione dei tecnici*. Milano, Mondadori.
- Caretti, P. 2010. *La “crisi” della legge parlamentare*. In *Osservatoriosullefonti.it*, 1/2010.
- Carli, M. 2007. *Materiali sulla qualità della normazione: Roma, 18 novembre 2005*. Firenze, FUP.
- Carli, S. 2021. *Anche l'Italia ha raggiunto i 100 milioni di cose connesse*. In *Affari e Finanza, La Repubblica*, 10 maggio 2021.

⁸⁹ Ministero dell'Università e della Ricerca, Ministero dello Sviluppo Economico e Ministro per l'Innovazione tecnologica e la Transizione Digitale (a cura di), *Programma strategico Intelligenza Artificiale 2022-2024*, novembre 2021, p. 32-33.

- Casonato, C. 2019. *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*. In *BioLaw Journal – Rivista di BioDiritto*, Special Issue, 2/2019: 711-725.
- Cassese, S. 2021a. *Lo stato della repubblica*. In *Il Foglio*, 11 maggio 2021.
- Cassese, S. 2021b. *Il circolo vizioso delle leggi*. In *Corriere della Sera*, 24 luglio 2021.
- Cherubini, M., Romano, F. 2014. *Informatica giuridica e pubblica amministrazione*. In Peruginelli, G., Ragona, M. (a cura di), *L'Informatica giuridica in Italia. Cinquant'anni di studi, ricerche ed esperienze*, Napoli, ESI.
- Ciampi, C., Archi, A., Biagioli, C., Cherubini, M., Fameli, E., Guidotti, P., Marinai, E., Mercatali, P., Nannucci, R., Ragona, M., Serrotti, L., Socci, F., Spinosa, P. 2000. *Studio di Fattibilità per la realizzazione del Progetto "Accesso alle norme in rete"*. In *Informatica e diritto*, 1/2000, Napoli, ESI.
- Ciampi, C., Guidotti, P., Ragona, M., Serrotti, L., Socci, F., Spinosa, P., Marchetti, A., Vitali, F., Rolleri, F., Megale, F., Seta, E. 2001. *Progetto NIR – Fase 2 "Accesso alle norme in rete". Recupero della normativa pregressa in formato XML e standards del progetto (DTD e URNs)*. In *Informatica e diritto*, 1/2001, Napoli, ESI.
- Clark, K., Luong, M.-T., Le, Q.V., Manning, C.D. 2020. *ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators*. ICLR. arXiv preprint arXiv:2003.10555.
- Costanzo, P. (a cura di) 2011. *La qualità della normazione nella dialettica governo-parlamento: strumenti e tecniche nella XVI legislatura*, Napoli, Jovene.
- D'Aloia, A. 2019. *Il diritto verso "il mondo nuovo". Le sfide dell'Intelligenza Artificiale*. In *BioLaw Journal – Rivista di BioDiritto*, 1/2019: 3-31.
- De Mattei, L., Cafagna, M., Dell'Orletta, F., Nissim, M., Guerini, M. 2020. *GePpeTto Carves Italian into a Language Model*. <https://arxiv.org/pdf/2004.14253.pdf>.
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. 2018. *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805.
- Floridi, L., Chiriatti, M. 2020. *GPT-3: Its Nature, Scope, Limits, and Consequences*. In *Minds and Machines*, 30.4: 681–694.
- Giachi, G. 1997. *Tecniche per l'analisi di fattibilità dei testi normativi*. Napoli, ESI.

- Giannini, M. 2021. *Pax draghiana e stanchi litigi della politica*. In La Stampa, 16 maggio 2021.
- Goldberg, Y. 2017. *Neural Network Methods for Natural Language Processing, Synthesis Lectures on Human Language Technologies*. Vol. 10, No. 1, Morgan & Claypool.
- Gomez-Perez, J.M., Denaux, R., Garcia-Silva, A. 2020. *A Practical Guide to Hybrid Natural Language Processing. Combining Neural Models and Knowledge Graphs for NLP*. Cham, Springer.
- Hochreiter, S., Schmidhuber, J. 1997. *Long short-term memory*. In Neural computation, 9(8): 1735– 1780.
- Huxley, A. 1997. *Il mondo nuovo. Ritorno al mondo nuovo*, Milano, Mondadori.
- Irti, N. 2021. *Viaggio tra gli obbedienti*. Milano, La nave di Teseo.
- Kaye, D. 2021. *Libertà vigilata. La lotta per il controllo di Internet*. Roma, Treccani.
- Lai, G., Xie, Q., Liu, H., Yang, Y., Hovy, E. 2017. *RACE: Large-scale ReAding Comprehension Dataset From Examinations*. arXiv preprint arXiv:1704.04683.
- Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R. 2019. *ALBERT: A lite bert for self-supervised learning of language representations*. arXiv preprint arXiv:1909.11942.
- Libertini, R., Mercatali, P., Romano, F. 2013. *Il rapporto sulla legislazione toscana 2012: luci e ombre tra rappresentanza, partecipazione e valutazione*. In Istituzioni del federalismo, fasc. 3-4: 917.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. 2019. *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv preprint arXiv:1907.11692.
- Mehr, H. 2017. *Artificial Intelligence for Citizen Services and Government*. Cambridge, Ash Center for Democratic Governance and Innovation.
- Mercatali, P., Romano, F. 2013. *I documenti dello Stato digitale. Regole e tecnologie per la semplificazione*. Borgo San Lorenzo, Ad.Es. Edizioni.
- Mercatali, P., Romano, F., Boschi, L., Spinicci, E. 2005. *Automatic Translation from textual representation of laws to formal models through UML*. In Proceedings of 18th Annual Conference on Legal Knowledge and Information Systems, JURIX 2005, Vol. 134: 71-80.
- Mercatali, P., Romano, F., Dini, L., Bolioli, A. 2002. *For the Automated Mark-up of Italian Legislative Texts in XML*. In Bench-Capon, T.J.M., Daskalopulu, A., Winkels, R.G.F. (eds.), “Legal Knowledge and Information Systems. Jurix 2002: The Fifteenth Annual Conference”, Amsterdam, IOS Press.

- Ministero dell'Università e della Ricerca, Ministero dello Sviluppo Economico e Ministro per l'Innovazione tecnologica e la Transizione Digitale (a cura di) 2021. *Programma strategico Intelligenza Artificiale 2022-2024*.
- Morwal, S., Jahan, N., Chopra, D. 2012. *Named entity recognition using hidden markov model (HMM)*. International Journal on Natural Language Computing (IJNLC), Vol. 1, No.4, December 2012: 15–23.
- Ogueji, K., Zhu, Y., Lin, J. 2021. *Small Data? No Problem! Exploring the Viability of Pretrained Multilingual Language Models for Low-resourced Languages*. In Proceedings of the 1st Workshop on Multilingual Representation Learning: 116-126.
- Osservatorio sulla legislazione 2021, *La legislazione tra Stato, Regioni e Unione europea. Rapporto 2021*.
- Pathak, A. 2021. *Comparative Analysis of Transformer based Language Models*. In CS & IT Conference Proceedings, Vol.11, n. 1.
- Picchi, M. 2012. *La qualità della normazione in Italia secondo l'OCSE*. In Osservatorio sulle fonti, fasc. 2/2012.
- Pietropaoli, S. 2009. *Tecniche e strategie di annotazione dei testi*. In Biagioli, C. (a cura di), *Modelli funzionali della legge. Verso testi legislativi autoesplicativi*, Firenze, EPAP.
- Queensland Cabinet, Better Regulation Taskforce (eds.) 2018. *Better Regulation Taskforce Regulatory review project*, December 2018.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. 2018. *Language models are unsupervised multitask learners*. Technical Report, OpenAi.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J. 2020. *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. arXivpreprint arXiv:1910.10683.
- Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P. 2016. *Squad: 100,000+ questions for machine comprehension of text*. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing: 2383–2392.
- Recchia, G., Dickman, R. (a cura di) 2002. *Istruttoria parlamentare e qualità della normazione*. Padova, Cedam.
- Rescigno, G.U. 1983. *Note preliminari sulle principali manchevolezze nella tecnica legislativa*. In AA.VV., *Fattibilità e applicabilità delle leggi*, Rimini.
- Rodotà, S. 2021. *Tecnologie e diritti*, II ed. (a cura di) G. Alpa, M.R. Marella, G. Marini, G. Resta, Bologna, il Mulino.

- Servizio studi del Senato e Ufficio rapporti con l'UE della Camera (a cura di) 2020. *Dossier per la Conferenza interparlamentare "Legiferare meglio da una prospettiva digitale"*. Roma, Senato della Repubblica.
- Shi, W., Demberg, V. 2019. *Next Sentence Prediction helps Implicit Discourse Relation Classification within and across Domains*. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, China, November 3–7, 2019. Association for Computational Linguistics: 5790–5796.
- Simoncini, A. 2019. *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*. In *BioLaw Journal – Rivista di BioDiritto*, n. 1/2019.
- Sisti, N. 2021. *Concetti, dati e analogie*, Roma, Aracne.
- Stradella, E. 2020. *AI, tecnologie innovative e produzione normativa: potenzialità e rischi*. In *DPCE online*, Vol. 44, n. 3/2020.
- Tafani, L. 2017. *La tecnica legislativa al servizio della qualità degli atti normativi*. In "Le parole giuste. Scrittura tecnica e cultura linguistica per il buon funzionamento della pubblica amministrazione e della giustizia. Atti del convegno di presentazione del progetto di ricerca e formazione, Palazzo Madama, 14 aprile 2016", Roma, Senato della Repubblica Italiana: 141-155.
- Task force sull'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale (a cura di) 2018. *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino*, Versione 1.0, Marzo.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. 2017. *Attention is all you need*. In "Advances in Neural Information Processing Systems.
- Vulić, I., Ponti, E.M., Korhonen, A., Glavaš, G. 2021. *LexFit: Lexical fine-tuning of pretrained language models*. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers): 5269-5283.
- Wallace, E., Feng, S., Kandpal, N., Gardner, M., Singh, S. 2019. *Universal Adversarial Triggers for Attacking and Analyzing NLP*. arXiv preprint arXiv:1908.07125.
- Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., Bowman, S.R. 2018. *GLUE: A multi-task benchmark and analysis platform for natural language understanding*. arXiv preprint arXiv:1804.07461.

- Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O., Bowman, S. 2019. *Superglue: A stickier benchmark for general-purpose language understanding systems*. arXiv preprint arXiv:1905.00537.
- Wei, J., Bosma, M., Zhao, V. Y., Guu, K., Yu, A. W., Lester, B., Du, N., Dai, A.M., Le, Q. V. 2021. *Finetuned language models are zero-shot learners*. arXiv preprint arXiv:2109.01652.
- Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., Zhu, J. 2019. *Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges*. In Tang, J., Kan, M.-Y., Zhao, D., Li, S., Zan, H. (Eds.), *Natural Language Processing and Chinese Computing, Lecture Notes in Computer Science*, Springer International Publishing, Cham: 563–574.
- Xue, L., Barua, A., Constant, N., Al-Rfou, R., Narang, S., Kale, M., Roberts, A., Raffel, C. 2021. *ByT5: Towards a token-free future with pre-trained byte-to-byte models*. arXiv preprint arXiv:2105.13626.
- Zeno-Zencovich, V. 2020. *Big data e epistemologia giuridica*. In Faro, S., Frosini, T.E., Peruginelli, G., (a cura di), *Dati e algoritmi. diritto e diritti nella società digitale*, Bologna, Il Mulino.
- Zhang, J., Zhao, Y., Saleh, M., Liu, P.J. 2020. *PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization*. Proceedings of the 37th International Conference on Machine Learning, Online, PMLR 119.
- Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., Sun, M. 2020. *How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence*. arXiv preprint arXiv:2004.12158.
- Zuboff, S. 2019. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*. Roma, LUISS University Press.
- Zuboff, S. 2021. "Colpo di stato digitale". In *Internazionale* n. 1404, 9 aprile 2021: 37-38.

