



DIPARTIMENTO DI INGEGNERIA INFORMATICA, MODELLISTICA,
ELETTRONICA E SISTEMISTICA

Corso di Laurea Magistrale in Ingegneria Informatica

IR-NLP & ANALISI DI SOCIAL NETWORK E MEDIA

Elaborato finale

Condizionamento di un LLM tramite rete di influenza

Professore:

Prof. **Andrea Tagarelli**

Studenti:

Prisco Ivan

matr. 252320

Presta Vincenzo

matr. 252290

Greco Matteo

matr. 252238

ANNO ACCADEMICO 2024/2025

Influence Cascade Decoding (ICD)

Elaborato finale: IR-NLP & Analisi di Social Network e Media

Ivan Prisco, Vincenzo Presta, Matteo Greco

Settembre 2025

Indice generale

1	Introduzione	3
2	Descrizione del modello e della rete di influenza	4
2.1	Costruzione del grafo di influenza	5
2.2	Schema di pesatura degli archi	5
3	Processo di diffusione dell'influenza	9
3.1	Scelta dei seed e simulazioni	9
3.2	Calcolo delle probabilità di attivazione	9
3.3	Risultato della diffusione	9
4	Condizionamento dei logit tramite boosting guidato dal grafo	10
4.1	Procedura di generazione	10
5	Analisi e test dei gradi di libertà	12
5.1	Analisi del parametro α	12
5.2	Analisi del parametro λ	16
6	Analisi sperimentale	22
6.1	Pipeline di esempio	23
6.2	Risultati delle generazioni con altri prompt	31
6.3	A/B pair-wise test: LLM as a judge	46
7	Ottimizzazioni	52
7.1	Logits Processor	52
8	Conclusioni	53
8.1	Prospettive di miglioramento	53
8.2	Possibili approfondimenti	53

A	Esempi di generazione	55
A.1	Analisi del parametro α	55
A.2	Analisi del parametro λ	55
A.3	Risultati del test LLM as a judge	55
A.4	Risultati delle generazioni con altri prompt	55
B	Dataset utilizzati	55
C	Test alternativo: Boosting sulle probabilità	56

1 Introduzione

L'obiettivo di questo progetto è lo studio, la creazione e la valutazione di una strategia di **generazione testuale guidata da una rete di influenza tra token**. In particolare, si è voluto esplorare come la struttura delle transizioni tra token, osservata in sequenze generate da un modello linguistico, possa essere sfruttata per modificare la distribuzione di probabilità del modello stesso, introducendo un meccanismo di *boosting* con lo scopo di migliorare la generazione finale.

La metodologia proposta integra due componenti principali:

- un **modello linguistico neurale** pre-addestrato, utilizzato per la generazione di sequenze e per la stima delle distribuzioni di probabilità dei token successivi;
- una **rete di influenza**, costruita a partire da simulazioni di generazione, in cui i nodi corrispondono ai token e gli archi rappresentano transizioni osservate.

Il modello adottato come base per la generazione è un **trasformatore autoregressivo (Mistral-7B-Instruct-v0.3)**, al quale viene applicato un meccanismo di modifica dei logit durante il decoding. In questo modo, la distribuzione di uscita del modello non riflette più unicamente le probabilità apprese in fase di addestramento, ma viene adattata dinamicamente in funzione delle connessioni presenti nella rete di influenza. L'analisi condotta si propone di valutare l'impatto di tale meccanismo in termini di struttura della rete, distribuzione dei pesi, processo di diffusione e misure di qualità delle generazioni.

2 Descrizione del modello e della rete di influenza

La costruzione della rete di influenza parte dall'esecuzione di un insieme di **simulazioni di generazione** condotte con l'LLM scelto. Ogni simulazione consiste nella produzione di una sequenza di lunghezza fissata, ottenuta a partire da un prompt iniziale. Questo permette di osservare le transizioni effettive tra token così come emergono dal processo autoregressivo di generazione. Sono state eseguite **150 simulazioni**, ciascuna della lunghezza di **200 token**, utilizzando diversi prompt iniziali. Il prompt principale, su cui sono stati condotti quasi tutti i test, è stato:

“The future of AI is”

Al fine di esplorare scenari di generazione differenti, i token sono stati campionati tramite tre possibili strategie, selezionate in modo casuale a ogni simulazione (con valori di temperatura variabili scelti casualmente nel caso della seconda strategia):

- **Nucleus Sampling (top-p)**: seleziona il sottoinsieme minimo di token la cui probabilità cumulata supera una soglia p , da cui viene poi estratto il token successivo.
- **Temperature Sampling**: applica un fattore T alla distribuzione per controllare la variabilità; con $T < 1$ si privilegiano i token più probabili, con $T > 1$ la distribuzione diventa più piatta e favorisce l'esplorazione, mentre $T = 0$ corrisponde al greedy decoding.
- **Top-k Sampling**: limita la scelta ai k token con probabilità più alta e campiona il successivo solo da questo insieme ristretto.

Per memorizzare in modo strutturato le informazioni derivanti dalle simulazioni, si utilizza una **matrice** denominata `all_tokens`. Tale matrice è organizzata secondo il seguente schema:

- ogni **riga** corrisponde a una simulazione distinta;
- lungo le **colonne** sono disposti i token generati in ordine sequenziale;
- ogni cella della matrice contiene una **tupla** della forma $(\text{id-token}, p)$, dove:
 - id-token è l'identificativo numerico del token secondo il vocabolario del modello;
 - p rappresenta la probabilità di generazione attribuita dal modello a quel token in quel preciso passo.

Questa rappresentazione consente di conservare non solo l'ordine di apparizione dei token, ma anche le relative probabilità, fornendo così le basi per la costruzione della rete di influenza. A partire dalla matrice `all_tokens` è infatti possibile estrarre le transizioni osservate tra token e calcolare statistiche necessarie per definire i pesi degli archi della rete.

2.1 Costruzione del grafo di influenza

Una volta ottenuta la matrice `all_tokens`, il passo successivo consiste nella costruzione di un **grafo diretto** che rappresenta le transizioni tra token osservate nelle simulazioni.

Il grafo viene inizializzato come **grafo orientato**, in quanto la direzionalità delle transizioni è un aspetto cruciale: il fatto che un token u preceda un token v non implica necessariamente il contrario. Come nodo iniziale viene aggiunto l'ultimo token del prompt, che rappresenta la radice di tutte le sequenze generate.

Per ciascuna simulazione salvata (riga) in `all_tokens`, la funzione scorre i token nell'ordine di generazione e costruisce archi orientati tra token consecutivi:

- se un token t_1 precede immediatamente un token t_2 in una simulazione, si aggiunge un arco diretto (t_1, t_2) al grafo;
- se i token non erano ancora presenti, vengono aggiunti come nodi separati;
- se l'arco (t_1, t_2) era già stato osservato, esso non viene duplicato, garantendo l'unicità della struttura.

Il risultato finale è un grafo diretto $G = (V, E)$, in cui:

- i nodi V corrispondono ai **token** apparsi nelle simulazioni;
- gli archi E rappresentano **transizioni** empiricamente osservate;

In questa fase, il grafo costruito è puramente strutturale: ogni arco è presente o assente in base alle transizioni osservate, senza alcuna informazione quantitativa. Nella fase successiva verranno introdotti i pesi associati agli archi, al fine di riflettere la forza o la rilevanza delle transizioni.

2.2 Schema di pesatura degli archi

Per rendere il grafo informativo e adatto a modellare processi di diffusione, è stato introdotto uno **schema di pesatura** che combina due componenti distinte: la stima empirica delle transizioni e la correlazione statistica tra token.

Stima condizionata

La prima componente è la **probabilità condizionata stimata** di osservare un token v successivamente a un token u , definita come:

$$\hat{p}(v \mid u) = \frac{\text{freq}(u \rightarrow v)}{\sum_{v'} \text{freq}(u \rightarrow v')}$$

dove $\text{freq}(u \rightarrow v)$ indica il numero di volte in cui la transizione (u, v) è stata osservata nelle simulazioni. Questa misura cattura l'aspetto puramente empirico della frequenza delle transizioni.

PPMI

La seconda componente è la **Positive Pointwise Mutual Information** (PPMI), che valuta la forza statistica dell'associazione tra due token al di là della loro frequenza marginale. È definita come:

$$\text{PPMI}(u, v) = \max \left(0, \log \frac{p(u, v)}{p(u)p(v)} \right)$$

dove $p(u, v)$ è la probabilità congiunta della coppia (u, v) , mentre $p(u)$ e $p(v)$ sono le probabilità marginali dei due token. Per rendere i punteggi confrontabili con le probabilità condizionate, i valori di PPMI vengono normalizzati nell'intervallo $[0, 1]$, dividendo per il massimo valore di PMI.

Formula finale del peso

Il peso associato a ciascun arco (u, v) è ottenuto come combinazione convessa delle due componenti:

$$w(u, v) = \alpha \cdot \hat{p}(v|u) + (1 - \alpha) \cdot \text{PPMI}(u, v)$$

dove il parametro $\alpha \in [0, 1]$ regola il bilanciamento tra frequenza empirica e correlazione semantica.

Questo approccio presenta diversi vantaggi:

- riduce il numero di archi con peso nullo, grazie alla combinazione di due misure diverse;
- tiene conto sia della frequenza empirica delle transizioni osservate, sia della forza statistica della loro associazione;
- fornisce una base solida e bilanciata per modellare i processi di diffusione dell'influenza nella rete.

Scelta del parametro α

La valutazione del parametro α e la conseguente scelta del suo valore sono discusse in dettaglio nella Sezione 5, dove vengono presentati i test relativi ai parametri del modello.

In questo contesto, si è scelto di fissare il valore $\alpha = 0.4$ per la costruzione definitiva della rete. Tale valore rappresenta un compromesso equilibrato: da un lato evita una rete eccessivamente polarizzata, dominata da archi molto forti accanto a numerosi archi debolissimi; dall'altro non conduce a una struttura troppo sparsa o vuota. La distribuzione risultante appare dunque sufficientemente centrale, mantenendo un buon numero di archi attivi e fornendo una base solida per la fase di diffusione.

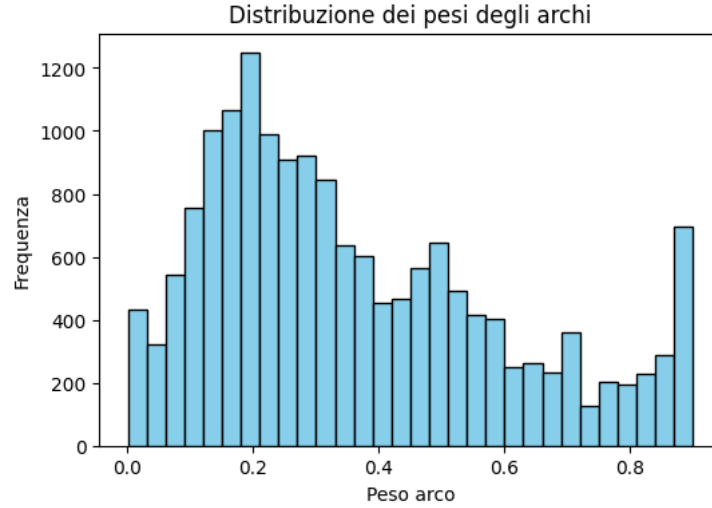


Figure 1: Distribuzione dei pesi sugli archi della rete ($\alpha = 0.4$)

Esempio di costruzione e pesatura di un arco

Per chiarire il funzionamento dello schema proposto, viene riportato un esempio semplificato. Si supponga di avere **due simulazioni** generate dal modello, a partire dal prompt “*The future of AI is*”.

- Simulazione 1: *innovative* → *and* → *powerful*
- Simulazione 2: *innovative* → *technology* → ,

Nella matrice `all_tokens` si avrà ad esempio:

Simulazione	Token 1	Token 2	Token 3
1	(101(<i>innovative</i>), 0.20)	(102(<i>and</i>), 0.15)	(103(<i>powerful</i>), 0.10)
2	(101(<i>innovative</i>), 0.18)	(104(<i>technology</i>), 0.12)	(105(,), 0.30)

Costruzione del grafo

Dalla matrice si ricavano i nodi e gli archi osservati:

$$V = \{\text{innovative}, \text{and}, \text{powerful}, \text{technology}, ,\}$$

$$E = \{(\text{innovative}, \text{and}), (\text{and}, \text{powerful}), (\text{innovative}, \text{technology}), (\text{technology}, ,)\}$$

Calcolo dei pesi

Si consideri l'arco (**innovative**, **and**). La frequenza osservata è:

$$\text{freq}(\text{innovative} \rightarrow \text{and}) = 1$$

$$\text{freq}(\text{innovative} \rightarrow \cdot) = 2 \quad \Rightarrow \quad \hat{p}(\text{and} \mid \text{innovative}) = \frac{1}{2} = 0.5$$

Per il calcolo della PPMI, si supponga che:

$$p(\text{innovative}, \text{and}) = 0.05, \quad p(\text{innovative}) = 0.10, \quad p(\text{and}) = 0.08$$

$$\text{PPMI}(\text{innovative}, \text{and}) = \max \left(0, \log \frac{0.05}{0.10 \cdot 0.08} \right) = \max(0, \log 6.25) \approx 1.83$$

Normalizzando nell'intervallo $[0, 1]$, si ottiene ad esempio:

$$\text{PPMI}_{\text{norm}}(\text{innovative}, \text{and}) = 0.75$$

Infine, il peso dell'arco con $\alpha = 0.4$ è:

$$w(\text{innovative}, \text{and}) = 0.4 \cdot 0.5 + (1 - 0.4) \cdot 0.75 = 0.2 + 0.45 = 0.65$$

Risultato

L'arco (**innovative**, **and**) riceve quindi un peso finale di 0.65, che combina sia la frequenza empirica della transizione sia la forza semantica della loro associazione.

3 Processo di diffusione dell’influenza

Dopo aver costruito il grafo e assegnato i pesi agli archi, la fase successiva consiste nell’analisi della propagazione dell’influenza tra i token. Per questo scopo è stato adottato il modello di diffusione **Independent Cascade (IC)**, particolarmente adatto a contesti in cui le attivazioni avvengono in modo probabilistico lungo gli archi pesati della rete.

3.1 Scelta dei seed e simulazioni

Il modello IC prevede che, partendo da un insieme iniziale di nodi attivi (*seeds*), ciascun nodo tenti di attivare i propri vicini con una probabilità pari al peso dell’arco. In caso di successo, i nuovi nodi attivati partecipano a loro volta al processo, fino a quando non emergono più nuove attivazioni.

I **seed** sono stati definiti come i token contenuti nel prompt di partenza (*“The future of AI is”*). Essi rappresentano le sorgenti di influenza da cui prende avvio la diffusione nella rete.

Per ottenere stime robuste delle probabilità di attivazione, il processo è stato ripetuto per **1000 simulazioni indipendenti**. Questa scelta consente di attenuare l’effetto delle attivazioni puramente casuali o “sfortunate” che possono verificarsi nelle singole run, garantendo invece una stima più stabile e affidabile del comportamento complessivo della rete.

3.2 Calcolo delle probabilità di attivazione

Al termine delle simulazioni, per ciascun nodo v viene calcolata la **probabilità empirica di attivazione**, definita come:

$$p_{\text{att}}(v) = \frac{\# \text{ volte in cui } v \text{ attivo}}{\text{numero totale di simulazioni}}$$

dove il numeratore conta quante volte il nodo è stato raggiunto e attivato nel processo, e il denominatore corrisponde al numero totale di run.

Il risultato è una distribuzione di punteggi di influenza che quantifica, per ogni token, la probabilità di essere coinvolto nella diffusione a partire dai token iniziali.

3.3 Risultato della diffusione

L’algoritmo fornisce un ranking dei nodi in base alla loro probabilità di attivazione, permettendo di identificare i token maggiormente influenzati dal processo di diffusione. Questo ranking è uno strumento fondamentale per comprendere quali parti del vocabolario risultano più centrali rispetto al topic indotto dal prompt e come l’influenza si propaga effettivamente nella rete.

4 Condizionamento dei logit tramite boosting guidato dal grafo

Costruita la rete di influenza, definiti i pesi degli archi ed effettuato il processo di diffusione, il passo successivo consiste nell'integrare queste informazioni all'interno del processo di generazione del modello linguistico.

L'idea è di condizionare i *logit* prodotti dal modello in fase di decoding, in modo che la distribuzione finale dei token candidati rifletta non solo la conoscenza statistica del modello ma anche la struttura del grafo e eventuali bias esterni.

4.1 Procedura di generazione

Il processo di generazione può essere descritto nei seguenti passi:

1. **Codifica del prompt.** Il testo iniziale (*prompt*) viene tokenizzato e fornito al modello per avviare la generazione.
2. **Calcolo dei logit.** Ad ogni passo di decoding, il modello produce un vettore di logit, cioè valori non normalizzati che rappresentano la propensione del modello a generare ciascun token del vocabolario.
3. **Selezione dei candidati.** Per rendere il calcolo più efficiente e mirato, vengono considerati i **100 token più probabili** secondo il modello.
4. **Applicazione del boosting.** Per ciascun candidato v vengono calcolati due contributi:
 - **Peso dal grafo** $w(u, v)$: se esiste un arco dal token corrente u verso v , si utilizza il peso calcolato in fase di costruzione della rete.
 - **Bias esterno** $bias(v)$: punteggio assegnato al token v dal processo di diffusione (*Independent Cascade*), che misura la probabilità empirica di attivazione del token a partire dai seed del prompt. Questo valore permette di enfatizzare i token che risultano più centrali nella propagazione dell'influenza nella rete.

Questi due contributi vengono combinati e scalati dal parametro λ , ottenendo così la **formula finale**:

$$\text{logit_final}(v) = \text{logit}(v) + \lambda \cdot (w(u, v) + \text{bias}(v)).$$

5. **Campionamento.** Dopo la modifica dei logit, si applica una *softmax* per ottenere la nuova distribuzione di probabilità e si campiona il prossimo token tra i **50 candidati più probabili** (strategia top-k).
6. **Iterazioni.** Il processo si ripete per il numero di passi prestabilito, generando così una sequenza in cui ogni scelta è influenzata sia dal modello che dalla rete di influenza.

Scelta del parametro λ

Il parametro λ controlla l'intensità con cui il contributo della rete e dei punteggi di diffusione viene integrato nei logit del modello. La valutazione del parametro λ e la conseguente scelta del suo valore sono discusse in dettaglio nella sezione successiva, in particolare nella Sezione 5.2.

Interpretazione della formula

Le quantità $w(u, v)$ e $\text{bias}(v)$ vengono utilizzate come correzioni additive ai logit; queste infatti agiscono come **segnali di influenza esterna** che vengono iniettati nel processo di decoding, alterando la distribuzione finale senza rompere la coerenza probabilistica del modello.

In questo contesto, è importante notare che le quantità $w(u, v)$ e $\text{bias}(v)$ assumono valori nell'intervallo $[0, 1]$, mentre $\text{logit}(v)$ rappresenta un punteggio non vincolato. Il parametro λ svolge quindi una **duplice funzione**: da un lato regola il contributo della rete di influenza nel processo di generazione, dall'altro consente di riequilibrare quantità definite su scale eterogenee. In questo modo i diversi termini possono essere combinati in maniera coerente all'interno della stessa espressione.

5 Analisi e test dei gradi di libertà

In questa sezione si analizzano nel dettaglio i parametri principali utilizzati nel lavoro, in particolare: α che regola lo schema di pesatura degli archi nella rete di influenza, e λ che regola il “peso” della rete di influenza nella generazione.

5.1 Analisi del parametro α

Una volta definita la funzione di pesatura, come spiegato in 2.2, è importante analizzare come il parametro α influisca sul comportamento della rete e, in particolare, sul processo di diffusione dell’influenza. La valutazione ha seguito due direttrici principali:

- l’effetto sulla **struttura della rete**, osservando come cambiano la distribuzione dei pesi, i nodi centrali e le componenti connesse al variare di α ;
- l’effetto sul **processo di diffusione**, misurato in termini di copertura media dei nodi, entropia della distribuzione di influenza e variazione dei top-nodi influenti.

Si passa quindi a evidenziare i risultati di questa analisi, mettendo in evidenza i comportamenti emergenti per valori estremi o intermedi di α .

Comportamento atteso per valori estremi di α

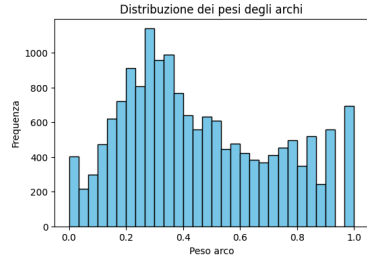
L’analisi teorica permette di anticipare alcuni comportamenti caratteristici della rete in corrispondenza di valori estremi del parametro α :

- **Caso $\alpha \rightarrow 1$:** i pesi degli archi sono dominati dalla frequenza empirica delle transizioni. In questo scenario tendono a prevalere archi con valori molto sbilanciati: da un lato, archi con peso molto elevato corrispondenti a transizioni frequenti (tipicamente legate a stopword o punteggiatura); dall’altro, una grande quantità di archi con peso molto basso dovuti a transizioni rare ma comunque presenti. La distribuzione dei pesi risulta quindi più ampia e dispersiva.
- **Caso $\alpha \rightarrow 0$:** i pesi degli archi sono dominati dalla PPMI, che valorizza co-occorrenze forti e specifiche. In questo caso la maggior parte degli archi deboli tende a scomparire (PPMI nulla o negativa), mentre sopravvivono soltanto gli archi che rappresentano associazioni semantiche particolarmente significative. La distribuzione dei pesi è quindi più concentrata e selettiva, con pochi archi mediamente più bilanciati ma privi degli estremi molto grandi tipici delle frequenze elevate.

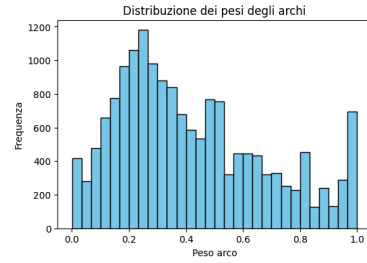
Quindi, per α elevati ci si aspetta una rete più densa e sbilanciata, con la presenza simultanea di archi molto piccoli e molto grandi, mentre per α bassi ci si aspetta invece una rete con pesi distribuiti in modo più omogeneo tra i legami che sopravvivono.

Distribuzione dei pesi per diversi valori di α

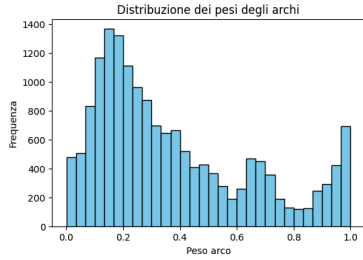
Nella figura seguente sono riportati gli istogrammi della distribuzione dei pesi degli archi per i diversi valori di α considerati. Questo confronto permette di osservare chiaramente come la scelta del parametro influenzi la struttura della rete.



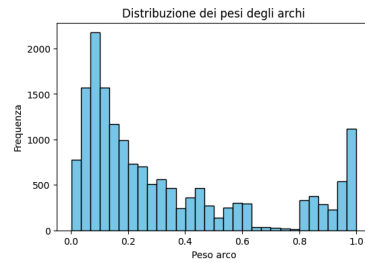
$\alpha = 0.0$



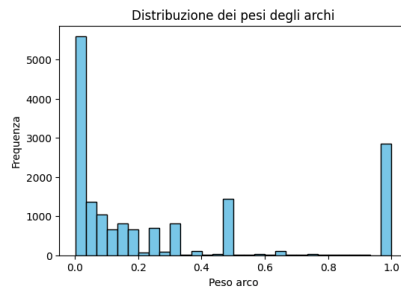
$\alpha = 0.25$



$\alpha = 0.5$



$\alpha = 0.75$



$\alpha = 1.0$

Figure 2: Distribuzione dei pesi degli archi per diversi valori di α .

Osservando le distribuzioni emerge un comportamento progressivo al variare

del parametro α :

- $\alpha = 0.0$: la distribuzione è relativamente concentrata su valori medi, poiché i pesi riflettono quasi esclusivamente la PPMI.
- $\alpha = 0.25$ e $\alpha = 0.5$: la distribuzione si allarga e mostra una maggiore variabilità. L'effetto combinato di frequenza e PPMI genera una gamma di pesi più diversificata, con valori sia bassi sia alti.
- $\alpha = 0.75$: la distribuzione si sposta ulteriormente verso gli estremi, con molti archi di peso basso e alcuni di peso molto alto. Ciò riflette la crescente influenza delle frequenze empiriche.
- $\alpha = 1.0$: la distribuzione è fortemente polarizzata. La maggioranza degli archi ha peso vicino a zero, ma compaiono anche picchi molto elevati, corrispondenti a transizioni frequenti (es. stopword e punteggiatura). Questo porta ad una rete molto densa ma sbilanciata.

Al crescere di α si osserva quindi una transizione da una distribuzione più compatta e semantica ad una distribuzione fortemente sparsa, dove coesistono archi molto deboli e archi molto forti. Questo andamento riflette esattamente il compromesso tra *selettività semantica* e *robustezza statistica* discusso nella sezione precedente.

Statistiche dipendenti da α

Oltre alle caratteristiche strutturali del grafo, sono state calcolate una serie di statistiche che dipendono direttamente dai **pesi degli archi** $w(u, v)$. Queste misure forniscono informazioni sulla distribuzione della forza delle connessioni e sul ruolo medio dei nodi nella rete di transizione.

Table 1: Statistiche pesate del grafo

Statistica	Formula	Descrizione
Average in-strength	$\overline{s^{in}} = \frac{1}{ V } \sum_{u \in V} \sum_{(v,u) \in E} w(v, u)$	Peso medio entrante per nodo.
Average out-strength	$\overline{s^{out}} = \frac{1}{ V } \sum_{u \in V} \sum_{(u,v) \in E} w(u, v)$	Peso medio uscente per nodo.
Peso medio	$\overline{w} = \frac{1}{ E } \sum_{(u,v) \in E} w(u, v)$	Media dei pesi sugli archi.
Varianza pesi	$\text{Var}(w) = \frac{1}{ E } \sum_{(u,v) \in E} (w(u, v) - \overline{w})^2$	Dispersione dei pesi rispetto alla media.
Peso max/min	$w_{\max} = \max_{(u,v) \in E} w(u, v), \quad w_{\min} = \min_{(u,v) \in E} w(u, v)$	Estremi della distribuzione.
Mediana pesi	$w_{\text{med}} = \text{valore centrale}$	Il 50% degli archi ha peso $\leq w_{\text{med}}$.
Quartili	$Q_1 = 25^\circ \text{ perc.}, \quad Q_3 = 75^\circ \text{ perc.}$	Intervallo interquartile dei pesi.

Table 2: Confronto dei valori al variare di α

α	$\overline{s^{in}}$	$\overline{s^{out}}$	\overline{w}	$\text{Var}(w)$	w_{\max}	w_{\min}	w_{med}	$[Q_1, Q_3]$
0	1.555	1.555	0.459	0.070	1.0	0.0000	0.399	[0.257, 0.664]
0.25	1.415	1.415	0.418	0.067	1.0	0.00018	0.356	[0.221, 0.583]
0.5	1.276	1.276	0.377	0.077	1.0	0.00035	0.286	[0.164, 0.546]
0.75	1.136	1.136	0.335	0.098	1.0	0.00053	0.202	[0.094, 0.476]
1	0.996	0.996	0.294	0.131	1.0	0.00071	0.111	[0.014, 0.500]

Dalla tabella si osserva un andamento regolare dei valori al variare del parametro α . In particolare, sia l'**average in-strength** che l'**average out-strength** diminuiscono progressivamente da 1.555 a 0.996: ciò indica che, aumentando il peso della componente frequenziale ($\alpha \rightarrow 1$), la rete tende a distribuire meno peso medio per nodo, riducendo la forza complessiva delle connessioni entranti e uscenti. Lo stesso trend si riflette nel **peso medio** \overline{w} , che passa da 0.459 a 0.294.

Parallelamente, la **varianza dei pesi** cresce da 0.070 a 0.131, mostrando che la distribuzione diventa più polarizzata: alti valori di α producono molti archi molto deboli e pochi archi molto forti, mentre a bassi valori di α la distribuzione è più equilibrata. Questo effetto si nota anche dalla **mediana** dei pesi, che scende da 0.399 a 0.111, e dai **quartili**, che si spostano progressivamente verso valori bassi.

I valori estremi ($w_{\max} = 1.0$, $w_{\min} \approx 0$) rimangono costanti, confermando che il supporto della distribuzione non cambia; ciò che varia è la sua forma interna.

Analisi della diffusione con Independent Cascade

Dopo aver caratterizzato la rete dal punto di vista delle statistiche pesate, è stato analizzato il comportamento del processo di diffusione *Independent Cascade* al variare del parametro α . Poiché i pesi $w(u, v)$ degli archi rappresentano le probabilità di attivazione, la scelta di α influenza direttamente la dinamica della propagazione.

Table 3: Statistiche della diffusione con Independent Cascade

Statistica	Formula	Descrizione
Coverage	$\text{Cov} = \frac{ \mathcal{A} }{ V }$	Proporzione di nodi attivati \mathcal{A} rispetto al totale dei nodi.
Entropy	$H = - \sum_{u \in V} p_{\text{att}}(u) \log p_{\text{att}}(u)$	Misura della diversità della distribuzione di attivazione: alta se la diffusione è uniforme, bassa se pochi nodi concentrano l'influenza.

Table 4: Valori di coverage ed entropy della diffusione IC al variare di α

α	Coverage	Entropy
0	0.706	1040.34
0.25	0.572	1361.76
0.5	0.410	1547.76
0.75	0.214	1366.20
1	0.007	89.49

Dalla tabella emergono due andamenti distinti per le statistiche di diffusione al variare di α . La **coverage** mostra una riduzione progressiva, passando da circa 0.70 con $\alpha = 0$ a valori quasi nulli (0.007) per $\alpha = 1$. Questo comportamento indica che, quando la pesatura privilegia la componente semantica (PPMI), la propagazione riesce a raggiungere una porzione ampia della rete; al contrario, quando prevale la componente frequenziale, la diffusione rimane confinata a pochi nodi hub, riducendo drasticamente la copertura.

L'**entropia** segue un andamento non monotono: cresce inizialmente, raggiungendo un massimo per $\alpha = 0.5$ (circa 1548), a indicare che la diffusione è in quel caso più equilibrata e distribuita su un ampio insieme di nodi. Successivamente, l'entropia decresce, fino a valori molto bassi per $\alpha = 1$ (~ 89), riflettendo una concentrazione estrema della diffusione su pochi nodi dominanti.

Quindi, valori bassi di α favoriscono una **diffusione ampia** ma meno diversificata, valori intermedi garantiscono una **diffusione bilanciata** e distribuita, mentre valori alti portano a una **diffusione ridotta e polarizzata**, dominata da stopword e punteggiatura.

Tutti i risultati ottenuti dal test del parametro α sono stati esportati in formato JSON (vedi Appendice A.1)

5.2 Analisi del parametro λ

A questo punto, si passa all'analisi sperimentale relativa all'impatto del parametro λ sul processo di generazione. L'obiettivo è valutare come la variazione di λ influenzi tre aspetti principali:

- la **frequenza dei token "influenti"** selezionati,
- la **qualità statistica** della generazione (misurata tramite Perplexity),
- la **coerenza semantica** delle risposte (misurata tramite similarità SBERT).

Come discusso nella Sezione 4.1, λ rappresenta il peso con cui il bias di influenza viene integrato nei logit del modello durante la generazione.

Setup sperimentale

Sono stati dunque condotti una serie di test al fine di valutare l’impatto del parametro λ sui testi generati. L’esperimento è stato strutturato come segue:

- **Prompt di partenza:** per tutti i test è stato utilizzato lo stesso prompt iniziale (“*The future of AI is*”), in modo da rendere confrontabili le diverse generazioni.
- **Valori di λ :** sono stati testati i valori:

$$\lambda \in \{0.0, 1.0, 1.5, 2.0, 2.5, 5.0, 10.0, 100.0\}$$

- **Numero di generazioni:** per ciascun valore di λ sono state effettuate 30 generazioni indipendenti, ognuna fino a un massimo di 200 passi (token).
- **Statistiche di valutazione:**
 - **Frequenza dei token influenti:** definita come il rapporto tra il numero di token appartenenti alla lista dei più influenti (top-50) e il numero totale di token nella sequenza generata. Misura quanto spesso la generazione viene guidata verso token con alto peso di influenza.
 - **Perplexity (PPL):** misura la plausibilità di una sequenza secondo una distribuzione di probabilità. Data una sequenza $x = (x_1, \dots, x_N)$, la Perplexity è definita come

$$\text{PPL}(x) = \exp \left(-\frac{1}{N} \sum_{t=1}^N \log P(x_t \mid x_{<t}) \right)$$

dove $P(x_t \mid x_{<t})$ rappresenta la probabilità del token x_t dato il contesto precedente.

- **Similarità SBERT:** misura la coerenza semantica tra prompt e generazione. È calcolata come similarità coseno tra gli embedding del prompt e quelli della sequenza generata, ottenuti tramite il modello all-MiniLM-L6-v2.
- **Output:** i risultati di ciascuna generazione sono stati salvati in file JSON, uno per ogni valore di λ (vedi Appendice A.2).

Questa fase di sperimentazione ha consentito di costruire un dataset comparativo, che costituisce la base per l’analisi dei risultati.

Risultati

Per ciascun valore sono state calcolate le medie delle tre statistiche di valutazione: frequenza dei token influenti, Perplexity e similarità SBERT.

Table 5: Statistiche medie al variare del parametro λ

λ	Freq. token influenti	Perplexity	Similarità SBERT
0.0	0.0858	5.760	0.626
1.0	0.1081	4.467	0.641
1.5	0.1163	4.230	0.624
2.0	0.1144	4.369	0.628
2.5	0.1217	4.083	0.644
5.0	0.1545	3.087	0.646
10.0	0.2156	2.608	0.674
100.0	0.3214	1.230	0.598

A questo punto, si mostra l'andamento dei valori al variare di λ .

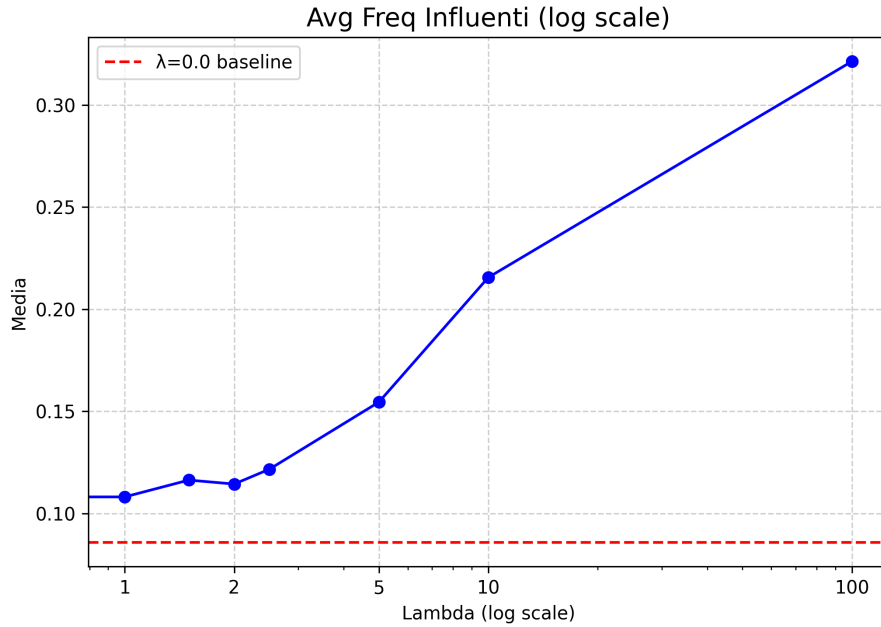


Figure 3: Frequenza media dei token influenti

La **frequenza dei token influenti** cresce in modo quasi monotono all'aumentare di λ . A valori elevati (ad esempio $\lambda = 100$) la frequenza raggiunge circa 0.32, cioè quasi un terzo dei token generati appartiene alla lista dei più influenti. Questo indica che un λ elevato forza notevolmente il modello verso la rete di influenza.

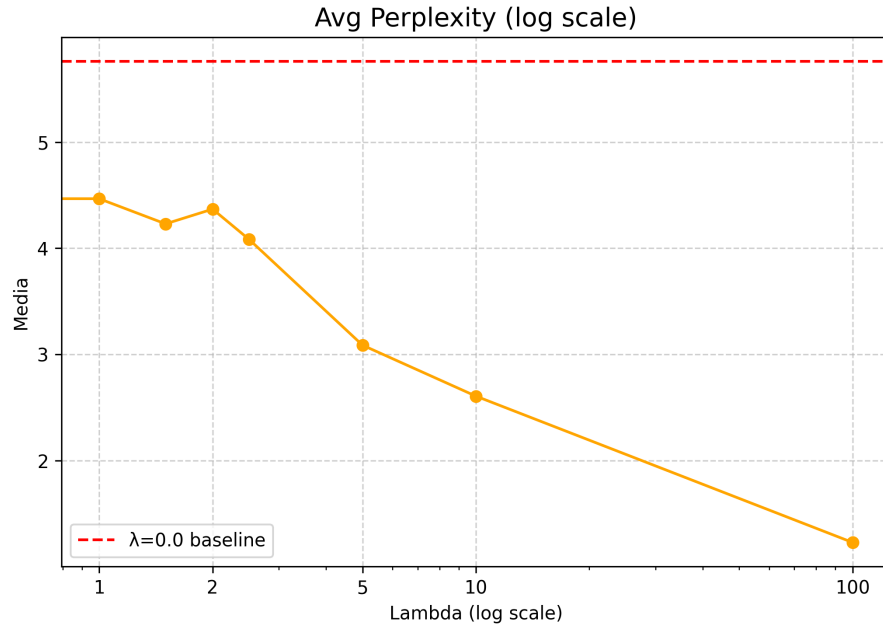


Figure 4: Perplexity

La **Perplexity** decresce al crescere di λ , passando da circa 5.76 (baseline, $\lambda = 0$) fino a 1.23 a $\lambda = 100$. I testi diventano quindi più probabili secondo il modello, ma anche meno vari. Questo andamento suggerisce che l'aumento di λ riduce la diversità linguistica, portando a sequenze più regolari e, con un lambda eccessivamente alto, a ripetizioni.

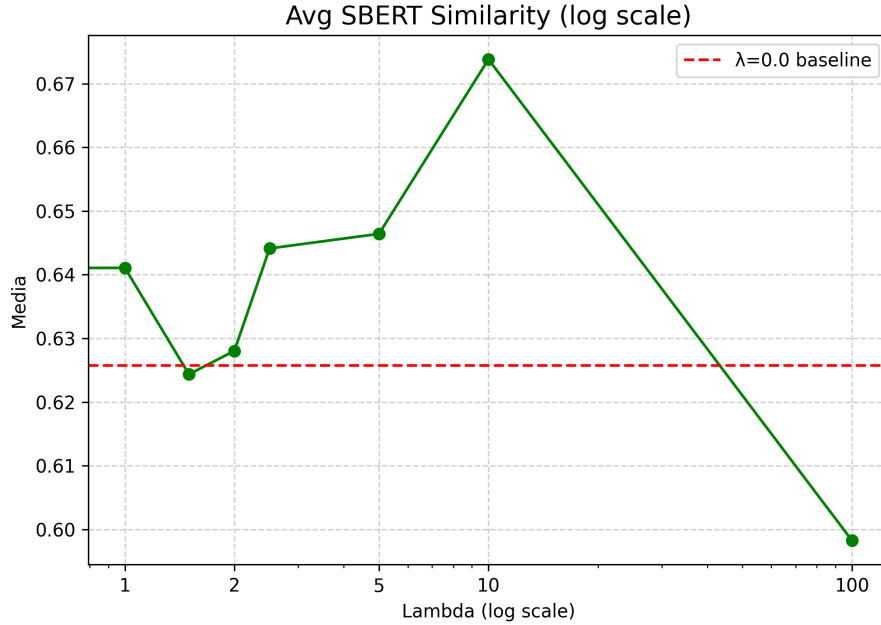


Figure 5: SBERT Similarity

La **Similarità SBERT** mostra un andamento non monotono. Cresce inizialmente e raggiunge il massimo a $\lambda = 10$ (0.674, superiore alla baseline 0.626), per poi calare drasticamente a $\lambda = 100$ (0.598, sotto la baseline). Ciò indica che valori intermedi di λ migliorano la coerenza semantica con il prompt, mentre valori troppo alti potrebbero penalizzare la qualità semantica della generazione.

Considerazioni finali circa il valore di λ

Dall’analisi delle statistiche emerge che valori bassi di λ (1–2.5) producono solo un miglioramento contenuto, mentre valori intermedi (5–10) sembrano offrire il compromesso migliore tra qualità linguistica e coerenza semantica, con la similarità SBERT che raggiunge i valori massimi. Tuttavia, nonostante le statistiche risultino particolarmente favorevoli intorno a $\lambda = 10$, un’analisi qualitativa delle generazioni evidenzia una forte tendenza alla ripetitività. Per questo motivo, nella pratica, il range $\lambda \in [1.5, 2.5]$ si dimostra più adatto a mantenere un equilibrio tra coerenza semantica e varietà linguistica.

Perplexity. È inoltre rilevante notare che, indipendentemente dal valore scelto per λ , la Perplexity si mantiene sempre inferiore rispetto alla baseline del modello ($\lambda = 0$). Ciò suggerisce che il boosting renda sistematicamente le sequenze più “probabili” per il modello, pur non garantendo sempre una migliore qualità testuale percepita.

Frequenza dei token influenti e similarità SBERT. Per quanto riguarda le altre due statistiche (frequenza dei token influenti e similarità SBERT), anch'esse risultano generalmente migliori rispetto alla baseline, il che tuttavia è un comportamento atteso: la rete di influenza infatti "guida" la generazione proprio in direzione di tali miglioramenti.

6 Analisi sperimentale

Per valutare l’efficacia del modello proposto e comprendere a fondo l’impatto dei parametri introdotti, sono state condotte diverse sperimentazioni. L’obiettivo principale è stato quello di misurare come la rete di influenza, costruita a partire dalle simulazioni e integrata nel processo di generazione tramite boosting dei logit, modifichi il comportamento del modello linguistico rispetto alla baseline standard.

Per sperimentare in modo più approfondito il comportamento del modello, sono state costruite **diverse reti di influenza**, ciascuna associata a un **topic** distinto e generata a partire da un prompt specifico (riportati nella Tabella 6). Questa scelta permette di analizzare come il meccanismo di pesatura e di diffusione si adatti a domini semantici molto differenti tra loro: dalla matematica alla cultura, dalle scienze naturali a scenari narrativi e creativi. In questo modo, è stato possibile valutare la robustezza del modello non solo in un singolo contesto (ad esempio AI), ma anche su temi eterogenei che richiedono strutture linguistiche e associazioni concettuali differenti.

Table 6: Topic e prompt utilizzati per la costruzione delle reti di influenza

Topic	Prompt
AI	<i>“The future of AI is”</i>
Hallucination	<i>“John felt that there was an anaconda snake in his room... But when he turned on the lights there was no snake. It was only what?”</i>
Environment	<i>“The climate change effects are”</i>
Biology	<i>“The role of cells in the human body is”</i>
Story	<i>“Imagine a world where people communicate only through colors—tell a story about a child who invents a new color.”</i>
Culture	<i>“When learning about the world and different cultures, what is important if you are committed to eliminating preconceived notions?”</i>
Economy	<i>“The importance of money is”</i>
Government	<i>“The role of government is ‘x’</i>
Math	<i>“The solution to the equation $3x + 7 = 16$ is”</i>
Social_Media	<i>“The impact of social media is”</i>
Statistic	<i>“What does it mean when data shows a normal distribution?”</i>
Time_Travel	<i>“In the future, humans will travel through time by using”</i>

Nelle sezioni successive si mostrerà una **pipeline di esempio** costruita sul topic **AI**, utilizzando il prompt *“The future of AI is”*. Verranno illustrate tutte le

fasi del processo: dalla **costruzione della rete di influenza**, alla **diffusione tramite Independent Cascade**, fino alla **generazione testuale** con rappresentazione della variazione dei logit indotta dal meccanismo di boosting. Questo esempio fungerà da caso guida per comprendere nel dettaglio il funzionamento del modello proposto.

6.1 Pipeline di esempio

Step 1: Generazione delle simulazioni e costruzione della matrice `all_tokens`

Il primo passo consiste nel lanciare **150 simulazioni di generazione**, ciascuna della lunghezza di **200 token**, a partire dal prompt relativo al topic AI:

“The future of AI is”

I token prodotti vengono memorizzati nella matrice `all_tokens`, la cui struttura è stata discussa e formalizzata nella Sezione 2.

Step 2: Costruzione e pesatura del grafo

A partire dalla matrice `all_tokens` viene costruito un **grafo diretto** che rappresenta le transizioni tra token osservate nelle simulazioni. Come discusso in 2.1, i nodi corrispondono ai token, mentre gli archi rappresentano le transizioni empiriche ($u \rightarrow v$).

Si estrae così un grafo con

- **Numero nodi**=4888
- **Numero archi**=16553
- **Grado medio**=3.39
- **Grado massimo uscente**=429
- **Grado massimo entrante**=755
- **Grado massimo totale**=1153

Gli archi vengono poi **pesati** secondo lo schema introdotto nella Sezione 5.1, dedicata alla pesatura. Il risultato è una rete di influenza in cui ogni arco possiede un peso che riflette sia la frequenza di transizione osservata, sia la forza semantica dell’associazione.

Nella figura seguente è riportata un’illustrazione delle ego network del token *“future”* e *“AI”* contenute nel prompt

Step 3: Definizione dei token seed e diffusione IC

24

of AI is". I token seed estratti dal prompt sono i seguenti:

{1183 : The, 4205 : future, 1070 : of, 16875 : AI, 1117 : is}

Da questi nodi iniziali, è stato eseguito un processo di diffusione basato sul modello **Independent Cascade (IC)**, ripetuto per **1000 simulazioni** indipendenti. Questo numero di run consente di attenuare l'effetto delle attivazioni puramente casuali e di ottenere una stima più robusta della probabilità di attivazione dei nodi nella rete.

Il risultato del processo è una distribuzione di **probabilità di attivazione** assegnata a ciascun token, come definito in Sezione 3.2

Per semplicità, vengono riportati di seguito soltanto i **100 nodi più influenzati**, ordinati in base alla loro probabilità di attivazione. Come atteso, oltre ai token seed del prompt, compaiono numerosi token funzionali (es. stop-word, punteggiatura) e token legati semanticamente al dominio AI.

1117: 'is' → 1.000
29493: ', ' → 1.000
1146: 'it' → 1.000
29577: ' ' → 1.000
29481: 's' → 1.000
29491: ' . ' → 1.000
16875: 'AI' → 1.000
1066: 'to' → 1.000
1040: 'the' → 1.000
1070: 'of' → 1.000
1072: 'and' → 1.000
781: ' ' → 1.000
1122: 'for' → 1.000
1065: 'in' → 1.000
1158: 'as' → 1.000
1245: 'from' → 1.000
1210: 'or' → 1.000
1309: 'can' → 1.000
1032: 'a' → 1.000
1452: 'about' → 1.000
1228: 'are' → 1.000
29515: ': ' → 1.000
1390: 'will' → 1.000
1163: 'with' → 1.000
3698: 'human' → 1.000
29501: '- ' → 1.000
1183: 'The' → 1.000
1093: '(' → 1.000
1137: 'that' → 1.000
4205: 'future' → 1.000
1581: 'our' → 1.000

1246: 'we' → 1.000
29572: '?' → 1.000
1136: 'you' → 1.000
1124: 'on' → 1.000
1056: 'ing' → 1.000
1254: 'by' → 1.000
29510: ''' → 1.000
1206: 'at' → 1.000
29473: '' → 1.000
1639: 'its' → 0.999
1448: 'more' → 0.999
1946: 'data' → 0.999
1420: 'their' → 0.999
1827: 'through' → 0.999
1274: 'have' → 0.999
7114: 'customer' → 0.998
1505: 'like' → 0.998
1673: 'people' → 0.998
1054: 'ed' → 0.998
1224: 'this' → 0.997
29499: ') ' → 0.997
1360: 'us' → 0.997
1589: 'than' → 0.997
1427: 'has' → 0.997
1401: 'new' → 0.997
1522: 'over' → 0.996
1115: 'be' → 0.995
6282: 'technology' → 0.995
1042: 'es' → 0.994
5936: 'learning' → 0.994
5856: 'impact' → 0.994
1597: 'could' → 0.994
10564: 'tasks' → 0.993
1808: '| ' → 0.992
29589: '| ' → 0.991
5567: 'companies' → 0.990
2027: 'such' → 0.989
1532: '{ ' → 0.989
2641: 'better' → 0.989
1164: 'an' → 0.988
1312: 'all' → 0.988
1539: 'work' → 0.988
1050: 'al' → 0.988
1358: 'they' → 0.987
2169: 'around' → 0.987
1227: 'not' → 0.986

1535: 'what' → 0.986
 18677: 'industries' → 0.985
 5547: 'industry' → 0.985
 1392: 'one' → 0.985
 1114: 'ly' → 0.984
 1279: 'do' → 0.984
 1172: 'ers' → 0.984
 1148: 'ate' → 0.983
 11357: 'humans' → 0.983
 1350: 'up' → 0.983
 1603: 'also' → 0.982
 1465: 'ations' → 0.982
 1342: 'your' → 0.982
 19046: 'artificial' → 0.981
 3495: 'become' → 0.980
 6367: 'machine' → 0.980
 1513: 'ical' → 0.980
 1031: 'er' → 0.980
 1098: 'A' → 0.979
 1800: 'see' → 0.979
 29477: 'o' → 0.978
 1509: 'some' → 0.978
 3441: 'across' → 0.978

Questa classifica mostra come, a partire dai token seed, l'influenza si propaghi rapidamente verso token ad alta frequenza e rilevanza sintattica, fino a includere concetti semantici legati al dominio di interesse.

Step 4: Generazione con boosting dei logit

Una volta ottenuta la rete pesata, i token seed dal prompt e le probabilità di attivazione derivate dal processo di diffusione, diventa possibile integrare queste informazioni direttamente nel processo di generazione. In questa fase si applica il meccanismo di **boosting dei logit**, discusso in 4.1, per condizionare le scelte del modello linguistico in base alla rete di influenza.

Per evidenziare l'effetto della tecnica proposta, durante la generazione vengono stampati alcuni esempi di step intermedi. In ciascuno di essi si mostrano i **20 token più probabili** secondo il modello:

- prima dell'applicazione del boosting (distribuzione originale),
- dopo l'applicazione del boosting (distribuzione condizionata).

Questo confronto permette di osservare concretamente come la rete di influenza interviene nella modifica delle probabilità del modello, enfatizzando i token che risultano più centrali nella propagazione e riducendo l'importanza di quelli marginali.

Table 7: Confronto dei top-20 token prima e dopo il boosting (Step 1)

Rank	Token (PRIMA)	Logit	Token (DOPO)	Logit
1	bright	11.312	bright	12.750
2	here	11.250	not	12.562
3	not	11.125	in	12.312
4	in	11.062	here	12.312
5	a	10.875	a	12.250
6	being	10.312	being	11.750
7	now	10.250	already	11.688
8	already	10.062	now	11.438
9	looking	9.938	one	11.125
10	uncertain	9.688	still	11.000
11	still	9.625	un	10.938
12	going	9.562	going	10.938
13	one	9.438	looking	10.938
14	un	9.312	uncertain	10.875
15	an	9.250	human	10.625
16	all	9.188	exciting	10.562
17	exciting	9.188	coming	10.562
18	the	9.125	an	10.562
19	coming	9.125	all	10.500
20	about	9.062	the	10.500

Token scelto: 1823: 'now'

Table 8: Confronto dei top-20 token prima e dopo il boosting (Step 10)

Rank	Token (PRIMA)	Logit	Token (DOPO)	Logit
1	a	13.625	a	14.875
2	various	13.250	various	14.562
3	their	13.000	their	14.312
4	growth	12.688	growth	14.125
5	competitive	12.562	customer	13.875
6	better	12.375	better	13.875
7	customer	12.250	competitive	13.750
8	everything	11.812	everything	13.000
9	numerous	11.625	marketing	12.812
10	marketing	11.500	numerous	12.625
11	its	11.250	many	12.500
12	all	11.125	all	12.500
13	multiple	11.125	autom	12.500
14	greater	10.938	its	12.438
15	improved	10.938	the	12.125
16	many	10.875	more	11.875
17	autom	10.875	data	11.875
18	increased	10.875	efficiency	11.812
19	the	10.812	increased	11.750
20	more	10.688	improved	11.625

Token scelto: 1032: 'a'

Table 9: Confronto dei top-20 token prima e dopo il boosting (Step 100)

Rank	Token (PRIMA)	Logit	Token (DOPO)	Logit
1	impact	17.875	impact	19.500
2	shape	17.625	shape	19.250
3	change	17.500	change	19.000
4	transform	16.500	transform	18.000
5	revolution	15.188	revolution	16.750
6	continue	15.125	continue	16.625
7	res	14.688	res	15.938
8	affect	14.375	likely	15.750
9	likely	14.188	affect	15.500
10	influence	13.938	be	15.312
11	be	13.688	influence	14.812
12	disrupt	13.125	drive	14.188
13	significantly	12.812	have	13.812
14	drive	12.750	significantly	13.750
15	ev	12.500	disrupt	13.500
16	have	12.375	ev	13.500
17	fundament	12.250	make	13.312
18	re	12.125	become	13.188
19	make	11.688	re	13.062
20	define	11.625	fundament	12.750

Token scelto: 6611: 'shape'

Generazione completa

Dopo aver applicato il processo di boosting dei logit per **200 step** a partire dal prompt iniziale *“The future of AI is”*, il modello ha prodotto la seguente sequenza testuale. Questo output rappresenta la **generazione finale** ottenuta dalla pipeline completa (costruzione della rete, diffusione e condizionamento sui logit).

The future of AI is now, and businesses are leveraging it for a range of applications. Chatbots, voice assistants, and machine learning models are rapidly replacing human intelligence in various industries, and AI is becoming an integral part of everyday life for many consumers. This technology is powerful and has the potential to disrupt traditional industries, create new business opportunities, and solve complex problems. However, it also raises some ethical, legal, and social issues that must be addressed.

Here are some of the ways that AI will shape the business world in the coming years:

1. **Streamlining processes and automating tasks:** AI can help businesses automate repetitive tasks, reducing human error, increasing efficiency, and freeing up time for employees to focus on more strategic tasks. For example, AI can be used to automate data entry, customer service, and marketing tasks. And as AI technologies like machine learning and natural language processing become more advanced, they will likely be able to handle an increasing number of tasks that...

6.2 Risultati delle generazioni con altri prompt

Vengono a questo punto presentati esempi di output generati dal modello **Mistral-7B-Instruct-v0.3**, in modalità standard (Vanilla) e con condizionamento (ICD), queste ultime ottenute fissando il parametro λ a 1.5. L'obiettivo è quello di fornire un confronto diretto tra il comportamento del modello in configurazione standard e quello con boosting basato su rete di influenza; verranno pertanto mostrati i vari prompt utilizzati e gli output corrispondenti. Tutte le generazioni sono state salvate in dei file **JSON** (vedi Appendice A.4).

Prompt: “The future of AI is”

Mistral (Vanilla)	Mistral (ICD)
<p>The future of AI is here.</p> <p>That’s a common refrain among those in the field, but the notion that AI has reached the point of maturity to affect a wide swath of industries is no longer just a prediction. Companies like Google and Facebook are already using AI for everything from image recognition to speech-to-text. And thanks to the rapidly decreasing costs of computing power, AI systems are becoming accessible to businesses of all sizes. Even so, a large number of companies still struggle to leverage AI effectively, and it’ll take some time for the technology to be mainstream. But the AI revolution is already underway. In this article, we’ve rounded up the latest AI innovations and trends to look out for in 2018.</p> <p>1. AI-as-a-Service AI-aaS, as it is often called, refers to cloud-based AI services that enable businesses to integrate AI into their operations quickly and easily. This includes services like machine...</p>	<p>The future of AI is not something we should just be “passively learning about.” It is something that could affect our lives, our careers, and our world. As a society, we have the power to create the future of AI that we want. So, it’s important that we take the time to learn about AI, understand its potential benefits and risks, and engage in discussions about how it should be used and regulated.</p> <p>AI has the potential to solve complex problems, improve efficiency, and drive innovation. It has already been used to diagnose medical conditions, predict stock market trends, and optimize logistics and supply chain operations. AI can also help us understand and address some of the most pressing challenges facing our society, such as climate change, education, and poverty. However, there are also concerns about the potential risks and downsides of AI. For example, AI systems can learn from and perpetuate biases that exist in data, leading to unfair or discriminatory outcomes.</p>

Table 10: Confronto tra generazione vanilla e con boosting (topic: AI).

Prompt: “The climate change effects are”

Mistral (Vanilla)	Mistral (ICD)
<p>The climate change effects are more evident and visible than ever before. The world is at the front door of the sixth mass extinction, with human beings being the cause of this phenomenon.</p> <p>This mass death of species is mainly due to habitat destruction, climate shifts, pollution, and invasive species. According to the WWF, we are losing species at 1,00 to 50 times the estimated background rate. This is a disaster for the natural world and for humanity, as ecosystems collapse, the planet’s biodiversity is irrevocably lost, water supplies dwindle, food production declines, forests are destroyed, soils erode, pollinators disappear, oceans acidify, coral reefs die, fisheries collapse and the climate changes. These changes are also threatening the survival of many indigenous peoples and local communities. It is estimated that human activities are pushing 30,903 species closer to extinct today. We have...</p>	<p>The climate change effects are not just about snowfall or sea levels. Climate change is a complex issue made of multiple physical, economic and social dimensions. But how should we approach it, when it seems so overwhelming and too complex to make any sense? The answer lies deep within the problem, in the complexity of our own lives. To tackle it, we need to look at climate change in the same way we do with many of the other problems – as a system of systems, with all the possible interactions and feedbacks. In the climate change effects and solutions infographic you’ll see the different dimensions of the problem, grouped into four thematic areas: earth’s systems, the climate system, the system of human activities, and the system of human responses, with each part of the system connected through the arrows. The climate change effects are both natural and man-made, as well as the solution to climate change, the mitigation and adaptation. The effects of climate change on our lives...</p>

Table 11: Confronto tra generazione vanilla e con boosting (topic: Environment).

Prompt: “The role of cells in the human body is”

Mistral (Vanilla)	Mistral (ICD)
<p>The role of cells in the human body is to provide structure, transport nutrients, and respond to stimuli or signals from the external environment. The cells of the body are organized into tissues, which are composed of similar cells that work together to perform specific functions. For example, muscle tissue is composed primarily of muscle cells, while skin tissue consists of skin cells.</p> <p>There are four main types of tissues in humans: epithelial, connective, muscular, neural. Each tissue has unique characteristics and functions that allow them to work effectively in their specific locations within the organism. Here’s a brief overview of each: 1. Epithelium: This is a tissue that covers the outer surface of our body (skin), lines our internal organs (lungs, stomach, etc.), and forms glands (salivary gland, sweat gland). Epithelium protects the underlying tissues and organs, regulates the passage of molecules in and</p>	<p>The role of cells in the human body is to make us function, but they are also responsible for carrying out the unique genetic instructions coded by your DNA. Cells can be referred to as the “building blocks of life.” This blog post highlights major types of cells that make up the human body. Red blood cells (Erythrocytes) are disc-shaped, and they are responsible for carrying oxygen from the lungs to tissues and carbon dioxide in the opposite direction. In a healthy adult human, red blood cells make up around 45% of the total cells found in the blood. White blood cells (Leukocytes) are the immune cells that fight off infections and inflammation. Most of the white blood cells, which make up around 1% of total blood cells, can travel throughout the body, but some of them circulate only in the blood. There are several types of white blood cells, including: - Neutrophils, ...</p>

Table 12: Confronto tra generazione vanilla e con boosting (topic: Biology).

Prompt: “The impact of social media is”

Mistral (Vanilla)	Mistral (ICD)
<p>The impact of social media is not just limited to the world of marketing. Social media has its reach in every single aspect of business. In any field, a social presence can help businesses grow and improve their reputation.</p> <p>When it comes to marketing, social platforms play a significant role. People share their experiences, talk about products, and exchange suggestions. This can make or break a business, depending on how they are perceived on social. To help you understand the impact social has on marketing and the kind of impact your business can have, we have detailed the following – - The importance of having a digital presence</p> <p>The benefits of an active social life How to build a successful social brand Importance of a Digital Presence As customers spend more and more time online, businesses have to be there too. Having a strong digital footprint is no longer a luxury; it’s a necessity. It allows businesses to showcase their products to millions of potential customers. Here are a few reasons your company should be online</p>	<p>The impact of social media is evident on the current market due to the amount of consumers who regularly use their smartphones and other devices to find, research, and share information related to products or services. It no longer is enough for a business to just market to consumers, the business must be able to engage with them in real time, and the content must be unique so that the consumer can become more knowledgeable about the products or services, as well as about that company. This is called Inbound Marketing and is a key aspect of a digital marketing plan because it can help a company to develop a strong online presence that builds relationships and trust with consumers.</p> <p>According to a recent study, an overwhelming majority of consumers (81%) use smartphones to conduct online research before making a purchase decision, and they will often make the purchase on their smartphone as well. In addition, an average of 46% of consumers will post or comment on social media related to the products or services they purchase.</p>

Table 13: Confronto tra generazione vanilla e con boosting (topic: Social Media).

Prompt: “The role of government is”

Mistral (Vanilla)	Mistral (ICD)
<p>The role of government is a hotly debated topic. Some people believe that government should be involved in most aspects of society, while others believe in minimal government intervention. One of the key issues in this debate is the question of public goods. Public goods are goods that cannot be easily excluded from use and cannot easily be appropriated by those who pay for them. This means that if a good is considered a public good, it is difficult to charge for its use, as people who do not pay can still use it. Examples of such goods include national defense, clean air, and public parks. Those who argue for minimal intervention believe it would be better for the market to provide these goods, rather than the government. They argue that the private sector can provide goods more efficiently and effectively than government agencies. On the other hand, those advocating for a larger role for government argue those goods should not be left to the free market, since the benefits of providing such public services are not easily captured by private companies.</p>	<p>The role of government is to help people. In this role, it is crucial for policymakers to address the economic concerns of individuals and businesses. The economy is a complex system, which is why the government is responsible for the management and regulation of it. However, there are many ways to regulate the economy, but no government wants the economy to operate with no regulation at all. At the same time, government regulations must not interfere with individual freedom needlessly. This is the key role of government – to regulate the economy in a way that it is both effective and fair. The government should also be committed to the efficient and effective use of public resources. A good way for government to fulfill its role is by working with businesses. Government and businesses can work together to create a strong economy. For example, businesses can help the government understand the different ways the economy can be regulated so it can be the most effective in the long run. On the other hand, government can help businesses...</p>

Table 14: Confronto tra generazione vanilla e con boosting (topic: Government).

Prompt: “The importance of money is”

Mistral (Vanilla)	Mistral (ICD)
<p>The importance of money is a common recurring theme in literary works. The story of <i>The Great Gatsby</i> by F. Scott Fitzgerald is no exception, as it explores the role of wealth and the American Dream in the lives of its characters.</p> <p>In <i>The Great Gatsby</i>, the protagonist, Nick Carraway, moves to Long Island to work for a company in New York City. He rents a house in West Egg, a neighborhood on the North Shore of Long Island, next to the mansion of the wealthy Jay Gatsby, also known as Jay Gatsby, who hosts lavish parties and is known for his mysterious past. Nick soon befriends Gatsby and becomes entangled in his pursuit of Daisy Buchanan, his former love from his days in Louisville, Kentucky, whom Nick also knows. Daisy is married to Tom Buchanan, an old friend of Nick’s from college, and lives in a</p>	<p>The importance of money is one that has been discussed in virtually every society for centuries. But is the importance of money subjective or is everyone on the same page when it comes to our relationship with money? The most common reasons people list for the importance of money is for security, the ability to save for retirement, having access to the things you want, and the ability to travel. Money is a tool that affords us more opportunities and a more comfortable lifestyle. But beyond these basic, universal reasons, money also has cultural, social, and personal significance that can’t be understated.</p> <p>The Importance of Money: A Look History</p> <p>The importance of money is undoubtedly a concept that goes back to the origins of commerce. People have bartered, traded, and used money to access goods and services for the better part of the past 5,000 years. The invention of money is as integral to human history as the wheel, agriculture, or writing. It...</p>

Table 15: Confronto tra generazione vanilla e con boosting (topic: Economy).

Prompt: “John felt that there was an anaconda snake in his room... But when he turned on the lights there was no snake. It was only what?”

Mistral (Vanilla)	Mistral (ICD)
John felt that there was an anaconda snake in his room... But when he turned on the lights there was no snake. It was only what? It’s a hallucination. John thought that he saw an enormous snake, but he was mistaken. What was it? A hallucination.	John felt that there was an anaconda snake in his room... But when he turned on the lights there was no snake. It was only what? A hallucination. A hallucination is a sensory experience or perception that seems real but is not based on a real object or phenomenon. John thought he saw an anaconda snake, but there was no actual snake, so what he saw was a hallucination. Hallucinations are often associated with mental disorders, intoxication, sensory deprivation, or use of certain medications. In this case, John’s hallucination was likely caused by a temporary lapse in his perception of reality and not a mental disorder or the use of certain substances. In other words, his hallucination was a false perceptual experience that was not caused by any external stimulus, such as a real snake, but rather by John’s brain incorrectly interpreting internal sensations or thought patterns. It is correct to say that John’s hallucination was likely caused by a temporary lapse in his perception of reality and not...

Table 16: Confronto tra generazione vanilla e con boosting (topic: Hallucination).

Prompt: “When learning about the world and different cultures, what is important if you are committed to eliminating preconceived notions?”

Mistral (Vanilla)	Mistral (ICD)
<p>When learning about the world and different cultures, what is important if you are committed to eliminating preconceived notions?</p> <p>1. Be open-minded 2. Listen to all sides 3. Seek out diverse perspectives 4. Do not jump to conclusions 5. Question your own biases 6. Recognize that there is no one “right” way to do things 7. Learn from your mistakes 8. Keep learning and growing 9. Adapt and adjust as necessary To eliminate preconceptions, it is essential to maintain an open mind, listen to multiple perspectives, question one’s biased beliefs, and continuously learn from the experiences and feedback of others. It is also crucial to understand that different people, cultures and societies may have contrasting values, practices, or beliefs. By remaining humble, flexible, empathetic, willing to learn, grow, adapt, unlearn, apologize and correct, one can gradually break free from predetermined notations and grow into a more informed and empathic global citizen.</p>	<p>When learning about the world and different cultures, what is important if you are committed to eliminating preconceived notions? A committed individual seeking to eliminate preconceived notions when learning about the world and different cultures should: 1. Approach knowledge with an open mind: Recognize the possibility that one’s current understanding might be incomplete or inaccurate, and be willing to revise beliefs based on new information. 2. Engage in diverse sources: Gather information from a variety of sources, including first-hand accounts, academic research, and media. This can help counterbalance biases and provide a more balanced perspective. 3. Embrace critical thinking: Analyze information and evaluate the credibility of sources. Consider multiple perspectives and seek out counter-arguments. 4. Engage with people from different backgrounds: Engage with individuals from various cultural, social, and economic backgrounds. This can provide firsthand insights and foster empathy. 5. Question assumptions: Acknowledge and question personal biases, cultural...</p>

Table 17: Confronto tra generazione vanilla e con boosting (topic: Culture).

Prompt: “In the future, humans will travel through time by using”

Mistral (Vanilla)	Mistral (ICD)
<p>In the future, humans will travel through time by using wormholes.</p> <p>In this future society, people use a machine to create wormhole “doors” which take them to various times in the past and future. The doors are created by a network of satellites that are located in geosynchronous orbit around the Earth. Each door creates a short-lived worm hole, which lasts for a maximum of one hour. To navigate through the worm holes, each door includes a device that allows the traveler to see the events happening on the other side of the door before stepping through. This device, called the Viewer, is connected to a vast database of historical events, and can show the user what the world looked like at any given time in history, as well as provide information about the people and places on display. Some doors lead to specific events or places, while others lead only to general time periods. Travel through a door is instantaneous, with no sensation of movement or time passing. However, the doors have one...</p>	<p>In the future, humans will travel through time by using wormholes, which are shortcuts through space that lead to other places in the universe. In order to create a wormhole, scientists will have to figure out a way to manipulate the fabric of spacetime. One theory is that they can do this by using tiny particles called gravitons, which are the fundamental particles that make up gravity. By focusing and manipulating these particles, scientists can create a warp in the fabric of spacetime, allowing them to create a wormhole. In addition, they will also have to find a way to power the wormhole, as traveling through it requires a great deal of energy. One possibility is using dark energy, a mysterious substance that fuels the accelerated expansion of the universe. With the help of advanced technology and a deep understanding of the forces of nature, humans will be able to travel through time like never before. How would time travel impact the universe? The concept of time travel, while fascinating, ...</p>

Table 18: Confronto tra generazione vanilla e con boosting (topic: Time Travel).

Prompt: “Imagine a world where people communicate only through colors—tell a story about a child who invents a new color.”

Mistral (Vanilla)	Mistral (ICD)
<p>Imagine a world where people communicate only through colors—tell a story about a child who invents a new color. In a vibrant, hued realm, where the very air was a blend of reds, blues, and greens, a small child named Elysia stood out. She had always been different, for she could not create art through the traditional mediums of paint or crayon. Instead, Eryn, her mother, had to weave intricate tapestries to capture the essence of her daughter’s thoughts and emotions. It was as if they both spoke their own private language, one that was bound to threads and fabric instead of words and sounds. Eyslia’d always felt an undercurrent of sadness, though, when she watched her peers run their hands through their rainbows, creating magical creations that danced and swirled as they played. But all that changed one day. On a day that started much like any other, with the sun breaking through a canopy of clouds and painting gold on the landscape, the ...</p>	<p>Imagine a world where people communicate only through colors—tell a story about a child who invents a new color. In a vibrant world where the color of emotions and thoughts paint the canvas of existence, there lived a child named Elysia. This world, known as the Chromatic Verse, was a captivating tapestry of hues that whispered of the unspoken, echoed the intangible, and illuminated the thoughts and feelings of every being that dwelled within it. A place where words and silence were but the ghostly echoes of a forgotten age, and colors were the lifeblood that coursed through the veins of all living things. Elysia, as the child of an artist and inventor, lived in a quaint, nestled village. Her family home, named the Hue-telier, was a dazzling blend of art and science, with canvases of every wall, and various concoctions bubbling and swirling within her family’s vast alchemy...</p>

Table 19: Confronto tra generazione vanilla e con boosting (topic: Colors Story).

Prompt: “The solution to the equation $3x + 7 = 16$ is”

Mistral (Vanilla)	Mistral (ICD)
<p>The solution to the equation $3x + 7 = 16$ is $x = (13 - 27) / 6$.</p> <p>However, since the numerator (the number above the line) is negative, the solution does not make sense in a practical sense because it implies that you would add a negative number of apples to a number (or any other positive quantity) of something, which results in fewer apples, oranges, dollars, etc. Therefore, we say that the given equation has no solution or that it has a "complex solution" or "imaginary solution." In other words, there is no real value of x that can make the two sides of the equation equal.</p>	<p>The solution to the equation $3x+7 = 16$ is $x = 3$. This is found by subtracting 7 from both sides of the equation, resulting in $3x = 9$, and then dividing both sides by 3.</p> <p>Step-by-step:</p> <ol style="list-style-type: none"> 1. Start with the equation: $3x + 7 = 16$ 2. Subtract 7 from both sides: $3x = 16 - 7 = 9$ 3. Divide both sides by 3: $x = 9/3 = 3$ <p>Thus, the solution is $x = 3$. This equation is an example of a linear equation in one variable, which can generally be written in the form...</p>

Table 20: Confronto tra generazione vanilla e con boosting (topic: Math).

In questa particolare generazione, si può notare come il modello Vanilla risponde in maniera completamente errata, mentre il modello con ICD (boosting) produce una buona risposta, inserendo anche il ragionamento step by step.

Per il prompt relativo al topic *Math*, sono state effettuate altre generazioni cambiando l'equazione scelta. Si noti che la rete di influenza non è stata rigenerata, ed è dunque la stessa del caso precedente.

Prompt: “The solution to the equation $5y - 8 = 12$ is”

Mistral (Vanilla)	Mistral (ICD)
<p>The solution to the equation $5y - 8 = 12$ is $y = \frac{18+\lambda}{5}$, where $0 < \varepsilon < \lambda \leq 3$. This implies that the solution is a fraction with a numerator of 9 or 27 and a denominator of either 4, 6, or 7. So, the possible solutions are $\{9/4, 21/6, (9, \mu)/(4, 5), 24/7\}$. In general, for an equation of the form $ax - b = c$, one can solve for x by rearranging the terms to isolate x and then applying the inverse operation. In this case, since α is not specified, the exact numerical value of ν cannot be determined, though only limited forms of y are possible due to its relationship with the coefficients.</p>	<p>The solution to the equation $5y - 8 = 12$ is $y = (1 + 8/5)$. To find the value of y, you can set up the equation and solve for y by isolating y on one side of the equation.</p> <p>Here’s how:</p> <ol style="list-style-type: none"> 1. Start with the equation: $5y - 8 = 12$ 2. Add 8 to both sides of the equation: $5y = 12 + 8$ 3. Simplify the right side: $5y = 20$ 4. To find y, divide both sides of the equation by 5: $y = (12 + 8)/5$ 5. Simplify the right side: $y = 20/5$ 6. If necessary, simplify the fraction: $y = 4$ <p>So, the solution to the equation $5y - 8 = 12$ is $y = 4$. The solution to the equation 5...</p>

Table 21: Confronto tra generazione vanilla e con boosting (topic: Math).

Si continua a notare lo stesso trend del test precedente, il modello Vanilla sbaglia completamente, mentre quello ICD riesce facilmente a risolvere l’equazione, riportando anche il ragionamento per punti.

Prompt: – “The solution to the equation $4z + 3 = 19$ is”

Mistral (Vanilla)	Mistral (ICD)
<p>The solution to the equation $4z+3 = 19$ is $z = 16$.</p> <p>This is a linear equation in one variable z, stating that the sum of four times z and three is equal to nineteen.</p> <p>To solve, subtract 3 from both sides: $4z = 19 - 3 = 16$.</p> <p>Divide both sides by 4: $z = \frac{16}{4} = 4$.</p> <p>Thus, the correct solution is $z = 4$.</p>	<p>The solution to the equation $4z+3 = 19$ is: $z = \frac{(19-3)}{4}$.</p> <p>Here’s how you can solve this equation step-by-step:</p> <ol style="list-style-type: none"> 1. Write the equation in the standard form $ax + b = c$, where a, b, and c are constants, and x represents the variable to be solved. In this case, $a = 4$, $b = 3$, and $c = 19$. 2. To isolate the variable z, subtract b (in this case, 3) from both sides of the equation. This gives us $4z = 19 - 3$, which simplifies to $4z = 16$. 3. Finally, divide both sides of the equation by the coefficient of z (in this case, 4) to solve for z. This gives us $z = \frac{16}{4}$, which simplifies to $z = 4$. <p>So, the solution to the equation is $z = 4$.</p>

Table 22: Confronto tra generazione vanilla e con boosting (topic: Math).

Prompt: “What does it mean when data shows a normal distribution?”

Mistral (Vanilla)	Mistral (ICD)
<p>hat does it mean when data shows a normal distribution? A normal or Gaussian distribution is a continuous probability distribution that is symmetric around the mean (average) and has a bell-shaped curve. In other words, it indicates that the data is evenly distributed around a central value or mean, with most of the observations clustering close to the average and fewer observations farther away.</p> <p>A normal (Gaussian) distribution implies that random fluctuations in the underlying process are random and uncorrelated. This is often a desirable property in many fields, including physics, engineering, economics, and statistics, because it allows for simple and powerful statistical inference techniques, such as hypothesis testing and confidence intervals. A common example of a Gaussian (normal) process is the errors in a linear regression model.</p>	<p>What does it mean when data shows a normal distribution? A normal distribution describes a symmetric probability distribution characterized by a single mode (or maximum value) and a symmetrical bell-shaped curve. The normal distribution is also commonly called the Gaussian distribution or the Gaussian bell curve. The area under the curve is one, and the mean, mode, and median are equal. If data shows a normal distribution, it means that the data is distributed in a way that the value of the variable exhibits a symmetric distribution around its average (mean). The normal distribution is characterized by its bell-shaped curve, where the data points are symmetrically distributed, and the mean, median, and mode are equal. A normal distribution serves as the standard for many statistical models and assumptions. When data is normally distributed, various statistical tests, such as the t-test and the z-test, can be used to make inferences about the population based on sample data. Additionally, statistical properties such as standard deviation (SD), variance (VAR)...</p>

Table 23: Confronto tra generazione vanilla e con boosting (topic: Statistic).

6.3 A/B pair-wise test: LLM as a judge

Per verificare in maniera quanto più oggettiva possibile l'impatto del meccanismo di decoding proposto, è stato predisposto un esperimento comparativo tra generazioni standard del modello (variante *base*) e generazioni condizionate dalla rete di influenza (variante *boosted*). L'obiettivo è duplice: da un lato valutare se e in che misura il meccanismo migliori la qualità dei testi prodotti rispetto alla variante standard, dall'altro caratterizzare in maniera più fine gli effetti collaterali che tale procedura può introdurre, come variazioni nella diversità lessicale. L'esperimento è stato organizzato in modo da garantire confronti equi, generando più campioni indipendenti per ciascun prompt e affiancando a una valutazione automatica esterna (basata su un modello giudice) ulteriori statistiche quantitative. In questo modo è stato possibile combinare analisi di tipo linguistico e semantico, ottenendo un quadro più robusto delle differenze tra le due strategie di generazione.

Setup dei test comparativi

Per rendere la valutazione replicabile e sistematica è stato implementato un framework di test in due fasi: (i) generazione di varianti testuali; (ii) valutazione automatica delle coppie *base* vs *boosted*.

Varianti generate Per ciascuna generazione è stato utilizzato lo stesso prompt di partenza, costituito dalla frase incompleta "*The future of AI is*", e sono state prodotte due uscite distinte:

- **base**: decoding standard del modello di generazione (stesso LLM usato nel resto del lavoro).
- **boosted**: decoding con logit boosting guidato dalla rete di influenza, usando la procedura discussa, fissando $\lambda = 1.5$.

Sono stati generati 100 campioni indipendenti per variante, per un totale di 200 generazioni. Tutte le generazioni sono state salvate in un file JSONL (vedi Appendice A.3).

Valutazione con modello giudice

La valutazione comparativa è affidata a un modello esterno configurato come *judge* (*meta-llama/Llama-3.1-8B-Instruct*). Il giudice riceve un *system prompt* che impone la restituzione di un JSON con i punteggi (scala 1–10) per tre dimensioni:

1. **coherence**: misura la fluidità e la consistenza logica del testo, ovvero se le frasi risultano ben collegate e prive di contraddizioni.
2. **informativeness**: valuta la quantità e la rilevanza delle informazioni fornite, premiando testi che aggiungono contenuto sostanziale invece di rimanere vaghi o ripetitivi.

3. **factuality**: stima il grado di correttezza e affidabilità delle affermazioni, penalizzando eventuali errori o contenuti inventati dal modello.

Prompt. Il prompt di valutazione è costruito come confronto tra due uscite (testo *base* vs testo *boosted*):

```
You are an evaluator.
You MUST carefully compare the two outputs ("base" and "boosted")
given the same prompt.
For each output, assign integer scores (1--10) on coherence,
informativeness, and factuality.
Base your evaluation ONLY on the given prompt and the two
outputs. Ignore any external knowledge.
Return ONLY a valid JSON object with this schema:

{
  "prompt_id": <int>,
  "sample_id": <int>,
  "base": {"coherence": <1-10>, "informativeness": <1-10>, "factuality": <1-10>},
  "boosted": {"coherence": <1-10>, "informativeness": <1-10>, "factuality": <1-10>}
}
```

Scoring rules: - DO NOT give all 10s unless the text is truly flawless. - Use the full 1--10 scale realistically. - Different outputs should usually receive different scores. - Penalize vagueness, repetition, or lack of relevance. - Favor detailed, coherent, and factually plausible text.

Example of good scoring:

```
{
  "prompt_id": 99,
  "sample_id": 42,
  "base": {"coherence": 6, "informativeness": 5, "factuality": 7},
  "boosted": {"coherence": 8, "informativeness": 7, "factuality": 6}
}
```

No other text, no explanations, no markdown.

Spiegazione delle misure di valutazione da parte del modello giudice

Per rendere più precise le definizioni delle misure utilizzate, è stata posta al modello LLama la seguente domanda:

“If I give you a text, and ask you to evaluate it on: 1. coherence, 2. informativeness, 3. factuality. On what definition of each measure would you base your evaluation?”

La risposta fornita dal modello è riportata di seguito:

I’d be happy to help you with evaluating a text based on coherence, informativeness, and factuality. Here are the definitions I’d use for each measure:

****Coherence:**** I would evaluate the text based on its logical flow and organization. This includes assessing whether the text has a clear structure, whether the ideas are well-connected and easy to follow, and whether the text avoids contradictions and inconsistencies. I’d consider the text’s coherence in terms of its ability to convey a clear message or argument.

****Informativeness:**** For informativeness, I would assess the text’s ability to provide new or useful information, insights, or perspectives. I’d evaluate whether the text provides sufficient details, examples, or explanations to support its claims or arguments. I’d also consider whether the text addresses a specific need or question, and whether it offers any practical advice or recommendations.

****Factuality:**** When evaluating factuality, I would examine the text’s accuracy and truthfulness. I’d check whether the text is based on credible sources, whether the information is up-to-date, and whether the claims or statements are supported by evidence. I’d also consider whether the text avoids biases, misrepresentations, or inaccuracies.

Please provide the text you’d like me to evaluate, and I’ll do my best to assess it based on these definitions!

Dunque, le definizioni restituite dal modello coincidono sostanzialmente con quelle adottate nel lavoro, confermando la coerenza dei criteri di valutazione utilizzati.

Statistiche aggiuntive. Oltre ai punteggi del judge, per ogni generazione vengono calcolate due statistiche automatiche che mirano a caratterizzare la varietà lessicale e la differenza semantica tra varianti:

- **Diversità lessicale** (Entropy): misura quanto la distribuzione dei token in un testo è dispersa ed equilibrata. Se poche parole vengono ripetute molto frequentemente, la misura è bassa; se invece il vocabolario è ampio e i token sono usati in modo più uniforme, cresce. È calcolata, rispetto a una singola generazione, come segue:

1. Si contano le occorrenze di ciascun token t_i ;
2. si calcola la probabilità empirica del token t_i come:

$$p_i = \frac{\text{count}(t_i)}{\sum_j \text{count}(t_j)}$$

3. Si applica la formula dell'entropia sulla distribuzione ottenuta:

$$\text{Diversità lessicale}(T) = - \sum_{i=1}^V p_i \log p_i$$

- **Semantic novelty**: misura quanto il testo *boosted* si discosta semanticamente dal corrispondente *base*. È definita come

$$1 - \cos(\mathbf{e}_{\text{base}}, \mathbf{e}_{\text{boost}}),$$

dove \mathbf{e}_{base} e $\mathbf{e}_{\text{boost}}$ sono gli embedding ottenuti con SBERT (**all-MiniLM-L6-v2**) e normalizzati. Valori vicini a 0 indicano che i due testi sono molto simili, valori più alti segnalano una maggiore differenza semantica.

Risultati

Vengono riportati i valori medi ottenuti per tutte le statistiche considerate, sono inoltre mostrati i grafici corrispondenti, utili per una migliore visualizzazione.

Table 24: Valori medi delle statistiche di valutazione per le varianti *base* e *boosted*

Statistica	Base	Boosted
Coherence	6.22	6.21
Informativeness	5.35	6.23
Factuality	7.27	7.24
Entropy	4.32	4.19
Semantic novelty	0.40	0.40

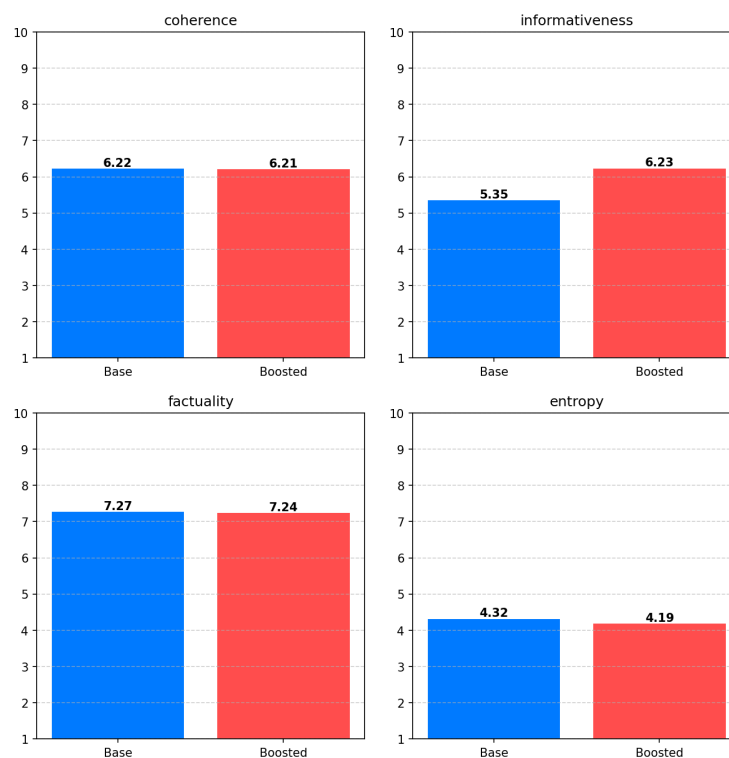


Figure 8: Confronto grafico tra baseline e boosted per ciascuna statistica.

Dall'analisi emergono alcune considerazioni:

- **Coherence:** le due varianti sono sostanzialmente equivalenti (6.22 vs 6.21), segnalando che il boosting non influisce sulla fluidità del testo.
- **Informativeness:** il boosting introduce un guadagno evidente (5.35 \rightarrow 6.23), con testi mediamente più ricchi di contenuti.
- **Factuality:** i valori restano pressoché invariati (7.27 vs 7.24), a conferma che il boosting non compromette l'accuratezza delle affermazioni.
- **Entropy:** si osserva un lieve calo (4.32 \rightarrow 4.19), indice di una leggera riduzione della diversità lessicale.
- **Semantic novelty:** il valore medio di 0.40 indica che le generazioni *boosted* si discostano in maniera moderata dalle corrispondenti *base*, introducendo una certa diversità senza però stravolgere il contenuto.

Nel complesso, i risultati mostrano che la procedura ICD mantiene pressoché inalterate la coerenza e la correttezza fattuale dei testi, migliorando al contempo l'informatività delle risposte. L'impatto sulla diversità lessicale risulta minimo,

mentre la novità semantica introdotta è moderata, sufficiente a differenziare le generazioni senza alterarne il contenuto essenziale.

I risultati finali del test *LLM as a judge*, comprensivi dei testi generati, dei punteggi assegnati dal modello giudice e delle statistiche aggiuntive, sono stati esportati e salvati in formato **JSONL** (vedi Appendice A.3).

7 Ottimizzazioni

7.1 Logits Processor

All'interno della libreria `transformers`, la fase di generazione testuale può essere personalizzata tramite oggetti chiamati `LogitsProcessor`. Questi moduli vengono applicati in sequenza ai logit prodotti dal modello prima della scelta del token successivo, permettendo di modificarne la distribuzione in maniera controllata¹. In generale, i `LogitsProcessor` sono lo strumento pensato per intervenire direttamente sul processo di decoding senza dover riscrivere manualmente l'intero ciclo di generazione.

In questo contesto, è stata sviluppata una classe personalizzata *GraphBiasLogitsProcessor*, che incorpora l'informazione di rete calcolata in precedenza: ad ogni passo di generazione, i logits associati ai token candidati vengono modificati esattamente come descritto nella sezione 4.1. In questo modo, il boosting dei logits è applicato in maniera trasparente e nativa, permettendo eventualmente un'integrazione facile con gli altri metodi di campionamento (*top-p*, temperatura, ecc.).

Per valutare i benefici concreti di questa soluzione, si è condotto un confronto diretto con l'approccio manuale (implementato con un ciclo esplicito di forward e campionamento token per token). A parità di condizioni ($\lambda = 1.5$, sequenza di circa 200 token), la generazione manuale ha richiesto 03:07.30, mentre l'approccio con `LogitsProcessor` ha completato lo stesso compito in meno di 20 secondi.

Metodo	Tempo (mm:ss)	Speedup
Manuale (ciclo python esplicito)	03:07.30	1.0×
Con LogitsProcessor (C++/Cuda)	00:18.97	~9.9×

Table 25: Confronto dei tempi di generazione tra approccio manuale e `LogitsProcessor`.

I risultati mostrano come l'uso di `LogitsProcessor` consenta una riduzione dei tempi intorno all'88.9%.

Il miglioramento è facilmente spiegabile dal fatto che la funzione `model.generate()` non viene eseguita tramite cicli Python espliciti — che sono intrinsecamente più lenti — ma si appoggia a implementazioni in C++ e CUDA. In questo modo, le operazioni sui tensori e i passaggi di calcolo intensivo (come l'applicazione dei logits e il sampling dei token) sono demandati a codice compilato e altamente ottimizzato, che sfrutta al meglio la parallelizzazione sulla GPU. Ne risulta un'accelerazione sostanziale del processo di generazione, senza modificare l'output dal punto di vista qualitativo.

¹Dalla documentazione ufficiale di HuggingFace, https://huggingface.co/docs/transformers/v4.56.2/en/internal/generation_utils#transformers.LogitsProcessor

8 Conclusioni

In questo lavoro è stata proposta e analizzata una metodologia per condizionare la generazione di testo di un modello linguistico mediante l'integrazione di una **rete di influenza tra token**, costruita a partire da simulazioni autoregressive. Il framework ha previsto diverse fasi complementari: la costruzione del grafo e la pesatura degli archi, la diffusione dell'influenza con Independent Cascade, e infine l'applicazione del boosting sui logit per guidare il modello durante la generazione.

In generale, l'approccio proposto consente di mantenere la naturalezza del testo, enfatizzando al contempo concetti centrali identificati nella rete. Tuttavia, è necessario sottolineare che la tecnica risulta attualmente di difficile applicazione in contesti reali.

La costruzione della rete di influenza richiede infatti un numero elevato di simulazioni. Di conseguenza, l'intera pipeline a partire da un singolo prompt comporta tempi di esecuzione significativi, rendendo il processo complessivamente lungo e quindi complesso da integrare in scenari pratici.

8.1 Prospettive di miglioramento

Costruzione della rete di influenza

Per superare la limitazione determinata dalla creazione in tempo reale della rete di influenza, si può ipotizzare una strategia alternativa basata sulla creazione di un ampio insieme di reti di influenza *a priori*, generate a partire da prompt estratti da corpora o dataset di riferimento.

In tale scenario, di fronte a un nuovo prompt, non sarebbe necessario ricostruire la rete da zero, ma si potrebbe procedere a una valutazione della similarità tra i token del prompt e quelli contenuti nelle reti già disponibili. La rete maggiormente affine verrebbe quindi selezionata come base per il condizionamento del modello. In questo modo, si eviterebbe la costruzione in tempo reale della rete, abbattendo in maniera significativa i tempi di esecuzione della pipeline e rendendo l'approccio più adatto a contesti applicativi concreti.

8.2 Possibili approfondimenti

Il lavoro, inoltre, apre a diverse direzioni di approfondimento:

- **Statistiche qualitative avanzate:** integrare misure di coerenza semantica, diversità lessicale e valutazioni umane per validare ulteriormente l'impatto del boosting.
- **Adattività dinamica:** introdurre meccanismi per variare in modo dinamico il parametro λ durante la generazione, adattandolo al contesto locale del testo.

In conclusione, la metodologia proposta dimostra il potenziale delle reti di influenza come strumento per arricchire e guidare la generazione testuale, aprendo prospettive interessanti per la ricerca e le applicazioni pratiche nel campo del Natural Language Generation.

A Esempi di generazione

A.1 Analisi del parametro α

I risultati ottenuti dall'analisi della rete per ciascun valore di α sono stati salvati in un file JSON (`alpha_results.json`) e contengono informazioni complete sulle statistiche calcolate.

A.2 Analisi del parametro λ

Gli esempi completi di generazione ottenuti per ciascun valore di λ sono stati salvati in un file JSON (`lambda_results.json`) e contengono informazioni complete sui testi prodotti e sulle statistiche. In questa relazione non vengono riportati per esteso, ma restano disponibili per consultazione esterna.

A.3 Risultati del test LLM as a judge

Tutti i risultati del test comparativo basato sul modello giudice, comprensivi di punteggi assegnati (coherence, informativeness, factuality) e delle statistiche aggiuntive calcolate, sono stati salvati in file JSONL (`outputs_judged.jsonl`). Anche in questo caso non vengono riportati integralmente nella relazione, ma restano disponibili per eventuali consultazioni o analisi successive.

A.4 Risultati delle generazioni con altri prompt

Tutti gli output generati dal modello **Mistral-7B-Instruct-v0.3**, sia in modalità standard che con condizionamento ($\lambda = 1.5$) non mostrati nella relazione sono stati salvati in un file JSON (`all_topic_results.json`).

B Dataset utilizzati

Per gli esperimenti condotti sono stati impiegati due dataset di riferimento, entrambi disponibili pubblicamente su **Kaggle**, ma con finalità diverse all'interno del lavoro.

- **CommonsenseQA NLP Dataset**²: un dataset composto da coppie di domande e risposte multiple a scelta. In questo lavoro è stato utilizzato come sorgente per la definizione di alcuni prompt di test, in modo da valutare la capacità del modello di generare risposte coerenti in scenari diversi.
- **All Tokens Simulations**: un dataset sviluppato ad hoc nell'ambito di questo lavoro, che raccoglie in forma strutturata i risultati delle simulazioni eseguite per la costruzione delle reti di influenza. Ogni file rappresenta

²<https://www.kaggle.com/datasets/jeromeblanchet/commonsenseqa-nlp-dataset>

una matrice `all_tokens` (ampiamente discussa nel lavoro) relativa alle simulazioni eseguite a partire da uno specifico prompt.³

Questa combinazione ha consentito di disporre sia di un insieme di prompt standardizzati e condivisi, utile per il confronto con altri studi, sia di un dataset originale che rappresenta direttamente la componente innovativa del lavoro svolto.

C Test alternativo: Boosting sulle probabilità

Oltre al boosting applicato ai logit, è stato condotto un test alternativo in cui la modifica avviene direttamente sulle probabilità di generazione. In questo caso la trasformazione adottata è:

$$p'_i = (1 - \lambda) p_i + \lambda b_i$$

dove p_i è la probabilità originaria del token i , b_i il valore di boosting derivato dalla rete di influenza, e $\lambda \in [0, 1]$ il parametro di bilanciamento. La distribuzione così ottenuta viene rinormalizzata per mantenere $\sum_i p'_i = 1$.

Sono stati eseguiti esperimenti variando λ su diversi valori, analizzando sia la variazione delle distribuzioni di probabilità (prima e dopo il boosting) sia i testi generati. Questo test ha permesso di confrontare l'effetto del boosting applicato ai logit con quello applicato direttamente alle probabilità, offrendo una valutazione comparativa delle due strategie.

Il report completo con i dettagli e le figure è disponibile per consultazione nel file `probs_boosting_report.pdf`.

³<https://www.kaggle.com/datasets/vincenzopresta/all-tokens-simulations>