



Armandillo

WHITE PAPER — APRILE 2026

La prima **AI civica** italiana al servizio della Pubblica Amministrazione

Un modello linguistico specializzato, un corpus aperto di oltre 1,4 milioni di documenti, una pipeline RAG trasparente. Tutto ospitato su infrastruttura italiana, tutto riproducibile.

Armando Passaro

Fondatore e autore di Armandillo
armandillo.it · passaroarmando@gmail.com

Versione 1.0
20 aprile 2026



Sintesi

Armandillo è la prima intelligenza artificiale civica italiana: un assistente conversazionale pensato per sindaci, segretari comunali, RUP, dirigenti e funzionari della Pubblica Amministrazione. Parla italiano amministrativo, cita le fonti ufficiali, funziona senza inviare dati a provider esteri.

■ Il problema

I modelli linguistici generalisti (GPT, Gemini, Claude) sono straordinari ma non conoscono il decreto di affidamento del Comune di Vigevano, non sanno come è strutturato un albo pretorio, non distinguono una delibera di giunta da una determina dirigenziale. Per la PA italiana questo è un limite operativo.

■ La nostra risposta

Armandillo combina **tre elementi**: (i) un modello linguistico da 7 miliardi di parametri fine-tuned su domande e risposte reali della PA, (ii) un corpus documentale italiano di oltre **1,47 milioni di chunk** normativi e amministrativi aggiornato quotidianamente, (iii) una pipeline RAG (Retrieval-Augmented Generation) che garantisce risposte basate su fonti verificabili.

■ I numeri oggi

1,47M

CHUNK INDICIZZATI

7.208

COMUNI MONITORATI

301K

DOCUMENTI RAW

137

FONTI NAZIONALI

DIFFERENZA CHIAVE

Mentre i modelli generalisti producono risposte statisticamente plausibili ma spesso inventate (allucinazioni), Armandillo **cita sempre il documento di origine**: delibera, determina, circolare, norma. L'operatore PA può verificare la fonte in un click.

■ A chi è rivolto

Enti locali (comuni, province, unioni, città metropolitane), società in-house, pubbliche amministrazioni centrali, società di consulenza alla PA, studi professionali che seguono enti pubblici, università e centri di ricerca interessati al tema PA-IA.

■ Perché adesso

Il **DDL Intelligenza Artificiale** (approvato in via definitiva nel settembre 2025) impone alle PA italiane di adottare sistemi IA tracciabili, spiegabili e rispettosi della normativa nazionale. Le linee guida AgID sull'uso dell'IA nella PA (2025) raccomandano esplicitamente **soluzioni italiane, open source ove possibile, con ospitamento nazionale**. Armandillo è nato già conforme a questi principi.

Un prodotto offerto da centrostudipa.it · Il Primo Osservatorio Nazionale Indipendente sull'IA e sull'Open Source nella PA



Architettura tecnica

■ Il modello: Armandillo 7B-PA

Armandillo si basa su un modello open-source da 7 miliardi di parametri (Qwen 2.5 Instruct come fondazione), sottoposto a **fine-tuning QLoRA** su un dataset proprietario di oltre 2.000 domande e risposte tratte dalla pratica amministrativa italiana (anno 2012 come baseline storico, aggiornamenti 2024-2026 per normativa vigente). Il modello è quantizzato **AWQ 4-bit** per girare su GPU RTX 3060 12 GB con costi operativi contenuti.

■ Il corpus: RAG italiano verticale

Il corpus è costruito da **tre filiere di harvesting indipendenti**:

- **Albi pretori** dei comuni italiani — atti ufficiali, ordinanze, avvisi, bandi (aggiornati quotidianamente).
- **Sezioni "Amministrazione Trasparente"** di 7.208 comuni con URL trasparenza verificati — consulenze, bilanci, pagamenti, sovvenzioni, bandi, incarichi.
- **137 fonti nazionali** organizzate in 23 categorie tematiche: ministeri, giurisprudenza, dati/statistiche, normativa, fondi/PNRR, appalti, ambiente, cultura, istruzione, sicurezza, tributi, urbanistica, welfare, demografici, CCNL, digitale, regioni.

■ Pipeline di indicizzazione (8 fasi sequenziali)

1. Discover Albi — scoperta nuovi albi comunali	2. Harvest Albo — crawling atti ufficiali
3. Harvest Trasparenza — 7.208 comuni × 200 pagine	4. Harvest Nazionali — 137 seed × 300 pagine
5. Anonymize — regex NER su dati personali	6. Extract — normalizzazione in JSONL
7. Push RAW — backup S3 (MinIO)	8. FAISS Index — embedding + ricerca vettoriale

■ RAG ibrido: semantico + keyword

Il motore di ricerca combina due approcci complementari: **FAISS** (indice vettoriale basato su embedding dense da 384 dimensioni, modello MiniLM-L6-v2) per la somiglianza semantica, e **SQLite FTS5** (BM25 ranking) per la corrispondenza lessicale esatta (sigle, numeri di protocollo, CUP, CIG). Il risultato è un retrieval più robusto rispetto ai RAG solo-vettoriali.

■ Infrastruttura

Tutto è ospitato in Italia, su infrastruttura proprietaria:

Ruolo	Hardware	Funzione
Inferenza chat	2× NVIDIA RTX 3060 12 GB	vLLM con modello Armandillo 7B-PA AWQ
Embedding / TTS	2× NVIDIA Tesla P100 16 GB	nomic-embed, coder-14B, XTTS v2 voce
Speech-to-text	2× NVIDIA Tesla P4 8 GB	Whisper large-v3-turbo ×4 istanze
Corpus RAG	TrueNAS 10 TB mirror	142k file JSONL, FAISS 2,2 GB
Reverse proxy	VM Ubuntu hardened	Apache + ModSecurity WAF, Let's Encrypt



Sicurezza, trasparenza, sovranità digitale

■ Nessun dato esce dall'Italia

Armandillo non si appoggia a OpenAI, Anthropic, Google, AWS Bedrock o qualsiasi altro provider estero. L'inferenza, l'embedding, lo storage del corpus, le interfacce utente, i database di autenticazione: tutto gira su GPU fisiche collocate in territorio italiano, di proprietà diretta. Nessun contratto di data processing con soggetti extra-UE.

■ Perimetro di sicurezza attivo

Ogni dominio dell'ecosistema passa da un reverse proxy hardened con:

- **WAF ModSecurity** + anti-DDoS mod_evasive
- **Fail2ban** con 10 jail (SSH, bot, shellshock, ModSec)
- **Threat intelligence AI** su 1.009 fonti cybersecurity (CERT, CVE feed, vendor, blog), oltre **64.500 frasi apprese** con embedding per auto-blocco IOC
- **Honeypot SSH + HTTP** con ban automatico
- **IDS custom** con anomaly detection (Isolation Forest)
- **VPN WireGuard** per l'accesso amministrativo

■ Anonimizzazione del corpus

Prima dell'indicizzazione, ogni documento passa attraverso una fase di **Anonymize** che rimuove via regex + NER i dati personali identificativi residui (codici fiscali, numeri di telefono, email private). L'utente finale vede solo dati già pubblici per legge (atti amministrativi), ripuliti da eventuali PII passate inavvertitamente dai siti origine.

I numeri del corpus (dato live)

1.476.882

CHUNK FAISS

301.448

FILE RAW

142.000

FILE NORMALIZZATI

384 dim

EMBEDDING

■ Aggiornamento quotidiano

La pipeline corpus gira in **ciclo continuo 24/7**: appena finisce un giro completo ne ricomincia uno nuovo. Il ciclo completo dura mediamente **200 ore** (harvest Albi 20h, harvest Trasparenza 48h, harvest Nazionali 30h, estrazione/indicizzazione 100h). Ogni giorno l'indice FAISS cresce di 10-20mila nuovi documenti.

WATCHDOG ANTI-BLOCCO

Per garantire che nessun ciclo si blocchi su siti comunali non rispondenti, ogni harvester ha un **watchdog a 30 minuti**: se un comune o un dominio nazionale non restituisce nuovi dati per mezz'ora, viene saltato e registrato. Il ciclo non si ferma mai su un singolo ente.

■ Voce e trascrizione

Armandillo è multimodale: accetta input vocali (Whisper large-v3-turbo, 4 istanze ridondate) e risponde con voce sintetica italiana naturale (XTTS v2 di Coqui AI, fine-tunata su voci italiane). Un operatore PA può parlare al telefono ad Armandillo mentre cammina per il corridoio del Comune.



Casi d'uso per la PA

■ 1. Supporto agli atti amministrativi

Un funzionario chiede: «Come motivo un affidamento diretto sotto soglia per un servizio culturale?» Armandillo risponde citando l'art. 50 del D.Lgs. 36/2023, linee guida ANAC, e mostra **esempi reali** di determine già pubblicate da altri comuni italiani in casi analoghi, linkando gli atti originali.

■ 2. Benchmarking inter-comunale

Un segretario comunale può chiedere: «Quanto spendono i comuni tra 5.000 e 10.000 abitanti per la manutenzione stradale?» Armandillo interroga i bilanci pubblicati in Amministrazione Trasparente e fornisce un quadro comparativo.

■ 3. Formazione continua del personale

Ogni operatore ha a disposizione 24/7 un assistente che spiega la normativa aggiornata, con il vantaggio di risposte calate nel contesto operativo: non «cosa dice la legge in generale» ma «cosa stanno facendo i comuni in questo mese».

■ 4. Trasparenza proattiva verso il cittadino

Le piattaforme **datipa.armandillo.it** e **trasparenzaincomune.armandillo.it** espongono a cittadini, giornalisti e associazioni dati pubblici altrimenti dispersi in migliaia di siti comunali. Il cittadino cerca, Armandillo trova.

■ 5. Gestione delle segnalazioni (D.Lgs. 24/2023)

Il modulo Whistleblowing PA permette agli enti di ricevere segnalazioni interne in forma protetta, con cifratura end-to-end lato server e anonimato opzionale per il segnalante. Integrato nell'ecosistema Armandillo.

■ 6. Adeguamento NIS2 (Direttiva UE 2022/2555)

Il modulo gap-analysis interroga 100 domande strutturate sulla compliance NIS2 e produce un piano di rientro con 9 policy template reali. Pensato per i soggetti essenziali e importanti individuati dal D.Lgs. 138/2024.

Roadmap 2026-2027

Periodo	Obiettivo
Q2 2026	Release pubblica chat Armandillo, Open WebUI con SSO, app mobile PWA
Q3 2026	Ampliamento corpus a province, regioni e città metropolitane (~700 nuovi enti)
Q3 2026	Fine-tuning v2 su 5.000 nuove Q&A raccolte dall'uso reale
Q4 2026	Moduli verticali: Bilancio Partecipato, AI Deliberazioni, Whistleblowing a self-registration
Q1 2027	Adesione di almeno 50 comuni pilota come partner, gratuito per fascia < 10k abitanti
2027	Pubblicazione del dataset Q&A PA con licenza aperta (CC BY-SA 4.0)

“ Armandillo non vuole sostituire il lavoro del pubblico dipendente. Vuole restituirgli il tempo rubato dalla burocrazia ripetitiva, e mettergli in mano uno strumento che conosce il suo mondo. ”

■ Come partecipare

- **Enti pubblici:** scrivere a passaroarmando@gmail.com per aderire al programma comuni pilota.
- **Ricercatori e giornalisti:** accesso API al corpus per lavori di ricerca (gratuito previo accordo).
- **Sviluppatori:** i moduli verticali (Bilancio Partecipato, AI Deliberazioni, Whistleblowing) saranno rilasciati come software libero.
- **Cittadini:** armandillo.it · chat.armandillo.it · datipa.armandillo.it