



Armandillo

WHITE PAPER — GIUGNO 2026

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

Un modello linguistico specializzato, un corpus aperto di oltre **1,26 milioni di chunk** indicizzati, una pipeline RAG trasparente. Tutto ospitato su infrastruttura italiana, tutto riproducibile.

Armando Passaro

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

Versione 2.0

28 giugno 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 dei Comuni, 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 **1,26 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

1,26M

CHUNK
INDICIZZATI

7.896

COMUNI
MONITORATI

260K

ALLEGATI
RACCOLTI

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). Coperti oggi **7.341** comuni.
- **Sezioni "Amministrazione Trasparente"** di 7.896 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 — 7.896 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)

1.261.306

CHUNK FAISS

260.098

ALLEGATI 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.

Infrastruttura: Proxmox · PostgreSQL 15 · FAISS · SQLite FTS5 · vLLM · Ollama · Docker · WireGuard



Aggiornamenti recenti

Cronologia delle ultime evoluzioni del progetto, in ordine cronologico inverso. Tutte le novità sono pubblicate live su armandillo.it/news.

NUOVA FUNZIONE

07/06/2026

Tre nuove piattaforme PA: Bilancio Partecipato, AI per gli Atti e Whistleblowing

L'ecosistema Armandillo si arricchisce di tre nuovi servizi verticali per la Pubblica Amministrazione: partecipazione civica, generazione assistita di delibere e segnalazioni anonime conformi al D.Lgs. 24/2023. Esteso inoltre il modulo NIS2

AGGIORNAMENTO

20/04/2026

Pipeline corpus più robusta: watchdog anti-blocco su tutti gli harvester

Risolto il problema dei siti comunali lenti o non rispondenti che bloccavano il ciclo di harvest per ore. Ora ogni comune o dominio ha 30 minuti per fornire dati, poi viene saltato automaticamente.

AGGIORNAMENTO

20/04/2026

Corpus superato 1,47 milioni di chunk: aggiornamento continuo 24/7

L'indice FAISS di Armandillo ha raggiunto 1.476.882 chunk (+215.000 rispetto al ciclo precedente). La pipeline ora gira in ciclo continuo: appena finisce un giro, ne parte subito un altro.

NUOVA FUNZIONE

20/04/2026

Pubblicato il White Paper Armandillo 2026

La prima AI civica italiana presenta architettura, numeri del corpus (1,47M chunk) e roadmap 2026-2027 in un documento istituzionale scaricabile liberamente.

AGGIORNAMENTO

07/04/2026

Aggiornamento automatico di tutte le VM: 63 macchine gestite da un unico workflow

Armandillo ora aggiorna automaticamente tutte le 63 macchine dell'infrastruttura (51 VM su 3 server Proxmox + VPS OVH) tramite un workflow n8n dedicato, senza intervento manuale.

COMMUNITY

07/04/2026

Ecosistema a 54 progetti: nuovi servizi per la PA e il turismo

L'ecosistema Armandillo ha raggiunto oltre 50 progetti attivi, con l'aggiunta di turismoinformune.it, turismopa.it e altri servizi verticali per la PA italiana.

AGGIORNAMENTO

07/04/2026

Corpus PA superato 1.1 milioni di chunk: conoscenza in crescita costante

L'indice FAISS di Armandillo ha raggiunto 1.120.943 chunk indicizzati, con un nuovo indice in costruzione che includerà ancora più documenti della PA italiana.

NUOVA FUNZIONE

29/03/2026

Armandillo ora cerca sul web: risposte anche su argomenti nuovi

Quando il corpus non ha la risposta, Armandillo cerca automaticamente sul web tra fonti PA ufficiali per fornire informazioni aggiornate.

AGGIORNAMENTO 29/03/2026

Pipeline corpus completata: aggiornamento automatico ogni notte

La pipeline di raccolta dati PA funziona ora in modo completamente automatico ogni notte, con 11 fasi che coprono harvesting, anonimizzazione, indicizzazione e rebuild.

NUOVA FUNZIONE 29/03/2026

25 leggi fondamentali PA nel corpus: 2.302 articoli consultabili

Il corpus di Armandillo ora include il testo integrale di 25 leggi fondamentali per la PA, per un totale di 2.302 articoli indicizzati articolo per articolo.

NUOVA FUNZIONE 29/03/2026

Ricerca ibrida: Armandillo trova la normativa esatta

Nuovo motore RAG v2.2 con ricerca semantica FAISS e keyword FTS5. Armandillo ora cita articoli di legge specifici con il testo originale.

NUOVA FUNZIONE 24/03/2026

Armandillo ora parla: voce italiana con XTTS v2 su GPU

Integrato il motore vocale XTTS v2 su GPU Tesla P100: 6 voci selezionabili per ascoltare le risposte.

NUOVA FUNZIONE 24/03/2026

TUEL, TUPI, CAD e 8 leggi fondamentali ora nel corpus di Armandillo

1.153 articoli delle principali leggi per gli enti locali scaricati e indicizzati articolo per articolo.

AGGIORNAMENTO 24/03/2026

Corpus superato 1 milione di chunk: conoscenza PA senza precedenti

Il motore FAISS ha raggiunto 1.056.414 frammenti indicizzati da 284.000 documenti PA, quasi raddoppiando la base di conoscenza precedente.

AGGIORNAMENTO 21/03/2026

Armandillo ora gira su RTX 3060: risposte piu veloci

Modello migrato su GPU RTX 3060 con context 8192 token

NUOVA FUNZIONE 21/03/2026

Ecosistema Armandillo: 31 progetti online per la PA

31 progetti attivi e interconnessi per la Pubblica Amministrazione

NUOVA FUNZIONE 21/03/2026

Chat Armandillo con accesso unico: un solo login per tutto

chat.armandillo.it ora usa l'autenticazione condivisa dell'ecosistema

NUOVA FUNZIONE 20/03/2026

In arrivo: Armandillo:7b-PA Rev 2.0 con 10.000+ esempi di training

In preparazione la versione 2.0 del modello AI specializzato PA, con dataset ampliato a oltre 10.000 esempi tra atti puliti, normativa vigente e Q&A di alta qualità.

AGGIORNAMENTO 20/03/2026

Corpus PA potenziato: oltre 860.000 frammenti di conoscenza

Il motore di ricerca semantica FAISS è stato ricostruito con oltre 860.000 chunks indicizzati da documenti reali di 7.905 Comuni italiani e fonti ufficiali.

AGGIORNAMENTO 20/03/2026

Qualità risposte migliorata: filtro anti-errori e italiano perfetto

Implementato un sistema di sanitizzazione avanzato che elimina artefatti, caratteri non latini e garantisce risposte sempre in italiano corretto.



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

Il progetto è partito a gennaio 2026 e ha già raggiunto i traguardi della prima metà dell'anno.

Periodo	Obiettivo	Stato
Gennaio 2026	Lancio prima versione di Armandillo, chat AI, corpus di 7.900 comuni indicizzati	FATTO
Febbraio 2026	RAG avanzato, sistema di feedback, upload allegati, pannello admin completo	FATTO
Marzo 2026	Auth ecosistema centralizzato (panel.armandillo.it), DatiPA, TrasparenzaInComune	FATTO
Aprile 2026	Modello proprietario Armandillo:7B-PA in produzione su vLLM (2× RTX 3060), SecureBox Difesa Attiva	FATTO
Maggio 2026	Harvest cortese su VPS dedicato, ponte Normattiva → corpus, indice FAISS > 1,26M chunk	FATTO
Giugno 2026	Bilancio Partecipato, AI per gli Atti (deliberazioni), Whistleblowing (D.Lgs. 24/2023). Modulo NIS2 (Direttiva UE 2022/2555) su HackPA	FATTO
Q3 2026	App mobile PWA (Open WebUI con SSO), self-registration enti su Whistleblowing e AI Atti	IN CORSO
Q3 2026	Ampliamento corpus a province, regioni e città metropolitane (~700 nuovi enti)	previsto
Q4 2026	Fine-tuning v2 su 5.000 nuove Q&A raccolte dall'uso reale, API pubblica per sviluppatori	previsto
Q1 2027	Adesione di almeno 50 comuni pilota come partner, gratuito per fascia < 10k abitanti	previsto
2027	Pubblicazione del dataset Q&A PA con licenza aperta (CC BY-SA 4.0), estensione ad ASL e altri enti	previsto

” 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

centrostudipa.it — Il Primo Osservatorio Nazionale Indipendente sull'IA e sull'Open Source nella PA
Helpdesk: helpdesk.armandillo.it · Documentazione tecnica: armandillo.it/progetto