



IA e Sostenibilità Computazionale

Valutazione comparativa tra Small Language Models, Machine Learning tradizionale e LLM per le
Piccole Medie Imprese e i Professionisti

Roberto Magnani

Asimmetria dei Modelli di Frontiera

Le Barriere d'Accesso

L'ascesa dei Large Language Models (LLM) ha generato pesanti barriere economiche e infrastrutturali per le PMI. Gestire queste architetture richiede costosi cluster cloud centralizzati, determinando una dipendenza strategica ed erodendo la marginalità delle piccole imprese.

L'Impatto Energetico

La gestione di modelli di frontiera può costare oltre 700.000\$ al giorno di sola elettricità. L'addestramento di un singolo grande modello emette CO2 pari all'intero ciclo di vita di 5 autoveicoli, richiedendo un passaggio urgente ad architetture energeticamente sostenibili.

| Il Valore dell'IA per le PMI

Sbloccare la Competitività

L'adozione mirata di intelligenza artificiale nelle PMI italiane promette benefici straordinari:

Esportazioni: Incremento atteso del volume export **+14%**.

Redditività: Incremento dei margini di profitto fino al **+19,5%**.

Verticali d'eccellenza:

Es.

Siderurgia/meccanica (+20 miliardi €)

farmaceutico (+13 miliardi €)

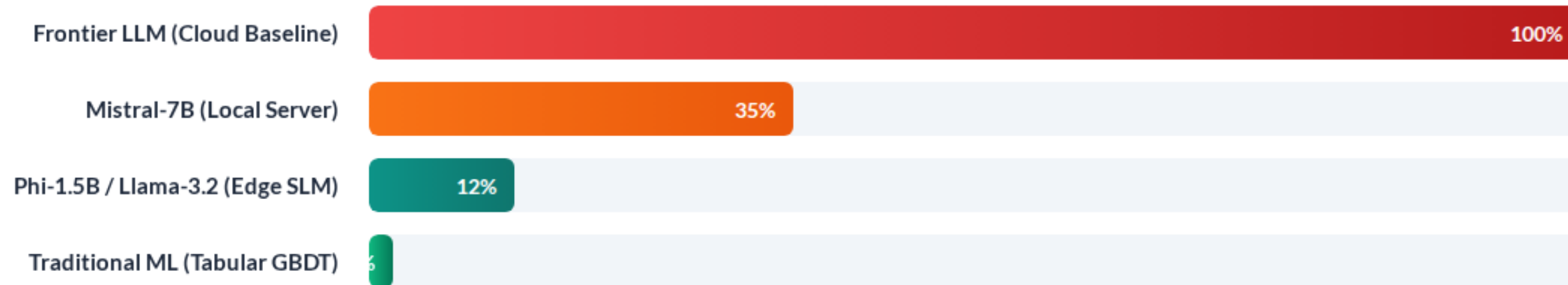
+122B€

Valore Aggiunto Potenziale per le PMI Italiane

Confronto: LLM vs SLM e ML

Parametro	Grandi Modelli (LLM)	Modelli Contenuti (SLM) & ML
Costo di Esercizio	Elevato e variabile (basato su chiamate API esterne)	Prevedibile, eseguibile localmente su hardware standard
Consumo Energetico	Elevato (~0,0022 kWh per query singola)	Ridotto del 60-70% rispetto alle architetture cloud
Investimento Custom	Da decine di migliaia a milioni di dollari	Da poche centinaia a migliaia di dollari (fine-tuning)
Latenza ed Elaborazione	Variabile, vincolata alla connettività internet	Ordine dei millisecondi, ottimizzata on-device (Edge)
Conformità Regulatoria	Massima (audit complessi sotto EU AI Act)	Minima, con totale sovranità e sicurezza sui dati

Efficienza Energetica dei Modelli



I dati empirici mostrano che gli SLM riducono l'impronta energetica fino all'88% rispetto ai LLM in cloud per compiti standard, mentre il Machine Learning classico richiede risorse infinitesime.

Mappatura dei Modelli: SLM-Bench



Llama-3.2-1B

Precisione: valido per accuratezza linguistica formale nel benchmark. Richiede un carico energetico leggermente superiore rispetto alla sua categoria parametrica.



Phi-1.5B

Ecosostenibile: Molto valido per consumi energetici ridotti in fase di inferenza. Assicura efficienza su task mirati.



Qwen2.5-Coder-3B

Codice Sorgente: efficienza nella programmazione. Nel 52% dei casi di successo genera codice più snello ed efficiente dei LLM cloud.



Gemma-3-1B

Bilanciato: Sviluppato da Google, offre un compromesso tra accuratezza formale ed efficienza energetica standard.

| Percezione dell'Utente e Bias

La trasparenza ecologica influisce profondamente sull'interazione uomo-macchina. Esplicitare il costo ambientale tramite sistemi di **Energy Consumption Disclosure (ECD)** genera dinamiche psicosociali cruciali:



Spinta Ecologica: Aumenta di 12 volte la scelta consapevole di SLM locali rispetto ai LLM.



Effetto Placebo Negativo: Gli utenti che sanno di usare un modello "eco" percepiscono una qualità inferiore, sebbene i risultati oggettivi siano equivalenti.



Necessità Formativa: È fondamentale formare il personale per allineare la percezione della qualità all'efficacia reale, non alla scala computazionale.



Sviluppo Software ed “Ensembles”

Sostenibilità del Codice

Il codice generato dai LLM monolitici è spesso strutturalmente ridondante e oneroso da eseguire. SLM come Qwen2.5-Coder generano codice snello, riducendo l'energia necessaria all'esecuzione del software del 52%. Un enorme vantaggio per startup e sistemi IoT a risorse limitate.

Ensemble di Agenti Specializzati

Il superamento dei LLM commerciali si realizza tramite architetture multi-agente basate su "Green AI by Default". Invece di usare un modello gigante, si coordina un gruppo di SLM specializzati su singoli micro-task operativi, eseguibili su workstation aziendali locali.

ML per Dati Tabulari



Alberi Decisionali vs Deep Learning: I modelli a gradiente incrementale (GBDT) come XGBoost e LightGBM superano sistematicamente le reti neurali profonde nell'analisi dei database aziendali reali, con significatività statistica pari a .
 $p < 0.001$



Adattamento a Funzioni Discontinue: I dati di vendita e contabilità presentano forti discontinuità spaziali. Il Deep Learning impone soluzioni eccessivamente "smussate", mentre gli alberi decisionali catturano i confini precisi del business.



Immunità al Rumore e Dati Categoricali: I modelli GBDT effettuano una selezione intrinseca delle variabili più importanti, mantenendo la precisione predittiva inalterata anche rimuovendo fino al 50% delle colonne irrilevanti.



Distillazione con M-RARU: Consente di distillare la conoscenza di un modello complesso "docente" trasferendola a un Random Forest "studente" ad altissima interpretabilità, azzerando i costi di calcolo.

| Manutenzione Predittiva

L'applicazione del Machine Learning tradizionale a dati di sensori fisici consente l'edge computing ad alta efficienza direttamente sui PLC degli impianti.

Risposte in tempo reale nell'ordine dei millisecondi, elaborazione locale senza dipendenza dal cloud e piena tracciabilità logica tramite SHAP/LIME rendono la manutenzione predittiva accessibile, sostenibile ed estremamente precisa per le PMI manifatturiere.



Conformità all'EU AI Act

Tipologia Modello	Soglia di Calcolo	Classificazione del Rischio	Implicazioni e Oneri per PMI
Frontier LLM	$> 10^{25}$ FLOP	Sistemico	Obbligo di red-teaming, reporting degli incidenti gravi e audit costanti.
Standard GPAI	$> 10^{23}$ FLOP	Standard GPAI	Forte rischio di responsabilità downstream in caso di personalizzazione commerciale.
SLM Open-Source	$< 10^{23}$ FLOP	Minimo	Esenzione dagli oneri burocratici complessi se i pesi sono aperti e modificabili.
Machine Learning	Non Applicabile	Nulla / Minimo	Nessun costo di conformità legale. Massima flessibilità operativa.

Roadmap Strategica per PMI – esempio didattico



TAKE AWAY

**La situazione geopolitica induce a una riflessione -
breve analisi strategica dei sistemi di intelligenza
artificiale per professionisti e PMI**



Il paradosso dell'IA

Articolo LinkedIn

<https://www.linkedin.com/pulse/la-situazione-geopolitica-induce-una-riflessione-breve-magnani-xffrf>

Vademecum per La Prontezza all'intelligenza Artificiale

Link commerciale al Vademecum

<https://ethosjob.it/editoria/prontezza-allintelligenza-artificiale-guida-pratica-in-5-punti-per-professionisti-e-piccole-medie-imprese/>



Domande e Risposte

Grazie per l'attenzione. La transizione verso l'IA sostenibile è la chiave della sovranità digitale.

|