

EdicolaGiuridica — Specifiche Migliorate (To-Be)

Versione proposta: 2026-03-27 · Target: scalabile, mantenibile da team umano + LLM, monetizzabile premium

1) Visione prodotto

Mission: trasformare l'informazione giuridica in decision intelligence affidabile e verificabile.

Posizionamento: piattaforma B2B/B2Pro per studi legali, uffici compliance, PA, aziende regolamentate.

2) Architettura target

- **Frontend:** Next.js/SSR ibrido, design system, route-level code splitting.
- **API gateway:** auth, RBAC, rate limit, WAF, audit.
- **Servizi separati:** ingest, enrichment, retrieval, synthesis, billing/tenanting.
- **Event bus + queue:** job asincroni per L3-L5 e task pesanti.
- **Storage:**
 - OLTP PostgreSQL (tenant-aware)
 - Object store per snapshot/documenti
 - Vector DB per RAG (metadata filtering serio)
 - Redis per cache e rate control
- **Observability:** OpenTelemetry, logs strutturati, cost telemetry LLM.

3) Requisiti di sicurezza (mandatory)

- Auth OIDC/JWT + refresh + MFA opzionale enterprise.
- RBAC per ruoli (viewer/analyst/admin) + ABAC per tenant/data scope.
- Rate limit multilivello (IP, API key, tenant, endpoint).
- Security headers completi (CSP, HSTS, frame-ancestors, Referrer-Policy).
- CORS a allowlist esplicita.
- Schema validation centralizzata, error catalog versionato.
- Audit trail immutabile delle azioni AI (prompt, fonti usate, output, versione modello).

4) RAG premium (plus rivendibili)

- Premium
- Enterprise
- Differenziazione

1. **RAG giuridico controllato:** retrieval su corpus normativo/giurisprudenziale curato + filtri per data vigenza, giurisdizione, sezione, materia.
2. **Citations-first answers:** ogni insight con citazioni puntuali e confidence score.
3. **What changed?:** diff semantico tra orientamenti/periodi/fonti.
4. **Tenant private knowledge:** knowledge base privata cliente (atti interni, policy, memo) con isolamento forte.
5. **Playbook operativi:** suggerimenti azionabili legati a ruoli (compliance officer, legale interno, contenzioso).
6. **Alert predittivi:** early-warning su norme/sentenze ad alto impatto per settore.

5) “Scala dell’informazione”: tenerla o cambiarla?

La scala L0-L5 è **utile come UX mentale** e storytelling commerciale. Va mantenuta, ma con metriche oggettive a supporto.

Metriche da affiancare (più robuste)

Metrica	Uso	Valore
---------	-----	--------

Source Reliability Score	Affidabilità fonte/cluster	Riduce bias e rumore
Normative Relevance	Pertinenza normativa per caso	Selezione priorità
Jurisprudential Stability Index	Stabilità o conflitto orientamenti	Misura rischio interpretativo
Operational Impact Score	Impatto su processi cliente	Drive azioni concrete
Freshness / Half-life	Obsolescenza insight	Evita decisioni su dati vecchi
Explainability Coverage	% output con citazioni verificabili	Trust e compliance

6) Manutenibilità umana + LLM

- Domain contracts stabili (OpenAPI + versioning semantico).
- Codice modulare per bounded context (news, codici, cassazione, rag-core).
- Test pyramid: unit + contract + e2e + eval LLM regressions.
- Playbook operativi in repo (runbook incident, cost burst, feed outage).
- Prompt/version governance: ogni change tracciata, A/B testabile, rollback facile.

7) Roadmap proposta

Fase A (0-30 giorni): Hardening

- Auth + rate limit + CORS + headers + input validation.
- Metriche base, alert costi LLM, budget guardrails.

Fase B (30-90 giorni): Scale core

- Queue asincrona, caching, separazione servizi, tenancy base.
- RAG v1 con citazioni verificabili.

Fase C (90-180 giorni): Premium

- Alert predittivi, playbook verticali, explainability dashboard, SLA enterprise.

8) Conclusione To-Be

Il concept è buono e vendibile. La chiave non è “più AI”, ma **AI governata**: affidabilità, tracciabilità, isolamento tenant e controllo costo. La scala L0-L5 può restare il front-end concettuale, mentre dietro servono metriche quantitative e pipeline industriale.