



AUTORITÀ PER LE
GARANZIE NELLE
AGCOM COMUNICAZIONI

Oggetto: Avviso pubblico "*Manifestazione di interesse ai sensi dell'art. 36, comma 2, d. lgs. n. 50/2016 per l'acquisizione di un servizio di tipo big data consistente nella messa a disposizione di un vasto database documentale di notizie (prodotte da varie fonti di informazione), per consentire all'Autorità di effettuare analisi dei dati sul sistema dell'informazione in Italia, attraverso l'elaborazione di serie storiche di occorrenze di parole chiavi e l'uso in-house di tecniche di NLP (Natural Language Processing) applicate al testo di milioni di notizie afferenti ad un arco temporale esteso, che viene tenuto sempre online.*"

Risposte a richieste di chiarimenti

D.1) Sono già disponibili i PDF delle testate o è a cura del fornitore procurarsi questi dati. Inoltre sono già presenti accordi riguardo l'utilizzo dei dati provenienti dalle testate giornalistiche in ambito ML da parte di Agcom o è il fornitore che deve stabilire accordi con le singole testate riguardanti il materiale protetto da diritto d'autore?

R.1) Il servizio oggetto della manifestazione di interesse prevede che sia cura del Fornitore stabilire accordi con le singole testate riguardanti il materiale protetto da diritto d'autore. Si precisa, inoltre, che i dati saranno utilizzati dall'Autorità esclusivamente a fini istituzionali, con particolare riguardo alle attività di monitoraggio e analisi del sistema dell'informazione, pertanto, senza alcuno scopo di natura commerciale.

D.2) Per quanto riguarda i Telegiornali (incluse le edizioni dei singoli TGR), altri programmi televisivi di informazione Giornalradio e altri programmi radiofonici di informazione, il fornitore deve rendere disponibile una trascrizione dei singoli TG tramite algoritmi di Speech to Text?

R.2) Il servizio richiesto prevede che il Fornitore renda disponibile una trascrizione dei flussi televisivi e radiofonici realizzabile tramite algoritmi di Speech to Text. Si precisa, inoltre, che le modalità di acquisizione dei contenuti che alimenteranno il database documentale sono a cura del Fornitore.

D.3) Siti web di Tv, Siti web di radio: che tipo di informazioni interessano in questo caso, dal momento che questi siti vengono già usati per reperire i tg, radio, etc descritti al punto precedente?

R.3) Per quanto riguarda i Siti web di Tv e Siti web di radio, ci si riferisce in particolare agli articoli testuali ivi pubblicati, laddove presenti.

D.4) Per quanto riguarda invece il Server che deve essere "messo a disposizione dal Fornitore", si fa riferimento a un server fisico da installare in sede o a un'altra soluzione?

R.4) Si precisa che Agcom intende disporre di un servizio di tipo cloud.