

zendata

Sommario

1. Introduzione al progetto	4
1.1 Panoramica	4
1.2 Funzionamento del tool	4
1.3 Fasi di progetto	5
Fase 1: Pilota	5
Fase 2: Tool in abbonamento (SaaS)	5
2. Architettura del tool	6
2.1 Panoramica dell'Architettura	6
Layer di Ingestion	6
Layer di Storage	6
Layer Applicativo	7
2.2 Autenticazione e Identità	8
2.3 Layer di Ingestion	8
2.4 Layer di storage	9
Azure Cosmos DB for MongoDB	10
Azure Blob Storage	10
2.5 Layer di AI	11
2.6 Layer Applicativo	12
3. Architettura della Dashboard	13
3.1 Panoramica dell'Architettura	13
3.2 Autenticazione e Controllo Accessi	14
3.3 Infrastruttura di Rete	14
3.4 Raccolta e Analisi delle Metriche	15

Chi siamo

ZenData è una startup dinamica e innovativa, specializzata nello sviluppo di soluzioni basate sull'intelligenza artificiale, per aumentare l'efficienza e stimolare la crescita.

Le nostre soluzioni abbracciano una vasta gamma di settori e reparti, spaziando dallo sviluppo di assistente virtuale per pubbliche amministrazioni e professionisti autonomi, fino all'analisi e all'ottimizzazione dei processi produttivi aziendali.

La nostra strategia è orientata alla progettazione di soluzioni che si integrano in maniera semplice e fluida nei processi operativi dei nostri clienti, garantendo un'adozione senza attriti e un impatto positivo immediato.

Siamo convinti che il nostro approccio all'innovazione, unito alla capacità di personalizzare le soluzioni in base alle specifiche esigenze dei nostri clienti, ci permetta di offrire un valore aggiunto significativo e di contribuire al successo dei nostri partner commerciali.



1. Introduzione al progetto

1.1 Panoramica

Il presente progetto prevede la realizzazione di un **tool di Intelligenza Artificiale** in grado di trasformare in modo automatizzato i contenuti formativi di Bureau Veritas, a partire da slide esistenti, in:

- un **copy testuale narrativo** coerente,
- un **video formativo completo**,
- e una serie di **test di valutazione dell'apprendimento**, con funzionalità di correzione e supporto alla valutazione.

L'obiettivo è semplificare e accelerare la produzione di corsi online, consentendo a Bureau Veritas di valorizzare il proprio patrimonio formativo con strumenti digitali evoluti e personalizzabili.

1.2 Funzionamento del tool

A partire dalle **slide fornite come input**, il tool esegue:

1. la generazione automatica di un **testo narrativo** (copy) ottimizzato per la fruizione video;
2. la creazione di un **video formativo** sincronizzato con il copy.
3. la produzione di **test/esercitazioni di verifica dell'apprendimento**, con:
 - domande a risposta chiusa e aperta;
 - funzionalità di correzione automatica ove possibile;
 - supporto alla valutazione manuale da parte del docente.

A completamento della soluzione, verrà inoltre realizzata una **Dashboard amministrativa** che consentirà agli utenti autorizzati di:

- gestire in modo efficiente i contenuti caricati (slide) e i materiali generati (copy, video, test);
- configurare i parametri di personalizzazione del processo di generazione;

- monitorare le performance del tool e l'utilizzo del sistema;
- analizzare i risultati degli studenti sui test generati.

La Dashboard sarà integrata nativamente nella piattaforma, garantendo coerenza dei dati e un'esperienza utente fluida.

1.3 Fasi di progetto

Fase 1: Pilota

Nella prima fase pilota, il progetto prevede la trasformazione di **2 corsi esistenti** di 8 ore ciascuno in **video corsi generati dall'AI** della durata di circa 5 ore ciascuno, corredati da **2-3 esercitazioni per corso** con relativa valutazione.

I contenuti saranno forniti a Bureau Veritas come **video finito** completo, con la possibilità di richiedere eventuali aggiustamenti su:

- lunghezza complessiva,
- tono di voce,
- stile narrativo.

Fase 2: Tool in abbonamento (SaaS)

In una seconda fase, il progetto prevede l'evoluzione in un **tool ad abbonamento mensile**, accessibile tramite interfaccia grafica (GUI).

In questa versione, l'utente potrà:

- caricare in autonomia le **slide di partenza**;
- generare il **copy** e successivamente il **video** e i relativi **test di valutazione**;
- personalizzare diversi parametri:
 - lunghezza del contenuto,
 - stile narrativo,
 - tone of voice,

- elementi visivi e formattazione;
- effettuare più versioni di prova e selezionare la versione finale più adatta alle proprie esigenze.

In questa seconda fase, il flusso operativo previsto seguirà le seguenti fasi principali:

- **Generazione dei contenuti formativi:** il tool consentirà la creazione automatica di contenuti formativi completi a partire dalle slide caricate dall'utente.
- **Esportazione dei contenuti in formato SCORM:** i contenuti generati verranno esportati in formato SCORM, così da poter essere caricati e fruiti all'interno della piattaforma di e-learning di Bureau Veritas.
- **Raccolta e valutazione dei test:** i risultati delle prove di valutazione svolte dagli studenti saranno quindi importabili nel modulo di valutazione AI incluso nel tool fornito.

Il tool sarà quindi composto da due moduli principali, integrati in un'unica soluzione:

- **Modulo di creazione dei contenuti:** consente di generare in modo automatizzato contenuti formativi completi a partire dalle slide di partenza, includendo copy testuale, video formativo e test di valutazione.
- **Modulo di valutazione:** permette di raccogliere e analizzare i risultati dei test svolti dagli studenti, utilizzando tecniche di intelligenza artificiale per supportare la valutazione e fornire insight utili all'ottimizzazione dei percorsi formativi.

2. Architettura del tool

La seguente sezione descrive l'architettura prevista per la realizzazione del tool su piattaforma cloud. L'infrastruttura sarà implementata su Microsoft Azure, garantendo scalabilità, sicurezza e flessibilità. L'adozione di una piattaforma cloud permetterà di ottimizzare le prestazioni, semplificare la gestione e garantire un'evoluzione continua della soluzione senza impatti sull'operatività.

2.1 Panoramica dell'Architettura

L'architettura si compone sostanzialmente di 4 layer funzionali:

Layer di Ingestion

- Gestisce il caricamento delle slide di partenza da parte dell'utente.
- Include componenti per:
 - estrazione automatica dei contenuti (testi, immagini) dalle slide caricate;
 - normalizzazione e preparazione del contenuto per la successiva generazione del copy.

Layer di Storage

- Supporta la gestione dei dati e dei contenuti generati. Comprende:
 - un database documentale (NoSQL) (es. Azure Cosmos DB for MongoDB) per:
 - gestire i contenuti delle slide caricate;
 - memorizzare le versioni del copy generato;
 - gestire i parametri di configurazione utente (tone of voice, lunghezza, stile, ecc.);
 - uno storage multimediale (Azure Blob Storage) per la gestione dei file video prodotti e degli asset correlati.

Layer di AI

- È il cuore del sistema e comprende:
 - modelli di NLP / LLM per:

- generare il copy testuale a partire dai contenuti estratti;
- ottimizzare e adattare il copy secondo i parametri scelti dall'utente (tone of voice, stile narrativo, lunghezza);
- modelli per la generazione video:
 - sintesi vocale (Text-to-Speech) con possibilità di personalizzazione della voce;
 - orchestrazione della creazione del video sincronizzato (immagini + narrazione);
- modelli per la generazione di test di valutazione:
 - domande a risposta chiusa (multiple choice)
 - domande a risposta aperta
 - supporto alla correzione automatica e al feedback.

Layer Applicativo

- Comprende le risorse cloud per la realizzazione delle componenti applicative:
 - Frontend: interfaccia utente (GUI) per:
 - caricamento slide;
 - configurazione dei parametri di generazione (copy, video, test);
 - gestione delle versioni generate;
 - download dei materiali finali.
 - Backend: orchestrazione dei flussi di lavoro AI, integrazione con i modelli e con il layer di storage.

2.2 Autenticazione e Identità

Per garantire la protezione dei dati e il controllo degli accessi, il sistema adotterà **Azure Active Directory (Azure AD)** per la gestione centralizzata delle identità e l'autenticazione sicura alle risorse Azure.

L'uso delle **Managed Identities** permette ai servizi Azure di autenticarsi in modo sicuro senza necessità di gestire manualmente credenziali statiche. In particolare, l'**App Service** utilizzerà una **System-Assigned Managed Identity** per ottenere token da Azure AD e

accedere in modo sicuro a **Key Vault**, **Cosmos DB** e al **database vettoriale**, sfruttando il **Role-Based Access Control (RBAC)** e politiche di accesso dedicate.

Per rafforzare la sicurezza della rete e ridurre la superficie di attacco, il sistema sarà isolato all'interno di una **Virtual Network (VNet) dedicata**. Questa configurazione permette di limitare l'accesso ai soli sistemi autorizzati e instradare il traffico attraverso connessioni private, evitando esposizioni pubbliche. Sarà prevista l'integrazione con la stessa Vnet per tool e dashboard.

L'architettura prevede:

- **VNet dedicata**
- **Integrazione dell'App Service con la VNet**, per connettersi internamente a **Cosmos DB**, il **database vettoriale** e **Key Vault** tramite IP privati.
- **Private endpoint per Cosmos DB e il database vettoriale**, assegnati a una subnet dedicata per garantire che il traffico resti confinato all'interno della VNet, senza esposizione pubblica.
- **Private endpoint per Key Vault**, che limita l'accesso solo ai servizi interni alla rete privata.

Questa strategia garantisce un'elevata sicurezza dell'infrastruttura, proteggendo i dati e minimizzando i rischi legati agli accessi non autorizzati.

2.3 Layer di Ingestion

Il **Layer di Ingestion** si compone di tutte le risorse software necessarie per permettere il caricamento e la preparazione dei materiali di partenza forniti dagli utenti (tipicamente **slide in formato PowerPoint o PDF**) per l'elaborazione da parte del tool.

Questo layer prevede l'utilizzo di servizi serverless (es. **Azure Functions**) e componenti di orchestrazione per automatizzare il processo di estrazione e preparazione dei contenuti.

In particolare, queste funzioni si occuperanno di:

- **Estrazione dei contenuti dalle slide:**
 - Parsing e conversione dei file (PowerPoint, PDF) in un formato intermedio strutturato (JSON, HTML, o equivalente).
 - Estrazione dei testi, dei metadati e delle immagini rilevanti.

- **Preprocessing dei contenuti:**
 - Pulizia e normalizzazione dei testi per rimuovere elementi non rilevanti.
 - Organizzazione dei contenuti in una struttura coerente per l'elaborazione successiva da parte del layer di AI (per la generazione di copy testuale e dei test).
 - Suddivisione in segmenti logici per facilitare la generazione modulare dei contenuti (es. segmenti corrispondenti a sezioni o capitoli del corso).
- **Calcolo di vettori di embedding:**
 - Calcolo degli embedding testuali dei contenuti estratti, da utilizzare per:
 - eventuale ricerca semantica nei contenuti caricati.
 - tracciamento delle versioni e confronto dei contenuti generati.
- **Storage dei contenuti preprocessati:**
 - Caricamento dei contenuti estratti e normalizzati all'interno del **database NoSQL** per la gestione operativa del tool.
 - Caricamento dei file originali e dei contenuti preprocessati all'interno dello **Azure Blob Storage**, per consentire il versioning e l'accesso ai materiali da parte dei moduli successivi (AI e Video Generation).

2.4 Layer di storage

Il **Layer di Storage** del sistema è progettato per gestire in modo scalabile e sicuro:

- i contenuti caricati dagli utenti (es. slide di partenza);
- i contenuti preprocessati ed elaborati dal tool;
- i contenuti generati (copy testuale, esercitazioni, video finali);
- i dati di configurazione e di utilizzo del sistema.

L'architettura di storage prevede i seguenti componenti principali:

Azure Cosmos DB for MongoDB

Utilizzato come **database NoSQL scalabile**, ha le seguenti funzioni:

- **Gestione dei contenuti preprocessati:**
 - testi ed eventuali immagini estratti dalle slide caricate;

- metadati associati (struttura logica, sezioni, titoli, timestamp di caricamento).
- **Gestione dei parametri di configurazione utente:**
 - tono di voce, lunghezza dei contenuti, stile narrativo, preferenze di generazione.
- **Gestione delle versioni dei contenuti generati:**
 - copy testuale prodotto;
 - test di valutazione generati (domande e risposte);
 - metadati relativi ai video (versioni generate, timestamp, parametri utilizzati).
- **Supporto alle funzionalità di reporting e analytics:**
 - tracciamento delle attività dell'utente all'interno del sistema;
 - tracciamento delle versioni e del ciclo di generazione dei materiali formativi.

Azure Blob Storage

Utilizzato come **storage multimediale** per:

- **Conservazione dei file originali caricati:**
 - slide in formato PowerPoint, PDF o altri formati supportati.
- **Conservazione dei contenuti generati:**
 - video formativi completi (versioni finali e intermedie);
 - eventuali asset multimediali associati (immagini, tracce audio, materiali accessori).
- **Gestione sicura dei file:**
 - versioning automatico dei contenuti;
 - accesso controllato e sicuro tramite integrazione con rete privata e Private Endpoint.

Sicurezza dei dati

Entrambi i componenti di storage garantiranno l'accesso sicuro ai dati attraverso:

- configurazione su **rete privata (Virtual Network)**;
- utilizzo di **Private Endpoint** per isolare i flussi di dati da Internet pubblico;
- applicazione delle best practice di sicurezza cloud.

2.5 Layer di AI

Il **Layer di AI** è il cuore del sistema e comprende i modelli e i servizi di intelligenza artificiale che abilitano le funzionalità di generazione automatica dei contenuti formativi.

In particolare, questo layer comprende:

- **Motore di generazione copy:**
 - utilizzo di **Large Language Models (LLM)** per generare un copy testuale coerente, fluido e didatticamente efficace a partire dai contenuti preprocessati delle slide;
 - possibilità di personalizzare il copy in funzione dei parametri configurati dall'utente (fase 2).
- **Motore di generazione video:**
 - integrazione di componenti di **sintesi vocale (Text-to-Speech)** per la creazione della narrazione audio sincronizzata con il copy generato;
 - orchestrazione della generazione video combinando il copy, la traccia audio e i contenuti visivi (slide, immagini) in un video formativo completo.
- **Motore di generazione test di valutazione:**
 - generazione automatica di **domande a risposta chiusa** e **domande a risposta aperta** a partire dal copy generato e dai contenuti delle slide;
 - supporto alla correzione automatica delle domande a risposta chiusa e semiautomatica per quelle a risposta aperta.
- **Motore di embedding:**
 - calcolo degli embedding testuali per consentire eventuali funzionalità avanzate di ricerca semantica nei contenuti caricati e/o generati.

Lo stack tecnologico di questo layer si basa sui **foundational models LLM disponibili su Azure** e su servizi di Text-to-Speech di nuova generazione integrati nella piattaforma.

2.6 Layer Applicativo

Il **Layer Applicativo** fornisce l'infrastruttura per la gestione della logica del sistema e per l'interazione utente.

Il sistema sarà distribuito attraverso un **backend centralizzato** e un'interfaccia utente web.

- **Backend applicativo:**
 - verrà ospitato su **Azure App Service**, eseguito all'interno di container Docker per garantire scalabilità, isolamento e semplicità di gestione;
 - il backend si occuperà di orchestrare i flussi di generazione dei contenuti (copy, video, test), di gestire l'interazione con il Layer di AI e il Layer di Storage e di fornire le API per il frontend.
- **Frontend applicativo:**
 - interfaccia utente (GUI) distribuita come applicazione web a pagina singola (SPA), con hosting statico su **Azure Blob Storage**;
 - consente all'utente di caricare le slide, configurare i parametri di generazione, monitorare l'avanzamento dei processi e scaricare i materiali generati.
- **Sicurezza e integrazione:**
 - l'accesso tra i layer avverrà tramite **Virtual Network (VNet)** e **Private Endpoints**, garantendo connessioni sicure su rete privata;
 - le chiavi di accesso e i segreti saranno gestiti tramite **Azure Key Vault** e **Managed Identity**, evitando l'uso di credenziali esplicite.
- **Monitoring e observability:**
 - verrà adottato **Azure Application Insights** per il monitoraggio dello stato dell'applicazione, delle performance e dell'utilizzo.

3. Architettura della Dashboard

La seguente sezione descrive l'architettura prevista per la messa in produzione della **Dashboard amministrativa**.

L'infrastruttura sarà implementata su **Microsoft Azure**, garantendo perfetta integrazione con il tool descritto precedentemente.

L'architettura della Dashboard è stata progettata per offrire un'interfaccia amministrativa completa che consenta:

- la **gestione efficiente dei contenuti formativi**,
- la **configurazione dei parametri di generazione**,
- il **monitoraggio delle performance del tool** e
- il **monitoraggio dei risultati degli studenti**.

3.1 Panoramica dell'Architettura

L'architettura della Dashboard si compone di quattro componenti principali:

1. Frontend Amministrativo:

-Offre funzionalità di:

- gestione dei contenuti formativi caricati (slide)
- visualizzazione e gestione dei materiali generati (copy, video, test)
- configurazione dei parametri di generazione
- visualizzazione di analytics e metriche

Design responsive ottimizzato per diversi dispositivi.

2. Backend API:

- Gestisce le richieste dal front end e coordina le operazioni sui dati

3. Storage condiviso:

- Utilizzo delle stesse risorse di storage del tool precedentemente descritto
- Include Cosmos DB, database vettoriale e Blob Storage
- Garantisce coerenza dei dati e riduce la duplicazione

Funzionalità principali supportate:

- Gestione completa del ciclo di vita delle slide (caricamento, elaborazione, eliminazione)
- Configurazione dei parametri di personalizzazione (copy, video, test)
- Monitoraggio delle performance del tool (metriche e analytics di utilizzo)
- Monitoraggio delle performance degli studenti (risultati dei test)
- Amministrazione sicura con autenticazione **Azure Active Directory (Azure AD)**.

3.2 Autenticazione e Controllo Accessi

Per garantire la protezione dei dati e il controllo degli accessi, il sistema adotterà le stesse politiche di sicurezza descritte nell'architettura del tool. Azure Active Directory (Azure AD) viene utilizzato per autenticare gli amministratori e i servizi che accedono alla dashboard, garantendo un accesso centralizzato e sicuro. L'uso delle Managed Identities consente all'applicazione di autenticarsi automaticamente con i servizi Azure senza necessità di gestire manualmente credenziali.

3.3 Infrastruttura di Rete

Verrà utilizzata una subnet dedicata per i componenti della Dashboard all'interno della VNet esistente. Verranno utilizzati gli stessi private endpoint configurati per il tool,, assicurando che il traffico resti confinato all'interno della rete privata.

3.4 Raccolta e Analisi delle Metriche

La Dashboard prevede la raccolta di **metriche** utili a:

- monitorare le performance del sistema e dei processi di generazione;
- individuare trend di utilizzo e ottimizzare il servizio nel tempo;
- analizzare i risultati ottenuti dagli studenti sui test generati.

Le metriche definitive potranno essere accordate in dettaglio in una seconda fase. I dati raccolti verranno resi disponibili attraverso dashboard interattive.

Voce	gg	€ / gg	Subtotale
Senior Frontend Dev	10	380	3 800
AI Engineer (backend + AI)	24	330	7 920
Project Manager	4	500	2 000
QA/DevOps (assorbito sull'AI Eng.)	3	330	990
Crediti Azure & API AI	—	—	500
Contingenza 10 %	—	—	1 421
Totale listino			15 631 €
Sconto fascia 15-30 k → 10 %			−1 563
Totale offerto PILOTA			≈ 14 100 €

Voce	gg	€ / gg	Subtotale
Senior Frontend Dev	45	380	17 100
AI Engineer (backend + AI)	65	330	21 450
Project Manager / Solution Architect	20	500	10 000
DevOps / QA hardening	10	330	3 300
Setup infra & licenze	—	—	1 500
Contingenza 15 %	—	—	8 202
Totale listino			61 552 €
Sconto fascia > 60 k → 20 %			-12 310
Totale offerto PROD			≈ 49 200 €

Voce	€/mese
Azure compute + storage (carico medio)	800
API LLM / Speech (≈ 10 k slide)	400
Manutenzione evolutiva & supporto (15 h)	660
Canone SaaS consigliato	2 500 €/mese

zendata



info@zendata.it

www.zendata.it

Via Nairobi 40
00144, Roma (RM)
info@zendata.it

zendata

WWW.ZENDATA.IT