



Università degli Studi di Padova

Laurea: Informatica

Corso: Ingegneria del Software

Anno Accademico: 2025/2026



**RUBBER
DUCK**

Gruppo RubberDuck

email: GroupRubberDuck@gmail.com

Verbale Riunione con Eggon

2025-10-30

Vers.	Data	Autore	Verificatore	Descrizione
0.1.0	2025-10-30	Filippo Guerra	Davide Testolin	Prima stesura

Indice

Informazioni generali	1
Ordine del giorno	2
Domande	3
Strategia di Implementazione AI e Fine Tuning	3
Formazione Tecnica e Risorse per il Team	3
Ambiente di Testing	3
Qualità e Testing	4
Esperienza Utente	4
Gestione degli errori di connessione	4
Riconoscimento dei documenti	5
Privacy e Sicurezza Dati	5
Automazione e Intervento Umano	5
Supporto e Allineamento Progetto	6
Repository Documentale	6
Integrazione con altri sistemi	6

Informazioni generali

- **Tipo di riunione** : Esterno
- **data**: 2025-10-30
- **luogo**: Riunione su Meet
- **Ora inizio**: 15:00
- **Ora fine**: 15:23
- **Scriba**: Filippo Guerra
- **Partecipanti**:
 - Filippo Guerra
 - Davide Testolin
 - Aldo Bettega
 - Davide Lorenzon
 - Gianpaolo Ferrarin (Eggon)
 - Luca Luzzolino (Eggon)

Ordine del giorno

- Presentazione e analisi con l'azienda Eggon delle domande elaborate dal gruppo durante la fase preparatoria.

Domande

Le domande poste sono state le seguenti:

Strategia di Implementazione AI e Fine Tuning

Nella pratica come avviene il fine tuning dell'AI Assistant Generativo? In che modo andremo ad orientare questo LLM a fornire stili/risposte il più accurate possibili? Quanto andremo realmente a toccare direttamente la parte di AI e in che modo?

Risposta 1

La strategia per il fine tuning dell'AI si basa su una scelta bilanciata di modelli LLM tramite AWS Bedrock, focalizzata su costi e performance. Il team userà AWS Bedrock per il fine tuning degli LLM, scegliendo il modello più adatto bilanciando performance e costi, con un esempio di costo di 300 mila euro per mezzo token su Cloud 4.5. Sarà necessaria un'analisi approfondita da parte del team incaricato per definire modalità e strumenti di tuning. L'approccio mira a mantenere alta accuratezza e sicurezza nella generazione dei contenuti. La scelta del modello iniziale è cruciale per un compromesso efficace tra costo e qualità. Il tuning sarà gestito dal team con supporto AWS, mentre la parte tecnica e operativa sarà oggetto di analisi e sperimentazione iniziale. La libertà di scelta dei modelli consente di sperimentare diverse soluzioni. L'obiettivo è integrare progressivamente l'AI nel sistema esistente tramite Proof of Concept (POC). L'approccio prevede un percorso di apprendimento e verifica prima dello sviluppo completo. Il tuning dinamico e statico sarà parte integrante per migliorare la qualità.

Formazione Tecnica e Risorse per il Team

Nella parte delle tecnologie siete stati molto precisi e rigorosi, ci sono delle risorse/consigli in particolare per apprendere l'utilizzo di questi strumenti?

Risposta 2

La formazione tecnica sarà basata su risorse online e corsi mirati per preparare il team all'uso di Angular, Ruby-on-Rails e stack backend, con supporto formale da parte di Nexum in corsi e documentazione. Non ci sono risorse interne specifiche, ma si può fornire materiale esterno come corsi Udemy. La formazione è pensata per essere flessibile e adattarsi alle competenze del team.

Ambiente di Testing

Viene detto che avremo un ambiente test a disposizione, da che cosa è costituito?

Risposta 3

Immagino un ambiente su cui poter deployare e tutto quello che vi serve per poter accedere alle funzionalità di AWS. Quindi un'organization in AWS all'interno della quale poter andare a sperimentare gli strumenti che ci sono a disposizione per quanto riguarda la parte di AI, con API, KLEE, Token, tutto quello che vi serve per.

Qualità e Testing

Viene detto che la qualità percepita deve essere $\geq \frac{4}{5}$ nel rating, come viene misurato questo parametro?

Risposta 4

La qualità percepita del sistema sarà misurata con rating utenti, e la fase di testing iniziale sarà interna per garantire il raggiungimento di standard elevati. Si mira a un rating di qualità superiore a 4 su 5, con raccolta feedback dagli utenti su ogni output e testing interno per validare i risultati dei prompt generati. La valutazione_G è un requisito_G in fase di studio e potrà evolvere. L'obiettivo è garantire contenuti affidabili e pertinenti prima del rilascio.

Esperienza Utente

A livello pratico, l'AI lavora dietro le quinte e comunica con l'app dei dipendenti tramite API, oppure gli utenti vedranno anche funzioni nuove nell'interfaccia_G (per esempio suggerimenti, alert o stati automatici: documento validato, in revisione, inviato...)? Si prevede la creazione di nuove sezioni anche nell'app dei dipendenti, come ad esempio una sezione "Notifiche" o "Documenti"? Avranno un modulo storico dove il dipendente può consultare tutti i documenti passati? Sarà possibile scegliere quali notifiche ricevere e in che modalità (app, email, push) oppure le impostazioni saranno predefinite?

Risposta 5

Farà tutto parte di analisi successive. In questo momento abbiamo pensato di concentrare l'analisi sulle funzionalità di assistenza diretta al consulente del lavoro. Contenuti che verranno generati a seguito dell'inserimento di documenti, dello split per esempio delle buste paga, verranno veicolati attraverso il canale di comunicazione convenzionale di Netcom, che sono i messaggi che vengono mandati dal server alle app. Per cui in questo momento nell'app non abbiamo previsto sezioni separate e nemmeno la possibilità gestione diversa da questa. Ci sono dei filtri di ricerca che permettono comunque all'utente di trovare per esempio messaggi che contengono allegati e in questo momento pensiamo di restare molto leggeri su questa parte ma di concentrarci di più sulla componente appunto di interazione con il consulente del lavoro.

Gestione degli errori di connessione

In caso di rete non disponibile o connessione instabile, è previsto un meccanismo di coda o buffer locale per garantire la consegna e l'elaborazione successiva?

Risposta 6

Al momento non è previsto in quanto il meccanismo è un'interfaccia_G web lato consulente del lavoro. Se il consulente non ha la connettività non raggiunge l'interfaccia_G web e quindi di conseguenza non può operare sulla piattaforma. Se cade la connessione durante l'attività la comunicazione si interrompe e non esiste un meccanismo offline sulla dashboard_G.

Riconoscimento dei documenti

Quando il Co-Pilot riconosce documenti (cedolini, CU, ecc.), come distingue le categorie? Utilizza regole testuali (keyword OCR), layout analysis o un modello ML addestrato? Può apprendere nuovi formati aziendali nel tempo?

Risposta 7

Noi abbiamo pensato di utilizzare diverse opzioni. Di sicuro i documenti standard quindi le CU, i cedolini che vengono generati dai vari software di studio sono molto standard, quindi hanno un layout che è facile analizzare, riconoscere ed estrarre, per esempio, con strumenti come Textract di AWS. Siamo veramente molto aperti a nuove opzioni e fa parte dell'attività di analisi che daremo in carico al team trovare la soluzione migliore, più economica e più sicura per fare questo tipo di analisi. La possibilità di apprendere nuovi formati è sicuramente qualcosa di utile ed interessante. Bisognerà valutare come implementarlo a livello di interfaccia_G, di gestione del tenant, quindi per esempio il consulente di lavoro che ha accesso e può caricare un formato specifico da far apprendere.

Privacy e Sicurezza Dati

In termini di governance e sicurezza, come gestite la privacy dei dati sensibili (soprattutto nel Co-Pilot, dove si elaborano cedolini e documenti personali)?

Risposta 8

La gestione della privacy e sicurezza dei dati personali sarà affidata alle funzionalità di AWS Bedrock e Textract, con attenzione a politiche di conservazione e accesso. AWS Bedrock sarà il fulcro per la gestione sicura dei dati personali, mentre Textract gestisce solo l'estrazione dati senza capacità AI avanzate. Textract salva i dati su bucket S3 privati con politiche di crittografia e possibile cancellazione automatica. La responsabilità legale e le politiche di retention saranno definite a livello organizzativo. La sicurezza si basa su strumenti consolidati di AWS per la protezione e la governance dei dati. L'analisi approfondita dei tool è prevista prima dell'implementazione per garantire piena compliance.

Automazione e Intervento Umano

In quali casi resta necessario l'intervento umano (es. revisione, validazione_G, correzione) e dove invece l'automazione è completa?

Risposta 9

L'intervento umano sarà richiesto in base a un sistema di confidence score che valuta automaticamente la qualità del riconoscimento documentale. Documenti con punteggio alto saranno processati automaticamente senza necessità di revisione. Sotto una soglia stabilita sarà previsto intervento umano per correzioni o validazioni. Questo approccio punta a bilanciare automazione e controllo di qualità. La pipeline di gestione sarà definita nel dettaglio nella fase di analisi e sviluppo.

Supporto e Allineamento Progetto

Che supporto fornirà l'azienda durante lo sviluppo del progetto? Con che scadenza si pensava di fissare eventuali incontri di allineamento? Preferite gestire questi incontri online o in presenza?

Risposta 10

Forniremo tutorial e documentazione per facilitare un avvio rapido. Supporto continuo sarà garantito con incontri di allineamento a cadenza concordata, preferibilmente online, per facilitare il progresso e risolvere dubbi. Si prevede un modello di lavoro simile a Scrum con sprint_G bisettimanali, ma con flessibilità sulle riunioni. Il team può stabilire la frequenza degli incontri, purché mantengano la costanza. Offriamo anche spazi fisici per incontri, ma preferiamo incontri in modalità digitale. L'obiettivo è un supporto calibrato sulle esigenze reali del team senza interferire con impegni esterni.

Repository Documentale

Quale tipo di repository bisogna utilizzare per la funzione di repository documentale accessibile dalla dashboard_G per gli utenti CdL per l'upload di documenti.

Risposta 11

La nostra idea è quella di utilizzare S3 come bucket storage per i documenti. In realtà è molto molto semplice.

Integrazione con altri sistemi

Noi dovremo integrare il sistema con un altro già esistente. Quanto sarà complicata questa integrazione?

Risposta 12

Sarà sicuramente necessaria una fase iniziale di acquisizione del dominio (i requisiti, le esigenze, gli attori). Successivamente sarà richiesta la comprensione dell'applicazione già esistente e delle sue funzionalità. Chiariti questi aspetti la complessità sarà di studiare i linguaggi di programmazione utilizzati e il codice degli applicativi messo a disposizione. Le integrazioni in sé, a livello funzionale, non saranno particolarmente complesse.
