# Piano di Progetto

## Riconoscimento Automatico del Parlato per Persone con Disartria

## 1. Introduzione

---

Nel contesto dell'assistenza ai disturbi del linguaggio di origine neurologica, proponiamo un progetto finalizzato allo sviluppo di un **sistema di riconoscimento e sintesi vocale** (anche noto come "speech-to-text-to-speech" o "sistema di riconversione vocale") che consenta di convertire il linguaggio di persone affette da disartria in testo e successivamente in parlato sintetizzato. La realizzazione di tale soluzione su misura potrebbe migliorare l'autonomia comunicativa, l'inclusione sociale e la qualità di vita di migliaia di persone, con un potenziale impatto sia sanitario che economico.

Il progetto si articolerà in due fasi principali:

1. **Fase Pilota**:
    - Obiettivo: dimostrare la fattibilità tecnologica del modello speech-to-text-to-speech su un campione selezionato di persone con disartria.
    - Attività chiave: studio preliminare, definizione delle caratteristiche del dataset, sviluppo di un prototipo ASR con sintesi vocale e validazione iniziale.
2. **Fase di Miglioramento ed Estensione**:
    - Obiettivo: potenziare e ampliare il sistema in base ai risultati del pilota, al fine di renderlo più robusto e applicabile su scala più ampia.

In allegato, presentiamo il dettaglio del piano, comprensivo di analisi dei rischi, strategie di mitigazione, budget e risorse necessarie.

## 2. Contesto e Motivazione

---

Molte patologie neurologiche (Parkinson, ictus, sclerosi multipla, traumi cranici, SLA, ecc.) possono causare **disartria**, con conseguente difficoltà nell'articolazione del parlato. È stato stimato che in Italia più di 100.000 persone presentino disartrie di vario grado, spesso con impatto notevole sulla comunicazione. L'esistente tecnologia di riconoscimento vocale fatica a interpretare correttamente queste emissioni vocali alterate, rendendo indispensabile lo sviluppo di soluzioni **custom**, adattate alle peculiarità del parlato disartrico.

Il potenziale impatto di un sistema ASR "inclusivo" per persone con disartria è considerevole:

- Miglioramento dell'autonomia comunicativa in contesti quotidiani (famiglia, lavoro, relazioni sociali).
- Riduzione del carico sull'assistenza socio-sanitaria.

# 3. Visione Generale del Progetto

---

## 3.1 Obiettivo Principale

Sviluppare e validare un **modello completo di riconversione vocale** (speech-to-text-to-speech) in lingua italiana specificamente addestrato per comprendere e trascrivere il parlato di persone con disartria di gravità moderata (o scelta diversa dopo uno studio preliminare), e successivamente convertire il testo in parlato sintetizzato chiaro e naturale.

## 3.2 Struttura in Due Fasi

1. **Fase Pilota**: Creazione di un prototipo iniziale, raccolta di dati su un gruppo ristretto e omogeneo di pazienti, analisi delle performance e validazione preliminare.
2. **Fase di Miglioramento ed Estensione**: Scalare il sistema, ottimizzare l'accuratezza, potenziare la robustezza a vari tipi di disartria e valutare un roll-out più ampio.

# 4. Fase Pilota (6 mesi)

---

## 4.1 Obiettivo

Verificare la **fattibilità tecnologica** dell'ASR disartrico e raccogliere indicazioni tecniche e cliniche fondamentali per il successivo sviluppo. L'ipotesi di base è che, attraverso un'adeguata raccolta di dati vocali e un'approfondita analisi delle caratteristiche del parlato disartrico, un modello di speech-to-text possa raggiungere **un'accuratezza sufficiente (es. Word Error Rate < 25-35%)** per consentire una comprensione funzionale del messaggio.

## 4.2 Attività Principali

1. **Studio Preliminare del Field (1° mese)**
    - Analisi dei vari sottotipi di disartria e definizione, in collaborazione con foniatri e logopedisti, dei **criteri di selezione** (es. disartria di grado moderato, target di intelligibilità 50-80%).
    - Identificazione del **gruppo pilota** di pazienti: si prevede di coinvolgere circa 50 persone con disartria di caratteristiche acustiche simili.
2. **Raccolta e Preparazione del Dataset (2°-5° mese)**
    - Registrazione di campioni vocali in lingua italiana, con frasi standard e conversazioni spontanee (100-200 frasi a partecipante).
    - Etichettatura dei dati.
    - Eventuali tecniche di data augmentation (rumore, variazioni di pitch, ecc.) per aumentare la varietà dei dati.
3. **Sviluppo Prototipo di Modello ASR e Sintesi Vocale (2°-5° mese)**
    - Utilizzo di un **modello pre-addestrato** sul parlato italiano standard (es. Whisper in italiano o altra soluzione neurale) come base per il componente speech-to-text.
    - **Fine-tuning** del modello sulle registrazioni disartriche raccolte.
    - Adozione di algoritmi di machine learning robusti (CNN-LSTM, Transformers, ecc.) con tecniche di regularization per evitare overfitting su un dataset relativamente ridotto.
    - Integrazione di un sistema di sintesi vocale (text-to-speech) di alta qualità per convertire il testo trascritto in parlato naturale.
4. **Analisi delle Prestazioni e Validazione (5°-6° mese)**
    - Valutazione del Word Error Rate (WER) e di metriche di intelligibilità automatica per il componente speech-to-text.
    - Valutazione della naturalezza e comprensibilità del parlato sintetizzato.
    - Raccolta di feedback qualitativo da parte dei partecipanti (grado di soddisfazione, usabilità).
    - Identificazione di eventuali **colli di bottiglia** tecnici o clinici (es. scarsa qualità audio, sovrapposizione con aprassia dell'eloquio, ecc.).

## 4.3 Timeline e Milestones

- **Mese 1**: Completamento dello Studio Preliminare, definizione criteri di inclusione, reclutamento partecipanti.
- **Mese 2-5**: Acquisizione dataset.
- **Mese 2-5**: Sviluppo prototipo e fine-tuning modello.
- **Mese 5-6**: Validazione interna e testing su test set disartrico.

## 4.4 Metriche di Successo

- **Quantitative**:
    - Word Error Rate (WER) < 25-35% su un test set di parlato disartrico (baseline su parlato disartrico tipicamente è molto più alta, >60-70%).
    - Tasso di miglioramento in confronto a un ASR standard non adattato (riduzione degli errori di almeno il 35%).
- **Qualitative**:
    - Feedback positivo da parte dei pazienti e dei clinici sulla comprensibilità del testo generato e sulla naturalezza del parlato sintetizzato.

# 5. Fase di Miglioramento ed Estensione (6-12 mesi)

---

**Obiettivo Generale:** Espandere e rafforzare il prototipo validato nella fase pilota, ottenendo un sistema di riconversione vocale (speech-to-text-to-speech) che funzioni in maniera robusta, adattandosi a una gamma più ampia di variabilità del parlato disartrico.

**Principali Aree di Intervento:**

- **Espansione e Raffinamento del Dataset:**
    - Ampliare il corpus dati includendo registrazioni da pazienti con diversi gradi e tipi di disartria provenienti da più centri clinici.
    - Integrare tecniche di data augmentation avanzate e implementare un sistema di apprendimento continuo per aggiornare il modello in base ai nuovi dati raccolti.
- **Ottimizzazione del Modello ASR e Sintesi Vocale:**
    - Applicare strategie di transfer learning avanzato e speaker adaptation per migliorare ulteriormente la precisione del riconoscimento.
    - Sperimentare e integrare architetture neurali ibride.
    - Migliorare la qualità e personalizzazione della sintesi vocale.

# 6. Analisi dei Rischi e Mitigazioni

---

1. **Disponibilità limitata di dati**:
    - *Mitigazione:* adottare strategie di data augmentation e collaborare con più centri di riabilitazione. Sviluppare politiche chiare su consenso e gestione della privacy.
2. **Elevata variabilità tra pazienti**:
    - *Mitigazione:* concentrare il pilota su un sottogruppo omogeneo di disartria; successivamente, estendere il modello in modo graduale.
3. **Rischio di overfitting**:
    - *Mitigazione:* ricorso a modelli pre-addestrati, regularization avanzata, cross-validation estesa.

# 7. Budget e Risorse

---

Le seguenti stime si riferiscono a due fasi di progetto (con una durata complessiva di 12-18 mesi). Le cifre possono variare in base a decisioni tecnologiche definitive.

## 7.1 Fase Pilota (6 mesi)

- **Ricerca e Sviluppo (Team tecnico)**
    - 2 Data Scientist: 6 mesi ~ 30.000 € cadauno
- *Totale stimato personale tecnico:* 60.000 €
- **Team Medico di Consulenza**
    - Sarà necessario un team di esperti consultabili composto da foniatri e logopedisti per l'intera durata del progetto
- **Infrastruttura IT**
    - Server e Cloud computing per training del modello (GPU, storage) ~ 20.000 €
    - Licenze software ed altri strumenti ~ 5.000 €
- *Totale infrastruttura:* 25.000 €

**Totale Fase Pilota:** 85.000 €

## 7.2 Fase di Miglioramento ed Estensione (6-12 mesi)

**Costi Stimati:** Il budget per questa fase è complessivamente stimato tra **150.000 e 250.000 €**.

**Principali Voci di Spesa:**

- **Team Tecnico:**
    - Maggior impegno dei Data Scientist/ML Engineer per l'integrazione di tecnologie avanzate (transfer learning avanzato, speaker adaptation e architetture ibride).
    - Specialisti in sintesi vocale per ottimizzare il componente text-to-speech.
- **Team Medico:**
    - Consulenza continuativa di specialisti in foniatria e logopedia.
- **Infrastruttura IT:**
    - Potenziamento dei server e dei servizi cloud.
    - Acquisizione di strumenti software avanzati.

# 8. Conclusioni

---

L'investimento proposto punta alla realizzazione di una **tecnologia pionieristica** in grado di colmare un vuoto significativo nell'ambito dell'inclusione digitale per persone con disartria. La struttura in due fasi (pilota e successiva estensione) permette di gestire gradualmente i rischi e consolidare progressi tangibili.

Il sistema di riconversione vocale completo offrirà non solo la trascrizione del parlato disartrico, ma anche la possibilità di comunicare attraverso una voce sintetizzata chiara e naturale, ampliando notevolmente le possibilità comunicative degli utenti.

Visti i **benefici potenziali** e l'assenza di soluzioni equivalenti in lingua italiana, riteniamo che il progetto offra **alte prospettive di ritorno** in termini di impatto sociale, opportunità di integrazione con sistemi sanitari e collaborazioni future (dispositivi di assistenza vocale, telemedicina, piattaforme di comunicazione alternativa e aumentativa).