

# Design, ingegnerizzazione e realizzazione di un sistema di dialogo basato su LLM nel dominio delle tecnologie assistive

Tesi di Laurea Magistrale



---

Relatore: Prof. Alessandro Mazzei

Co-Relatori: Dott. Pier Felice Balestrucci, Dott. Michael Oliverio

Candidato: Dott. Stefano Vittorio Porta

2 Aprile 2025

Università degli Studi di Torino, Dipartimento di Informatica - Anno Accademico 2023/2024

# 1. Contesto

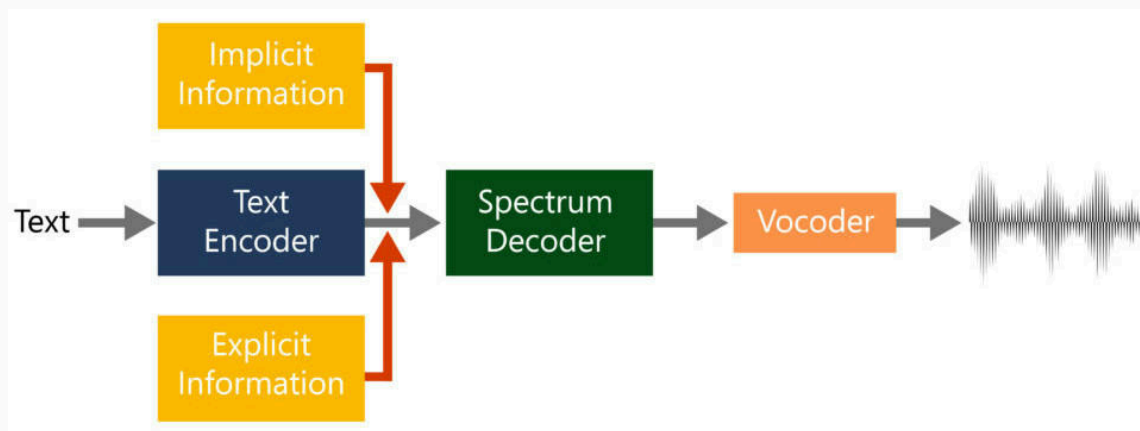


Iniziamo ad ambientarci:

- **Lettura di contenuti testuali:** da 30 anni sono disponibili sistemi di sintesi vocale integrati in smartphone e computer.

Iniziamo ad ambientarci:

- **Lettura di contenuti testuali**: da 30 anni sono disponibili sistemi di sintesi vocale integrati in smartphone e computer.
- Essenziali per le persone con **disabilità visive**: consentono di accedere a contenuti testuali in modo autonomo e senza l'ausilio di un lettore umano.



- Pagine web. I sistemi TTS sono in grado di
  - Leggere il testo
  - Interpretare la struttura del documento
  - Questo permette di seguire il flusso di lettura per fornire un'esperienza di qualità

## 1.2 Problema!



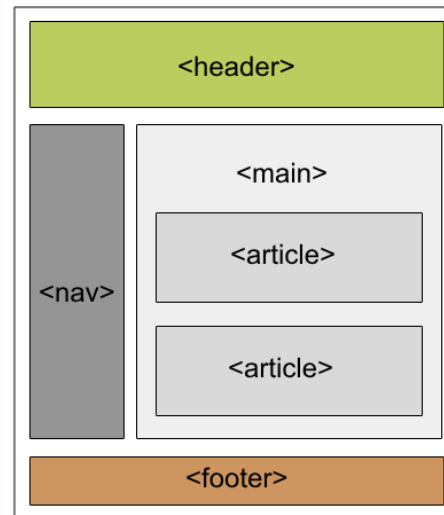
Tutto funziona, a patto che...

## 1.2 Problema!



Tutto funziona, a patto che...

- La pagina web sia strutturata in modo semantico

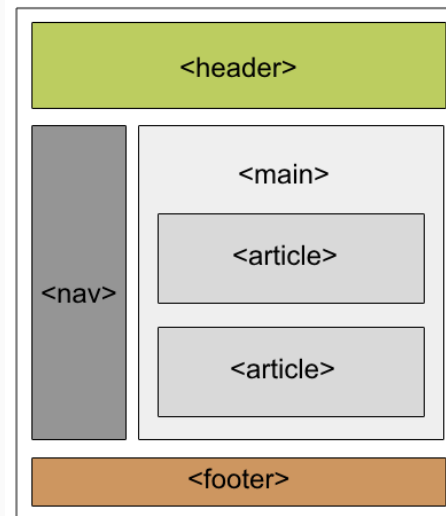


## 1.2 Problema!



Tutto funziona, a patto che...

- La pagina web sia strutturata in modo semantico



- Le immagini siano accompagnate da un testo alternativo

```
<img alt="..."> o <img aria-label="...">
```

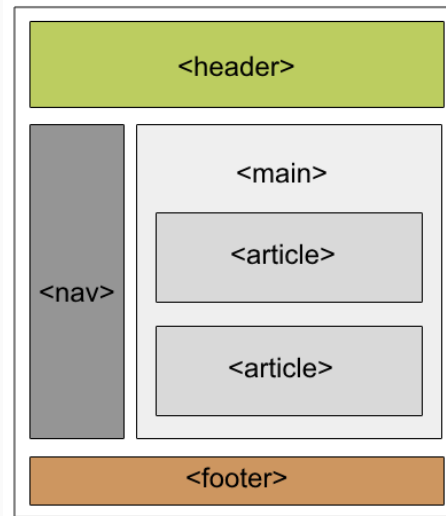


## 1.2 Problema!



Tutto funziona, a patto che...

- La pagina web sia strutturata in modo semantico



- Le immagini siano accompagnate da un testo alternativo

`<img alt="...">` o `<img aria-label="...">`

**Queste due condizioni dipendono da chi prepara il contenuto!**

I sistemi di TTS non sono in grado di interpretare il significato di un'immagine o di un elemento puramente visivo.

---

<sup>[1]</sup>WebAIM, «The WebAIM Million: The 2024 Report on the Accessibility of the Top 1,000,000 Home Pages». [Online]. Disponibile su: <https://webaim.org/projects/million/>

## 1.2 Problema!



I sistemi di TTS non sono in grado di interpretare il significato di un'immagine o di un elemento puramente visivo.



---

<sup>[1]</sup>WebAIM, «The WebAIM Million: The 2024 Report on the Accessibility of the Top 1,000,000 Home Pages». [Online]. Disponibile su: <https://webaim.org/projects/million/>

## 1.2 Problema!



I sistemi di TTS non sono in grado di interpretare il significato di un'immagine o di un elemento puramente visivo.



Se non viene fornita un'alternativa testuale contenente delle informazioni utili, l'utente non potrà comprendere appieno il contenuto della pagina o di uno specifico elemento!

---

<sup>[1]</sup>WebAIM, «The WebAIM Million: The 2024 Report on the Accessibility of the Top 1,000,000 Home Pages». [Online]. Disponibile su: <https://webaim.org/projects/million/>

## 1.2 Problema!



I sistemi di TTS non sono in grado di interpretare il significato di un'immagine o di un elemento puramente visivo.



Se non viene fornita un'alternativa testuale contenente delle informazioni utili, l'utente non potrà comprendere appieno il contenuto della pagina o di uno specifico elemento!

Una di quattro immagini sul web non ha una descrizione testuale o non è informativa<sup>[1]</sup>.

---

<sup>[1]</sup>WebAIM, «The WebAIM Million: The 2024 Report on the Accessibility of the Top 1,000,000 Home Pages». [Online]. Disponibile su: <https://webaim.org/projects/million/>

## 1.3 Anche peggio...

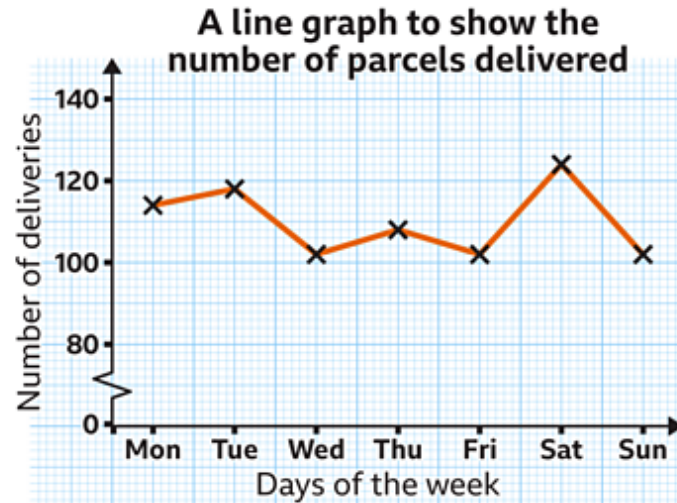


Se per le immagini possiamo utilizzare tecniche di **Computer Vision** o **Reti Neurali** per generare automaticamente un testo alternativo, per le rappresentazioni grafiche di dati (grafici, diagrammi, mappe) non è così semplice.

## 1.3 Anche peggio...



Se per le immagini possiamo utilizzare tecniche di **Computer Vision** o **Reti Neurali** per generare automaticamente un testo alternativo, per le rappresentazioni grafiche di dati (grafici, diagrammi, mappe) non è così semplice.

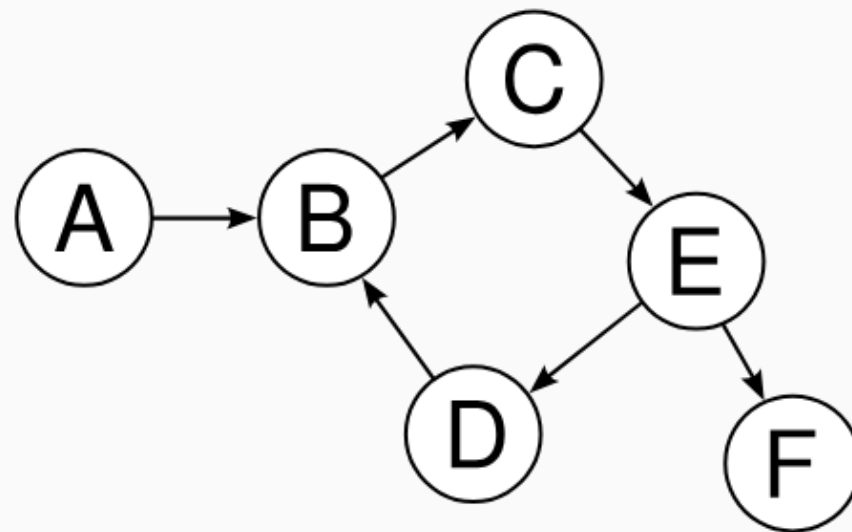
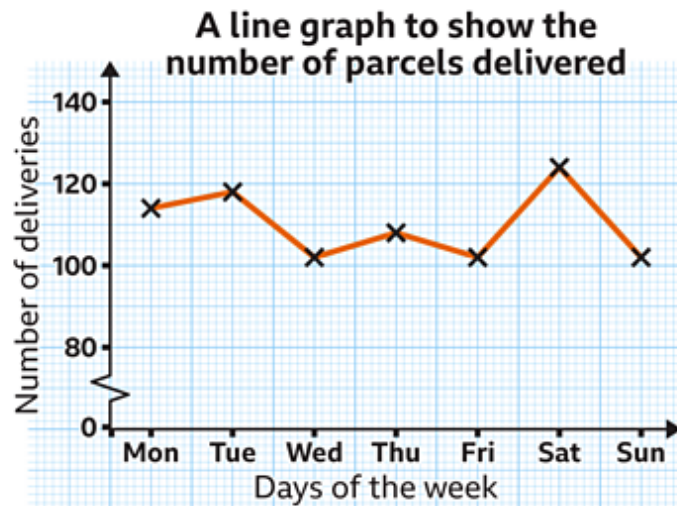




## 1.3 Anche peggio...



Se per le immagini possiamo utilizzare tecniche di **Computer Vision** o **Reti Neurali** per generare automaticamente un testo alternativo, per le rappresentazioni grafiche di dati (grafici, diagrammi, mappe) non è così semplice.





- Il Progetto NoVAGraphS si propone di rendere più accessibili questi contenuti, mediante la costruzione di sistemi di dialogo (Chatbot).

- Il Progetto NoVAGraphS si propone di rendere più accessibili questi contenuti, mediante la costruzione di sistemi di dialogo (Chatbot).
- Con essi è possibile interagire per ottenere informazioni sui dati presenti in grafi o strutture simili, per avere una comprensione **profonda** del contenuto.

- Il Progetto NoVAGraphS si propone di rendere più accessibili questi contenuti, mediante la costruzione di sistemi di dialogo (Chatbot).
- Con essi è possibile interagire per ottenere informazioni sui dati presenti in grafi o strutture simili, per avere una comprensione **profonda** del contenuto.
- Il Progetto originale fa fondamento su AIML (Artificial Intelligence Markup Language), un linguaggio di markup per la creazione di chatbot.

```
<category>  
  <pattern>CIAO</pattern>  
  <template>Ciao! Come posso aiutarti oggi?</template>  
</category>
```

```
<category>  
  <pattern>MI CHIAMO *</pattern>  
  <template>  
    Ciao <star/>, piacere di conoscerti!  
  </template>  
</category>
```

```
<category>
  <pattern>IL MIO COLORE PREFERITO È *</pattern>
  <template>
    <think>
      <set name="colore"><star/></set>
    </think>
    Ok, ricorderò che il tuo colore preferito è <star/>.
  </template>
</category>
```

```
<category>
  <pattern>QUAL È IL MIO COLORE PREFERITO</pattern>
  <template>
    Il tuo colore preferito è <get name="colore"/>.
  </template>
</category>
```

- Le strategie di wildcard e pattern matching restano **prevalentemente letterali**: Se una frase si discosta dal pattern previsto, il sistema fallisce il matching

- Le strategie di wildcard e pattern matching restano **prevalentemente letterali**: Se una frase si discosta dal pattern previsto, il sistema fallisce il matching
- Sono disponibili ridotte funzionalità per la gestione di *sinonimi*, semplificazione delle *locuzioni* e *correzione ortografica*



- Le strategie di wildcard e pattern matching restano **prevalentemente letterali**: Se una frase si discosta dal pattern previsto, il sistema fallisce il matching
- Sono disponibili ridotte funzionalità per la gestione di *sinonimi*, semplificazione delle *locuzioni* e *correzione ortografica*
- La **gestione del contesto** (via <that>, <topic>, <star>, ecc.) è rudimentale

- Le strategie di wildcard e pattern matching restano **prevalentemente letterali**: Se una frase si discosta dal pattern previsto, il sistema fallisce il matching
- Sono disponibili ridotte funzionalità per la gestione di *sinonimi*, semplificazione delle *locuzioni* e *correzione ortografica*
- La **gestione del contesto** (via <that>, <topic>, <star>, ecc.) è rudimentale
- L'integrazione (via <sraix>) con **basi di conoscenza esterne** (KB, database, API) è possibile implementando funzioni personalizzate, ma è di difficile gestione

- Le strategie di wildcard e pattern matching restano **prevalentemente letterali**: Se una frase si discosta dal pattern previsto, il sistema fallisce il matching
- Sono disponibili ridotte funzionalità per la gestione di *sinonimi*, semplificazione delle *locuzioni* e *correzione ortografica*
- La **gestione del contesto** (via <that>, <topic>, <star>, ecc.) è rudimentale
- L'integrazione (via <sraix>) con **basi di conoscenza esterne** (KB, database, API) è possibile implementando funzioni personalizzate, ma è di difficile gestione
- Le risposte generate sono **statiche e predefinite**, e non possono essere generate dinamicamente in base a dati esterni o a contesti più ampi in modo automatico

- Sviluppare un sistema di dialogo che superi le limitazioni di AIML evidenziate

- Sviluppare un sistema di dialogo che superi le limitazioni di AIML evidenziate
- Integrare tecniche di **Natural Language Understanding** (NLU) e **Retrieval-Augmented Generation** (RAG) per migliorare l'esperienza d'uso

- Sviluppare un sistema di dialogo che superi le limitazioni di AIML evidenziate
- Integrare tecniche di **Natural Language Understanding** (NLU) e **Retrieval-Augmented Generation** (RAG) per migliorare l'esperienza d'uso
- Assicurare una elevata facilità di estensione e personalizzazione per diversi domini e applicazioni

## **2. Natural Language Understanding**

---

Il primo elemento dello stack di NLP rispetto ad AIML che vogliamo migliorare è il riconoscimento delle intenzioni dell'utente.



Il primo elemento dello stack di NLP rispetto ad AIML che vogliamo migliorare è il riconoscimento delle intenzioni dell'utente.

- Non useremo più un sistema basato su pattern matching ed espressioni regolari
- Riconosceremo la categoria di interazione affidandoci ad un classificatore basato su LLM
- Le parti variabili della frase (slot) verranno estratte tramite un sistema di Named Entity Recognition (NER)

- Essendo un task supervisionato, bisogna partire con l'etichettatura dei dati.

- Essendo un task supervisionato, bisogna partire con l'etichettatura dei dati.
- Il dataset utilizzato proviene dalle precedenti pubblicazioni del progetto NoVAGraphS, e contiene 350 interazioni degli utenti prodotte durante precedenti sperimentazioni.

- Essendo un task supervisionato, bisogna partire con l'etichettatura dei dati.
- Il dataset utilizzato proviene dalle precedenti pubblicazioni del progetto NoVAGraphS, e contiene 350 interazioni degli utenti prodotte durante precedenti sperimentazioni.
- L'annotazione:
  - Inizialmente è stata effettuata automaticamente

- Essendo un task supervisionato, bisogna partire con l'etichettatura dei dati.
- Il dataset utilizzato proviene dalle precedenti pubblicazioni del progetto NoVAGraphS, e contiene 350 interazioni degli utenti prodotte durante precedenti sperimentazioni.
- L'annotazione:
  - Inizialmente è stata effettuata automaticamente
  - Successivamente è stata completamente riveduta ed effettuata manualmente

La classificazione automatica è stata effettuata tramite prompting:

La classificazione automatica è stata effettuata tramite prompting:

1. Due LLM diverse (*Gemma2*, *Llama3.1*) sono state eseguite localmente

La classificazione automatica è stata effettuata tramite prompting:

1. Due LLM diverse (*Gemma2*, *Llama3.1*) sono state eseguite localmente
2. Ciascuna ha ricevuto tutte le interazioni (una per una) assieme alla lista delle possibili classi



La classificazione automatica è stata effettuata tramite prompting:

1. Due LLM diverse (*Gemma2*, *Llama3.1*) sono state eseguite localmente
2. Ciascuna ha ricevuto tutte le interazioni (una per una) assieme alla lista delle possibili classi
3. È stata selezionata la classe con majority vote

## 2.2 Classificazione

La classificazione automatica è stata effettuata tramite prompting:

1. Due LLM diverse (*Gemma2*, *Llama3.1*) sono state eseguite localmente
2. Ciascuna ha ricevuto tutte le interazioni (una per una) assieme alla lista delle possibili classi
3. È stata selezionata la classe con majority vote

ID	gemma2:9b	gemma2:9b	llama3.1:8b	llama3.1:8b
0	START	START	START	START
1	GEN_INFO	GEN_INFO	GEN_INFO	GEN_INFO
2	SPEC_TRANS	SPEC_TRANS	TRANS_BETWEEN	TRANS_BETWEEN
3	Please provide the interaction. : START	START	START	START
...	...	...	...	...
287	REPETITIVE_PAT	REPETITIVE_PAT	REPETITIVE_PAT	REPETITIVE_PAT
288	TRANS_DETAIL	TRANS_DETAIL	TRANS_DETAIL	GEN_INFO
289	GRAMMAR	GRAMMAR	FINAL_STATE	FINAL_STATE

- In seguito a una analisi dei dati è risultato che le classi fossero troppo sbilanciate, e troppo generiche.
- Questo non avrebbe aiutato il modello che sarebbe stato addestrato a riconoscere le classi con precisione e affidabilità.

- In seguito a una analisi dei dati è risultato che le classi fossero troppo sbilanciate, e troppo generiche.
- Questo non avrebbe aiutato il modello che sarebbe stato addestrato a riconoscere le classi con precisione e affidabilità.
- Sono state ridefinite le classi, dividendole in due livelli di granularità:

- In seguito a una analisi dei dati è risultato che le classi fossero troppo sbilanciate, e troppo generiche.
- Questo non avrebbe aiutato il modello che sarebbe stato addestrato a riconoscere le classi con precisione e affidabilità.
- Sono state ridefinite le classi, dividendole in due livelli di granularità:
  - Le **classi principali** da 21 sono state ridotte a 7

- In seguito a una analisi dei dati è risultato che le classi fossero troppo sbilanciate, e troppo generiche.
- Questo non avrebbe aiutato il modello che sarebbe stato addestrato a riconoscere le classi con precisione e affidabilità.
- Sono state ridefinite le classi, dividendole in due livelli di granularità:
  - Le **classi principali** da 21 sono state ridotte a 7
  - Sono state introdotte le **classi secondarie** per ogni classe principale, per un totale di 33.

### Classi principali

Classe	Scopo	Numero di Esempi
transition	Domande che riguardano le transizioni tra gli stati	77
automaton	Domande che riguardano l'automa in generale	48
state	Domande che riguardano gli stati dell'automa	48
grammar	Domande che riguardano la grammatica riconosciuta dall'automa	33
theory	Domande di teoria generale sugli automi	15
start	Domande che avviano l'interazione con il sistema	6
off_topic	Domande non pertinenti al dominio che il sistema deve saper gestire	2

Classi secondarie per la classe primaria dell' **Automa**

Sottoclassi	Scopo	Numero di Esempi
description	Descrizioni generali sull'automa	14
description_brief	Descrizione generale (breve) sull'automa	10
directionality	Domande riguardanti la direzionalità o meno dell'intero automa	1
list	Informazioni generali su nodi e archi	1
pattern	Presenza di pattern particolari nell'automa	9
representation	Rappresentazione spaziale dell'automa	13

(esempio)



### **3. Retrieval Augmented Generation**

---

## **4. Ingegnerizzazione del sistema**

---

## **5. Conclusioni**

---

Grazie per l'attenzione!  
Domande?