



19 Maggio 2026
Seminario permanente Credit

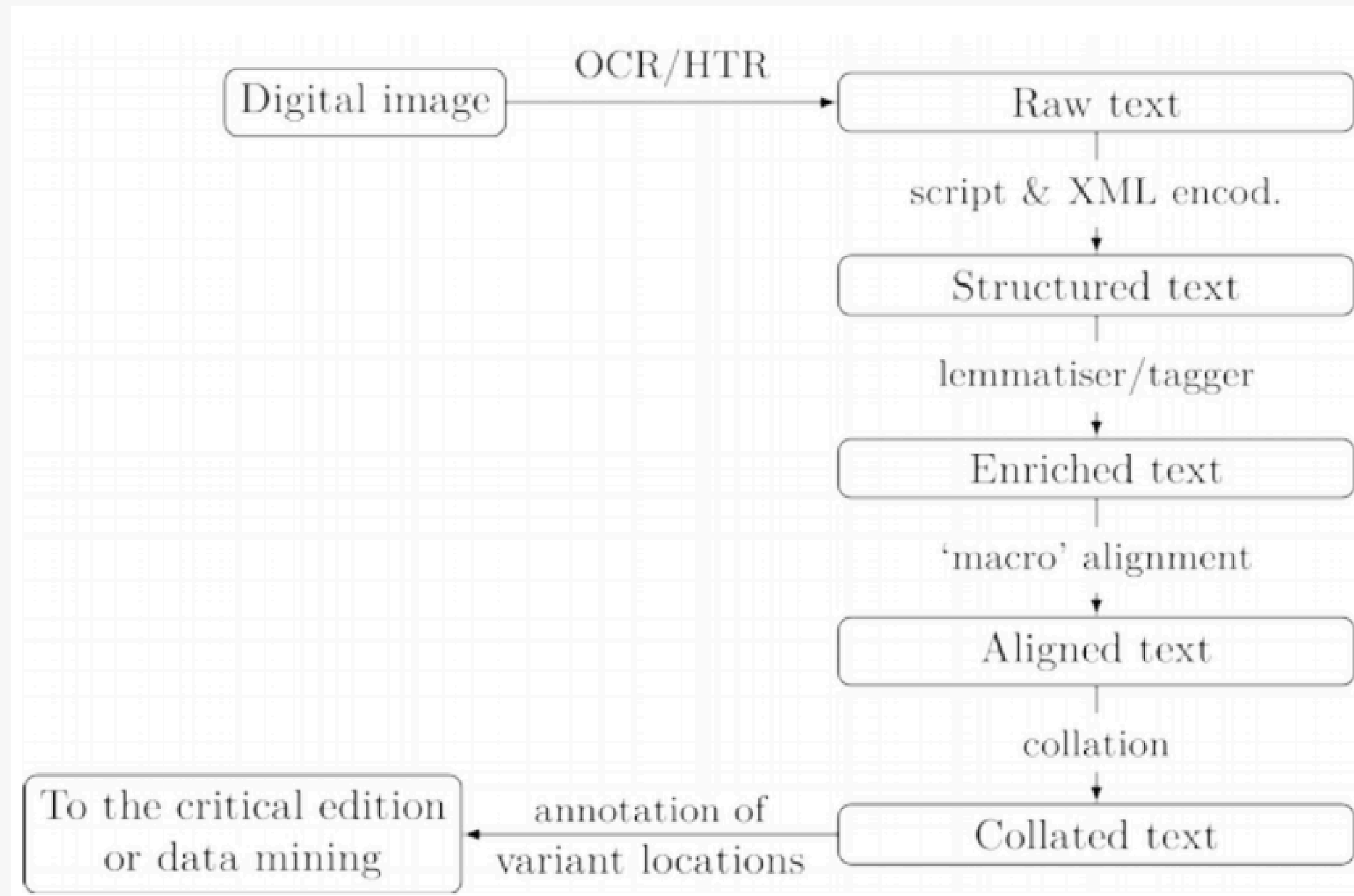


UNIVERSITÀ DI PADOVA
Dipartimento
di Studi linguistici
e letterari

AI E MACHINE LEARNING TRA FILOLOGIA DIGITALE E LINGUISTICA COMPUTAZIONALE: Sfide, limiti e nuove prospettive.

Giacomo Costa, PhD student
Università degli studi di Padova
giacomo.costa.4@phd.unipd.it

Serena Malatesta, PhD student
Università degli studi di Padova - Université de Tours
serena.malatesta@phd.unipd.it



Perché lemmatizzare i testimoni di una tradizione manoscritta?

- la lemmatizzazione precedere il lavoro di allineamento per la collazione automatica;
- l'analisi delle varianti si pone a un livello più alto di astrazione: non si confrontano più le forme flesse ma i loro esponenti lessicali.



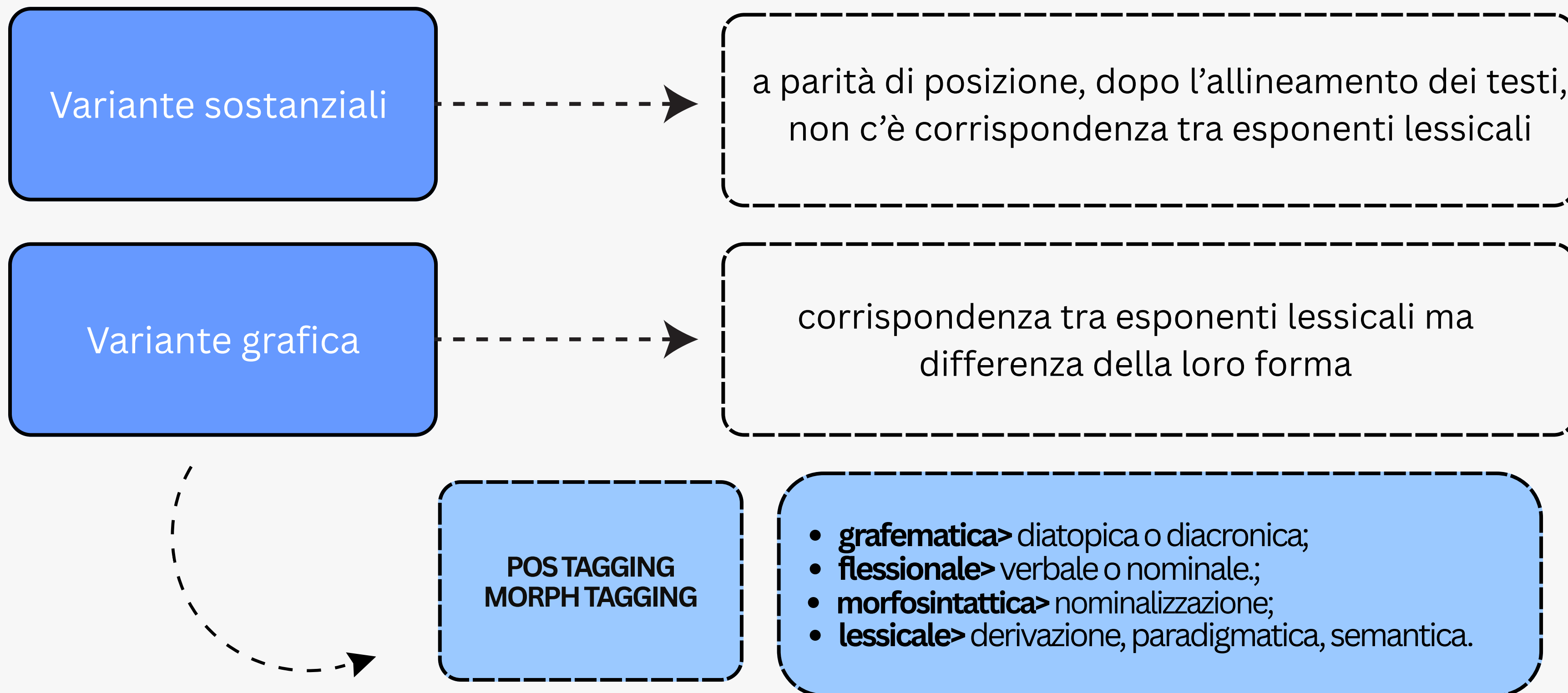
- facilita l'allineamento dei testimoni;
- “superamento” distinzione tra varianti puramente grafiche e varianti sostanziali.



“
One substantial conclusion to be found in the data, and one that reinforces finding made previously, is that 'insignificant' variation is really not that insignificant at all. We have seen that some philologists prefer to exclude it entirely; others [...] include the information but give it as low a weighting as possible. This experiment, together with several others, strongly suggests that pur practices for handling this sort of 'insignificant' variation are in dire need of revision.
”

Tara L. Andrews , *Analysis of Variation Significance in Artificial Traditions Using Stemmaweb*, in «Digital Scholarship in the Humanities», 31(3), 2016: 523–39 (10.1093/llc/fqu072), p. 538.

Con la lemmatizzazione



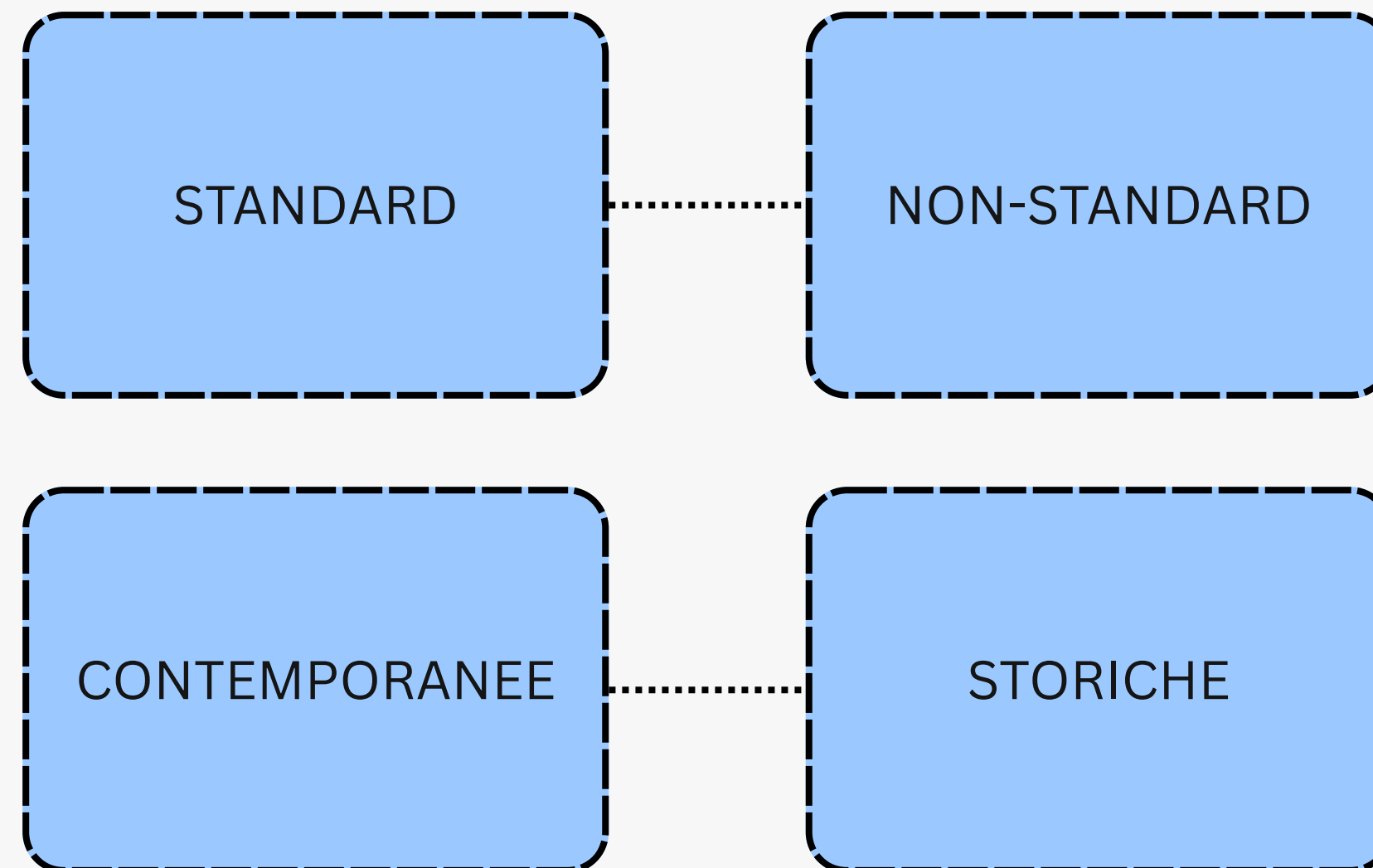
Tradizioni manoscritte ad alta variazione linguistica

Huon d'Auvergne franco-italiano

<p>ms. B (1441), v. 9006 De ces parolle le queens Hue l'en mercie.</p>	<p>Ms Br (seconda metà del XIV secolo), v. 165 De tieles paroles le quens Ugon l'en mercie.</p>
<p>ms. P (XIV-XV), v. 3914 De queste parlar lo conte Ugo merçie</p>	<p>ms. T (1441), v. 8676 de queste parole lo conte Ugone li rende merzede</p>

Quali sono i costi, in termini economici e di tempo, per un processo di questo tipo?

- Lo stato dell'arte di modelli ML nel campo dell'NLP si limita principalmente alle lingue contemporanee e standard.

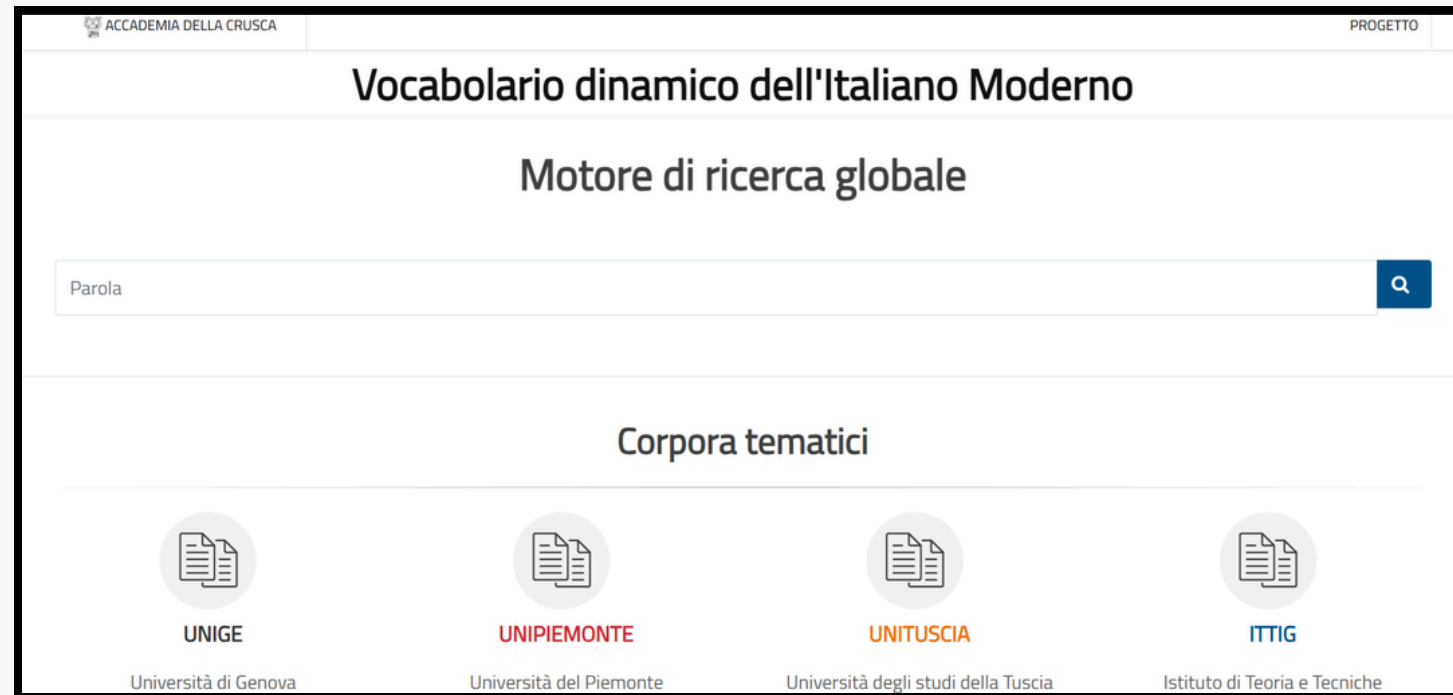


Qual è la strada più percorsa?

Adattare un modello computazionale addestrato per una lingua contemporanea al target del nostro corpus

- normalizzazione grafica;
- estensione del lessico;
- estensione del corpus di addestramento della lingua moderna con un *sample* rappresentativo della varietà storica in questione.

Il progetto VoDIM



**Impossibilità di normalizzazione
delle varietà storiche
dell'italiano**

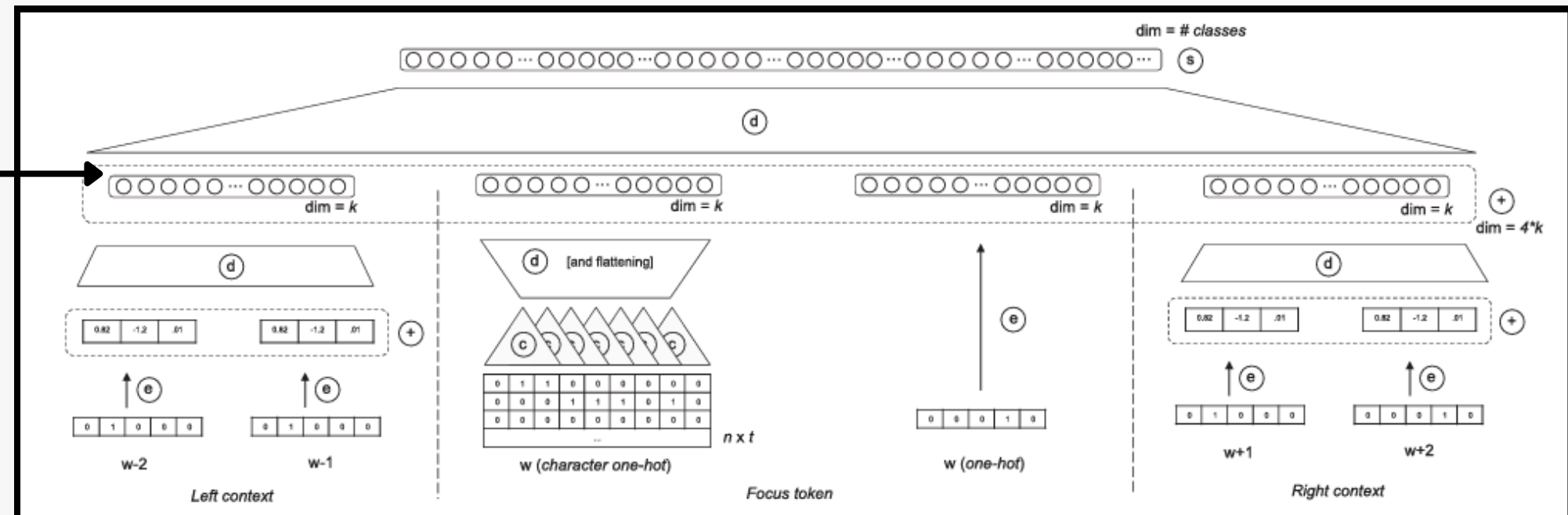
CLTK



**Necessità di normalizzazione
grafica di alcune varietà del
francese antico**

NLP e Deep Learning

Primo tentativo: 2016, costituzione di quattro dataset per il Middel Dutch



convoluzioni 1D sui caratteri

word embeddings

A differenza dei metodi tradizionali, le reti profonde imparano automaticamente quali combinazioni di caratteristiche sono più utili.

Una soluzione alternativa e integrata?

FrlngE (The French in/of Italy: Code-MixiNG in Medieval Europe)

adattamento di un modello già esistente senza normalizzazione

addestramento di un modello ad hoc

PYRRHA

- interfaccia di post-correzione basata su algoritmi di Pie;
- no rule based: adatta al trattamento di sistemi linguistici ad alta varietà grafica.

PRO:

- si utilizza l'interfaccia senza lavorare direttamente sul modello computazionale;
- risultati ottimi per il francese antico, non pienamente soddisfacenti per il franco-italiano (variazione non solo grafica).

CONTRO:

- completa dipendenza dal proprietario del software;
- mancati aggiornamenti da parte del centro di ricerca.

- integrazione del lemmario con quello del TLIO
- etichette specifiche

L'interfaccia di Pyrrha

311	Ni	sens si garnie Ni per mer ni	2
		LEMMA ni2	POS ADVneg
		MORPH MORPH=empty	Save
312	per	si garnie Ni per mer ni per	161
		LEMMA par	POS PRE
		MORPH SPEC=it MORPH=empty	Save
313	mer	garnie Ni per mer ni per terre	12
		LEMMA mer	POS NOMcom
		MORPH NOMB.=s GENRE=f CAS=r	Save
314	ni	Ni per mer ni per terre de	66
		LEMMA ni2	POS ADVneg
		MORPH MORPH=empty	Save
315	per	per mer ni per terre de l'	161
		LEMMA par	POS PRE
		MORPH SPEC=it MORPH=empty	Save
316	terre	mer ni per terre de l' eisue	44
		LEMMA terre	POS NOMcom
		MORPH NOMB.=s GENRE=f CAS=r	Save
317	de	ni per terre de l' eisue n'	364
		LEMMA de	POS PRE
		MORPH MORPH=empty	Save
318	l'	per terre de l' eisue n' a	201
		LEMMA	POS
		MORPH	

Alcune considerazioni finali

- lemmatizzazione: vantaggi per l'allineamento di tradizioni testuali caratterizzate da un'alta variazione linguistica;
- vantaggi sul piano dell'analisi semi-automatica delle varianti.

↓
MA

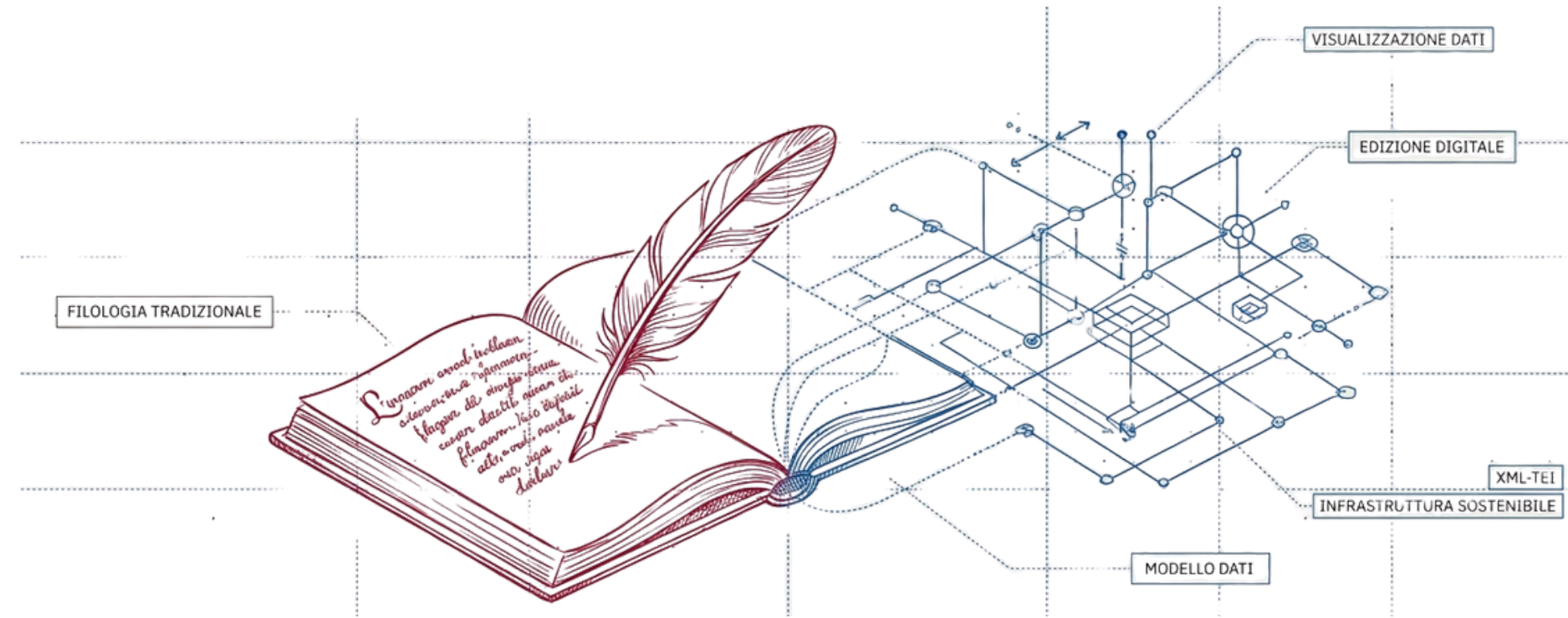
- problema della normalizzazione grafica dei testi;
- poca disponibilità di modelli computazionali già addestrati per sistemi linguistici storici e non standard;
- limiti nell'adattamento di modelli già esistenti per lingue contemporanee;
- un filologo non ha competenze e risorse per lo sviluppo di un modello ad hoc.

**Necessità di sviluppo di un
“ecosistema” ingegneristico che
integri personale di ricerca con:
competenze in:**

- filologia
- linguistica storica
- linguistica computazionale
- ingegneria informatica

+ molti soldi e molto tempo:)))

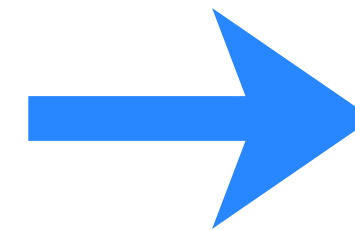
Il workflow editoriale: la visualizzazione



Serena Malatesta, PhD student
Università degli Studi di Padova - Université de Tours
serena.malatesta@phd.unipd.it

“ Il digitale ha una funzione trasformativa sia sul metodo di produrre l'edizione, sia sul modo in cui questa viene utilizzata.

Pierazzo E. - Mancinelli T., *Che cos'è un'edizione scientifica digitale*, Roma, Carocci Editore, 2020.



Dove fermarsi?

Riprodurre ogni singolo dettaglio è impossibile e inutile. Per questo, l'editore deve porre dei limiti precisi che non sono mancanze, ma scelte interpretative fondamentali. La sfida scientifica sta nel selezionare i confini giusti per rendere il testo un modello utile allo studio e alla ricerca.

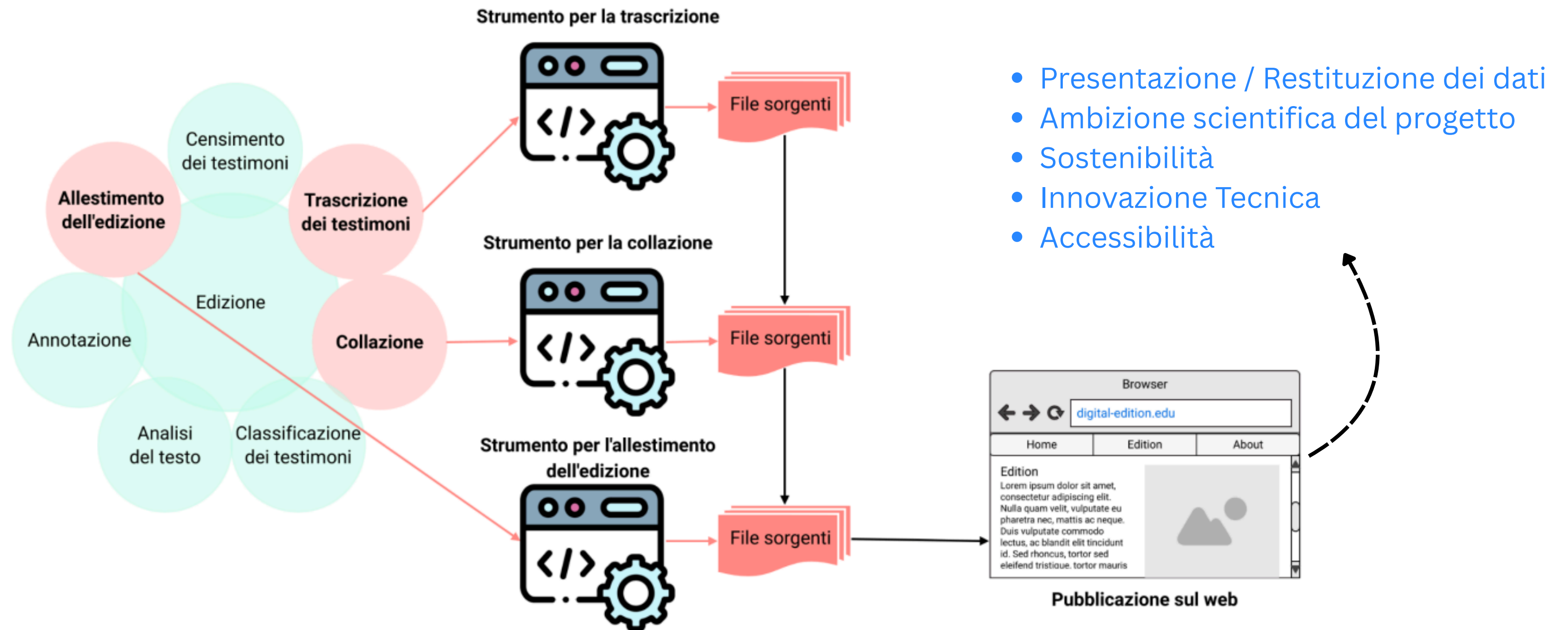
La griglia decisionale

“ Decision are taken according to five parameters:

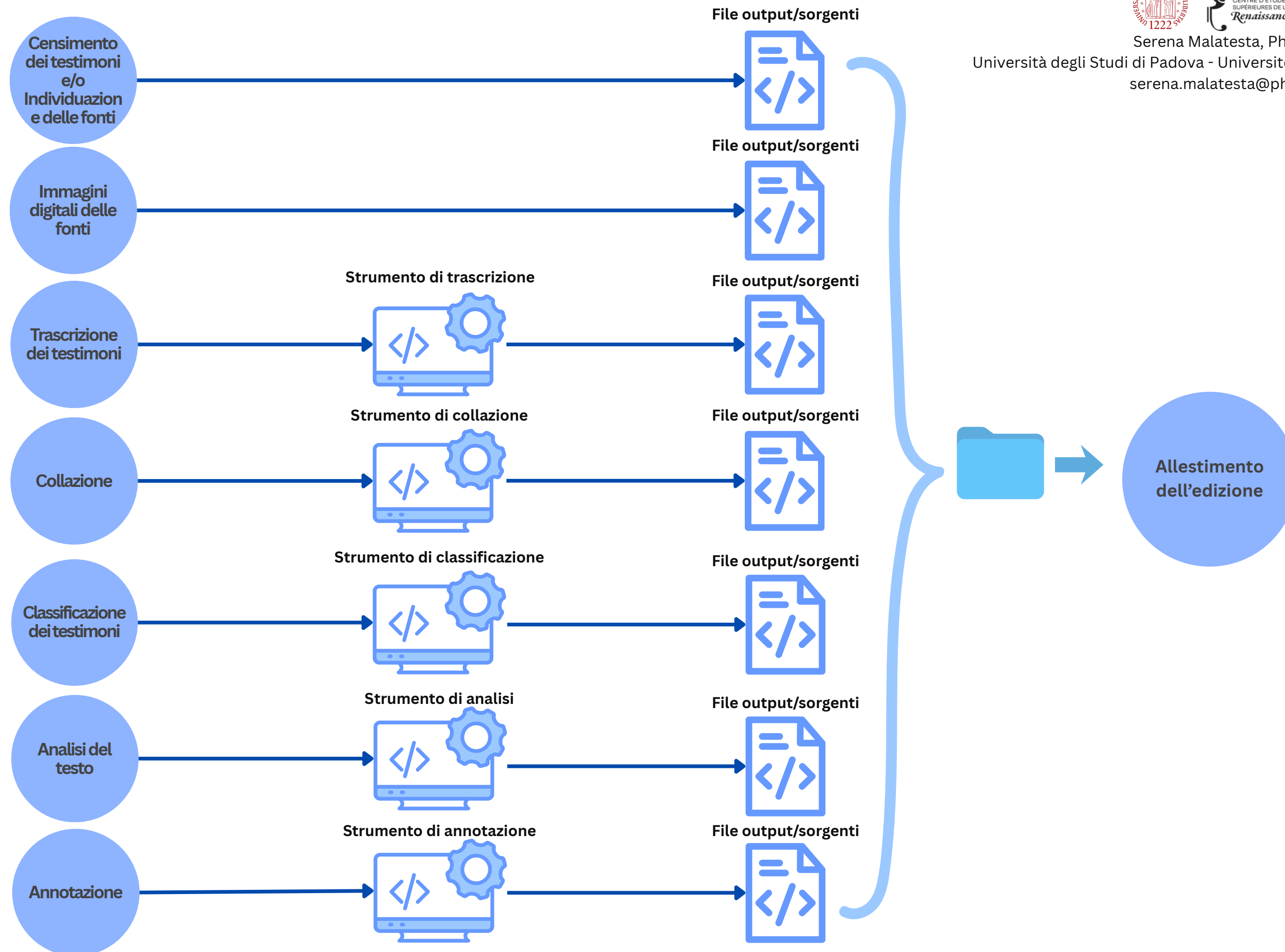
- (1) The purpose of the edition (or the needs of the editors);
- (2) The needs of the others (prospected readers, scholars);
- (3) The nature of the document;
- (4) The capabilities of the publishing technology;
- (5) The costs of encoding/the amount of time available for the job.

Pierazzo E., *A rationale of digital documentary editions*, *Literary and Linguistic Computing*, vol. 26, n. 4, 2011.

	Purpose of edition	Readers	Nature of document	Publishing technology	Costs
Document	✓		✓	✓	✓
Topology	✓	✓	✓		✓
Handwriting		✓		✓	
Orthography			✓		✓
Reading Facilitators	✓	✓		✓	
Shorthands			✓	✓	
Genesis	✓	✓		✓	
Textuality	✓		✓		✓
Semantic	✓	✓		✓	✓
Linguistics	✓		✓	✓	
Decorations		✓	✓		✓
Other				✓	



EDIZIONE =





Serena Malatesta, PhD student
Università degli Studi di Padova - Université de Tours
serena.malatesta@phd.unipd.it

COME PUBBLICARE E VISUALIZZARE L'EDIZIONE DIGITALE?

COME PUBBLICARE E VISUALIZZARE L'EDIZIONE DIGITALE?

- Acquisire competenze informatiche avanzate
- Affidarsi a strumenti preconfezionati
- Delegare la fase di sviluppo a specialisti

- Acquisire competenze informatiche avanzate
 - **Vantaggi:** Controllo diretto sull'intero processo
 - **Svantaggi:**
 - Costo umano elevato (tempo, formazione, frustrazione);
 - Rischio di sbilanciamento verso figure ibride;
 - Non è sostenibile pensare di diffondere uniformemente le competenze.

● Affidarsi a strumenti preconfezionati

Edizioni prêt-à-porter

Pierazzo E., *Quale infrastruttura per le edizioni digitali? Dalla tecnologia all'etica*, Textual Cultures, vol. 12, n.2, 2019, pp. 5-17.

● **Vantaggi:**

- Riducono concretamente i tempi e i costi di pubblicazione;
- Garantiscono un alto livello di standardizzazione e accessibilità iniziale.

● **Svantaggi:**

- Rigidità strutturale, le scelte progettuali devono essere adattate alle funzionalità predefinite dal sistema;
- Rinuncia di personalizzazioni significative, con il rischio di non soddisfare le esigenze scientifiche del progetto.

● Delegare la fase di sviluppo a specialisti

Edizioni haute couture

Pierazzo E., *Quale infrastruttura per le edizioni digitali? Dalla tecnologia all'etica*, Textual Cultures, vol. 12, n.2, 2019, pp. 5-17.

● **Vantaggi:**

- Risultati fortemente innovativi;
- Massima personalizzazione per le esigenze specifiche del progetto.

● **Svantaggi:**

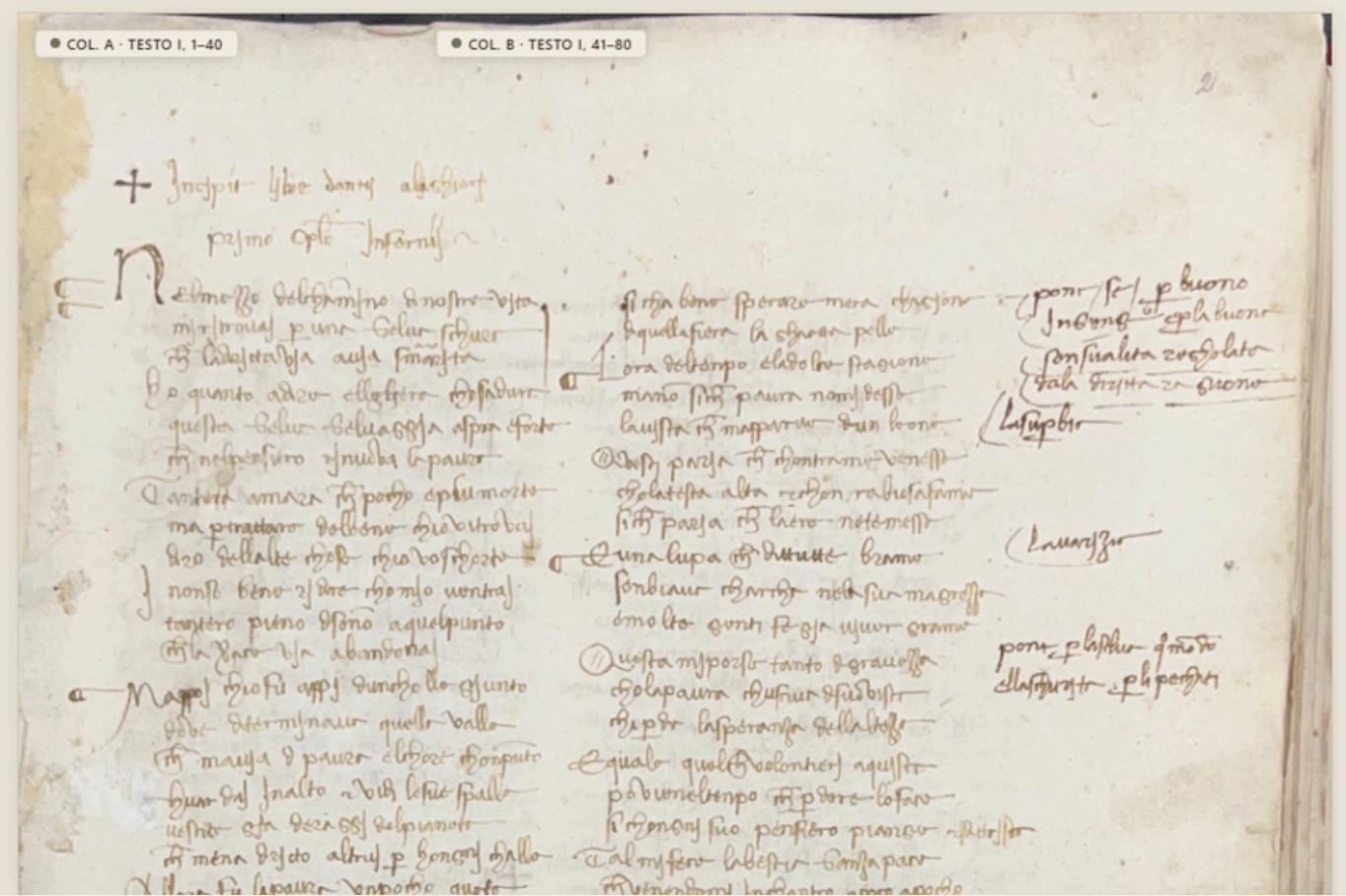
- Costi insostenibili;
- Lavoro concentrato in tempi brevi, vincolato alla durata dei fondi e alla necessità di personale specializzato;
- Disuguaglianza strutturale;
- Esclusione dei giovani.

Files

main

Go to file

- assets
- facsimile
- .gitkeep
- data
- README.md
- app.js
- index.html
- styles.css



PRIMO CAPITOLO INFERNII

- [c. 2r] COL A
- Nel mezzo del cammino di nostra vita
 - mi ritrovai per una selva scura,
 - ché la diritta via avia smarrita.
 - Oh quanto a dire e gli era cosa dura
 - questa selva selvaggia, aspra e forte
 - che nel pensiero rinnova la paura!
 - Tant'è amara che poco è più morte;
 - ma per trattare del bene ch'io vi trovai,
 - dirò dell'altre cose ch'io v'ho scorte.
 - Io non so bene ridire com'io v'entrai,
 - tant'ero pieno di sonno a quel punto
 - che la verace via abandonai
 - Ma ppoi ch'io fu' a ppiè d'un colle giunto

Inf. I, 1 **Nel mezzo del cammin di nostra vita**

In questo luogo intende l'autore che egli era nel torno di trentatrè anni, insino al qual tempo crescono le virtù del corpo e da indi innanzi cominciano a mancare. E però dice *nel mezzo* quasi come sia in fra 'l crescere el menovare; e ancora perché quel tempo può essere appellato mezzo corso di nostra vita umana assai giustamente. E però dice il profeta:

«Anni Sion aranea meditabuntur: dies annorum nostrorum in ipsis LXX anni»

², cioè, sì come radi pensano, gli anni nostri stanno in LXX, sì che ne' XXXIII o ne' trentacinque l'uomo è nel mezzo di sua vita.

Inf. I, 2 **Mi ritrovai per una selva scura**

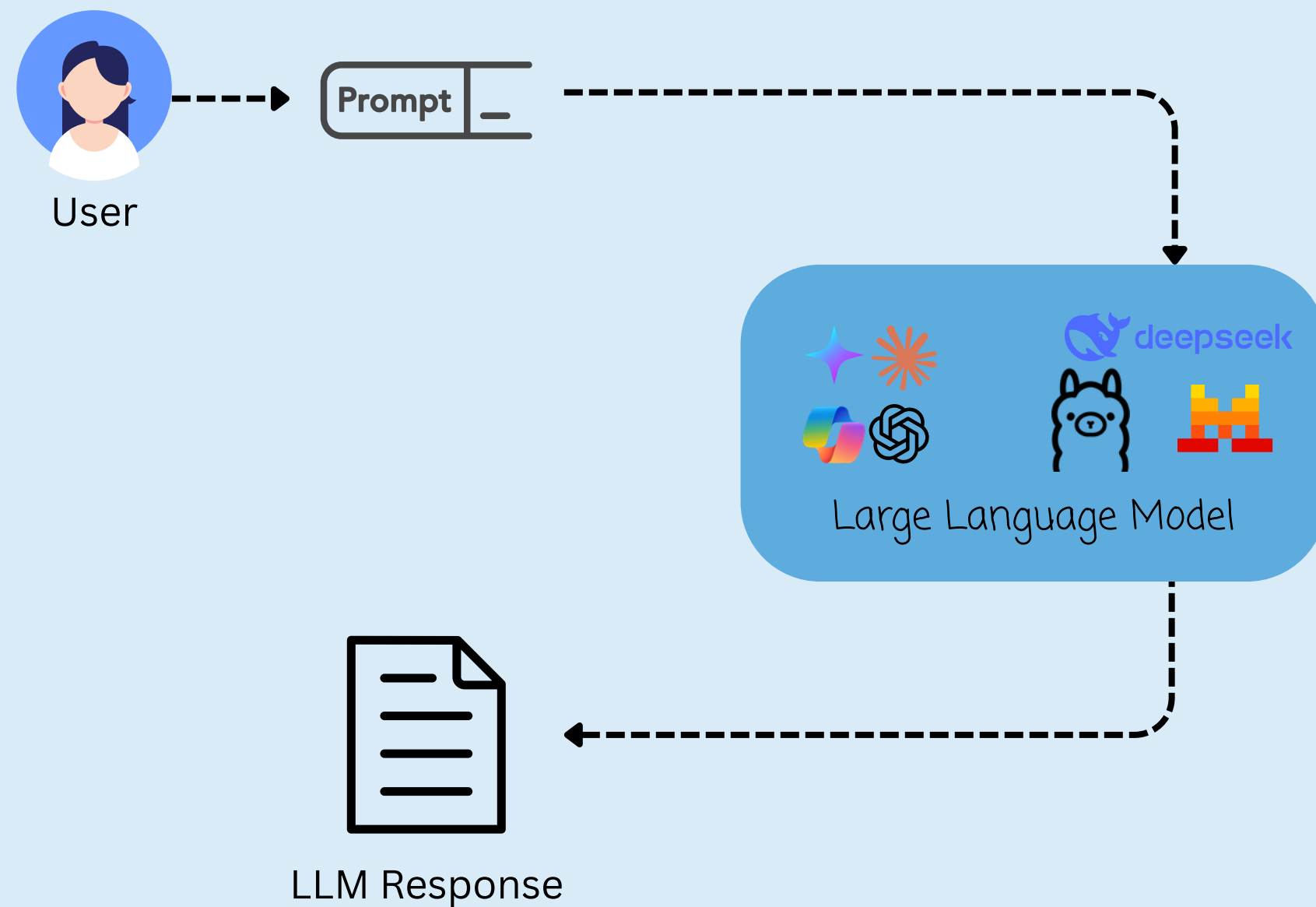
In questo luogo dice l'autore che esso era vizioso e peccatore. E la diritta via, ciò sono l'opere di Dio, aveva abbandonate e disprezzate.

Inf. I, 13 **Ma ppoi ch'io fu a ppiè d'un colle giunto**

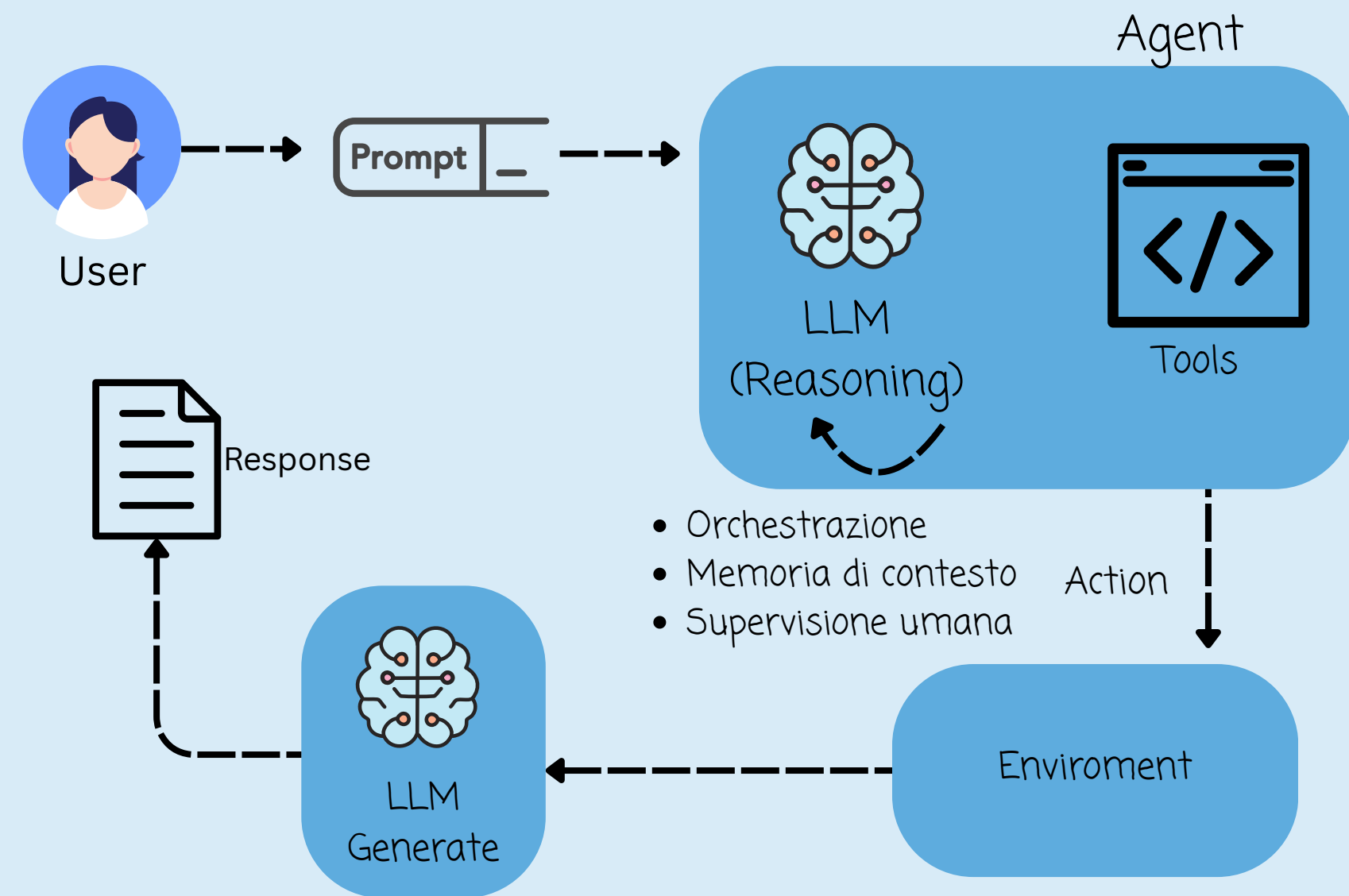
In questa parte vuol dire l'autore che apressandosi esso a le vertudi e cessandosi da' vizi, vide insù la sommità del monte gli splendori del sole, cioè considerando in sé medesimo, conobbe e' vizi e volsesi acostare a le virtù e a le cose buone. E però dice la Scrittura parlando di Dio:

PROPOSTA: AI AGENTICA?

Basic Responder



ReAct Pattern



PROPOSTA: AI AGENTICA?

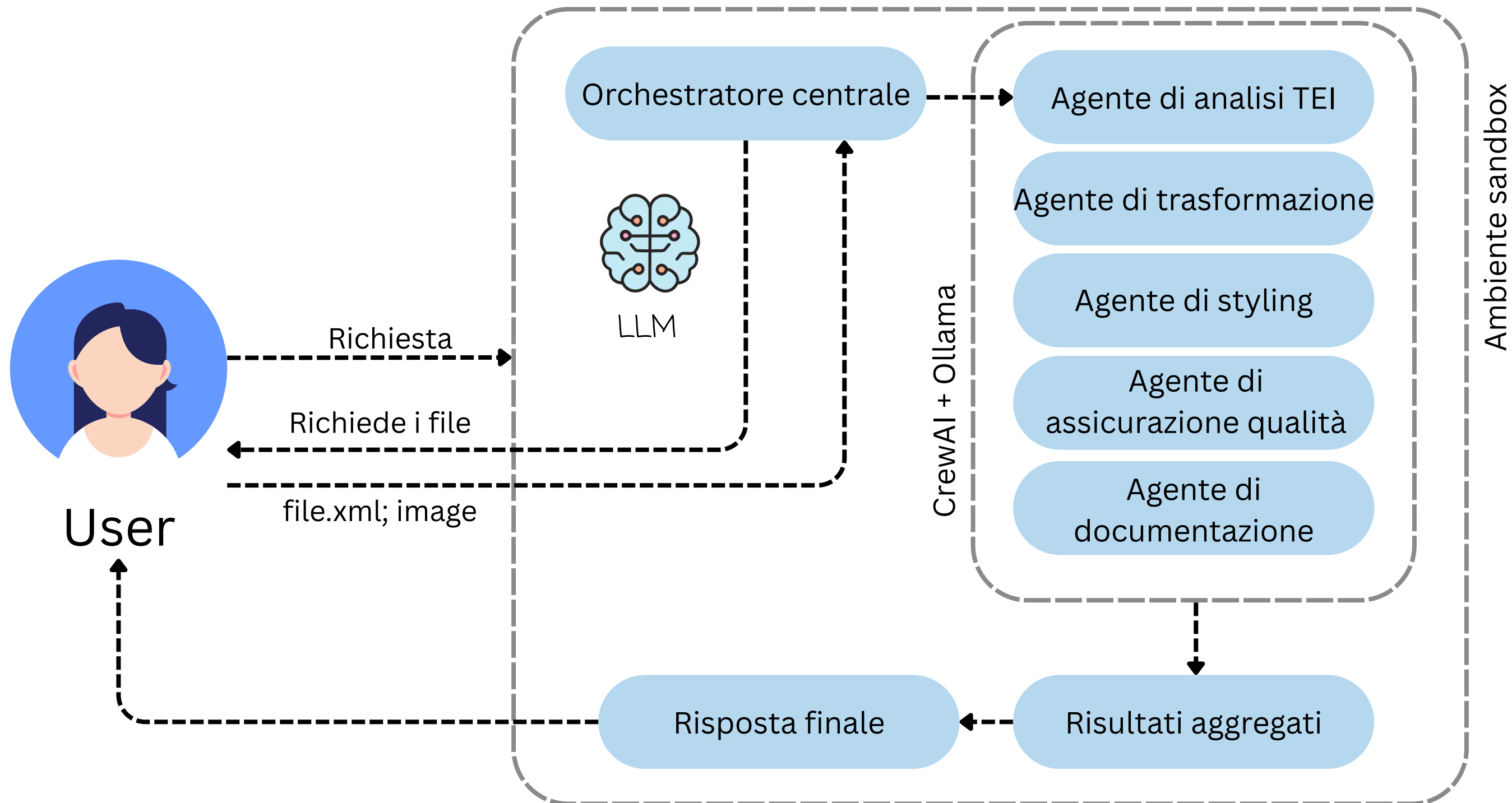
- **Vantaggi:**

- Efficienza e scalabilità: automazione di task ripetitivi e gestione di workflow complessi;
- Riduzione del carico per operazioni tecniche.

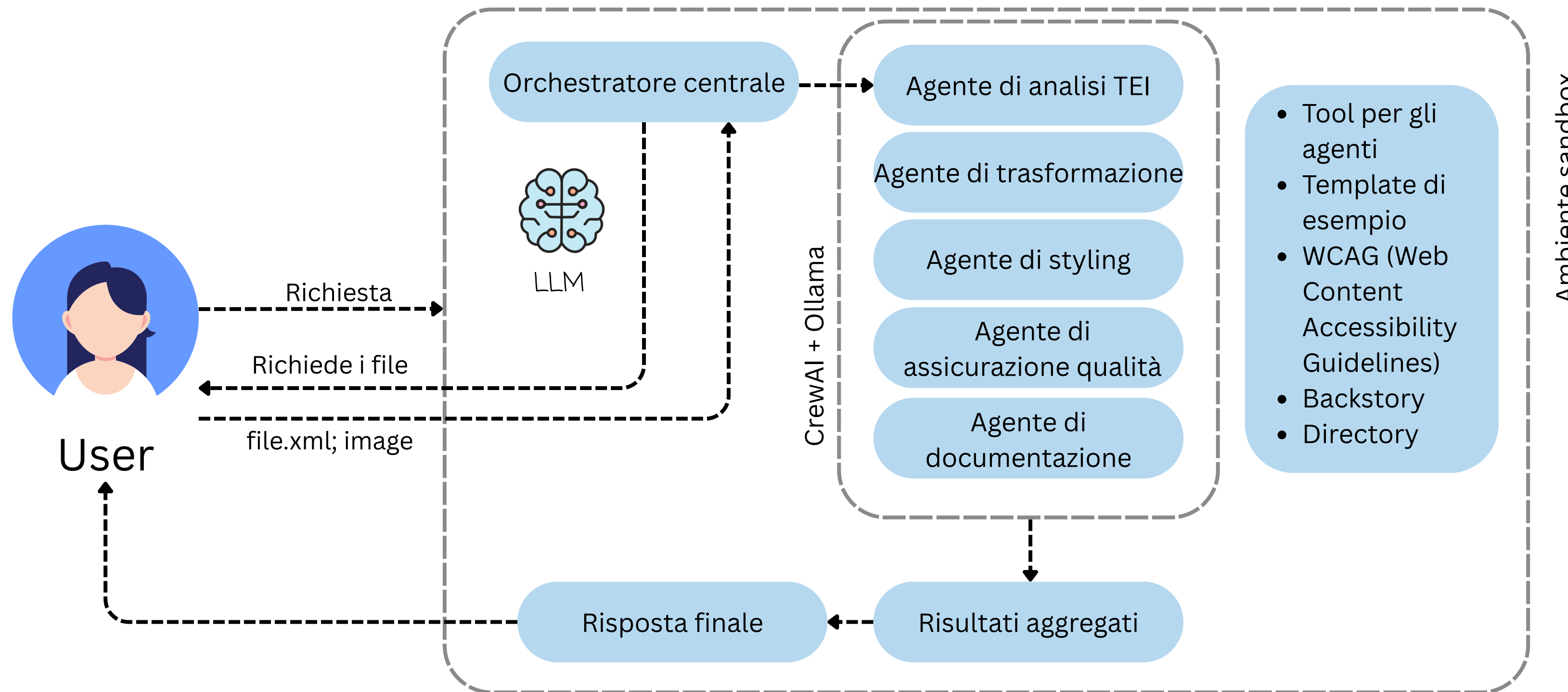
- **Svantaggi:**

- Sicurezza: consentire l'accesso root al sistema espone al rischio di operazioni distruttive;
- Opacità decisionale: mancanza di trasparenza nelle logiche interne;
- Dipendenza tecnologica: rischio di legarsi a tecnologie proprietarie e obsolete.

FRAMEWORK DI ORCHESTRAZIONE MULTI-AGENTE



FRAMEWORK DI ORCHESTRAZIONE MULTI-AGENTE



VISUALIZZAZIONE CLI (COMMAND LINE INTERFACE)

```
44 ## Installazione locale
45
46 ```bash
47 # 1. Ollama
48 curl -fsSL https://ollama.com/install.sh | sh
49 ollama pull qwen2.5:7b
50 ollama pull nomic-embed-text # per memoria semantica (opzionale)
51 ollama serve &
52
53 # 1.1 Ollama - Win Powershell
54 # 1. Installa Ollama
55 winget install Ollama.Ollama
56
57 # Riavvia il terminale dopo l'installazione se necessario
58
59 # Scarica i modelli
60 ollama pull qwen2.5:7b
61 ollama pull nomic-embed-text
62
63 # Avvia Ollama
64 ollama serve
65 ollama run qwen2.5:7b
66
67
68 # 2. Pacchetto
69 cd tei-editor-agentic
70 python -m venv .venv && source .venv/bin/activate
71 pip install -r requirements.txt
72
73 # 2. Vai nella cartella progetto
74 cd tei-editor-agentic
75
76 # Crea ambiente virtuale
77 python -m venv .venv
78
79 # Attiva ambiente virtuale
80 .\.venv\Scripts\Activate.ps1
81
82 # Installa dipendenze
83 pip install -r requirements.txt
84
85 # 3. Avvio
86 python run_editor.py # apre http://localhost:7860
87 python run_editor.py --headless # versione CLI senza Gradio
88 python run_editor.py --check # solo verifica setup
```

```
(.venv) PS C:\Users\user\tei-editor-agentic\tei-editor-agentic> py run_editor.py --headless
✓ Dipendenze installate
✓ Ollama attivo. Modelli disponibili: nomic-embed-text:latest, qwen2.5:7b, llama3.2:latest

Modalità CLI – chat con la crew CrewAI

Viewer pronto: C:\Users\user\tei-editor-agentic\tei-editor-agentic\workdir\tei-viewer
Modello Ollama [qwen2.5:7b]:
AGENTI DISPONIBILI:
- TEI/Data Agent
- UI Agent
- QA Agent
- Memory Agent

✓ Crew pronta con qwen2.5:7b. Scrivi 'esci' per terminare.

Tu > █
```

```
🔥 Crew Execution Started

Crew Execution Started
Name: crew
ID: fa6cf37b-873c-4abb-8532-8b7a99b06644

📌 Task Started

Task Started
Name: ## Contesto progetto stabile
{}

## Richiesta dell'utente
fai un check dei sistemi
```

VISUALIZZAZIONE GRADIO - WEB INTERFACE PER MODELLI MACHINE LEARNING



Configurazione

Modello Ollama

qwen2.5:7b

URL Ollama

http://localhost:11434

Abilita memoria CrewAI (ChromaDB)

Inizializza crew

Stato crew

✓ Crew pronta: modello=qwen2.5:7b, memoria=on

Upload file di progetto

Carica TEI XML

RomandAlexandre_Digitale_Uff.xml 302.4 KB ↓

Esito upload TEI

✓ Caricati in workdir/data/: RomandAlexandre_Digitale_Uff.xml

Carica immagini manoscritte

Ms_Correr_1493_003.jpg 380.6 KB ↓ ×

Ms_Correr_1493_002.jpg 368.8 KB ↓ ×

Ms_Correr_1493_001.jpg 303.2 KB ↓ ×

Esito upload immagini

✓ 3 immagini in workdir/data/immagini/

Memoria di progetto (YAML)

Modifica direttamente le preferenze persistenti del progetto.

project_memory.yaml

Memoria progetto vuota.
Verrà popolata al primo salvataggio.

TEI Editor Agentico

Chat con una crew CrewAI (Router + TEI/Data + UI + QA + Memory) che modifica in tempo reale il tuo sito di edizione critica. LLM locale via Ollama. Anteprima a destra, aggiornata a ogni reload.

Anteprima sito — <http://localhost:8765/>

Premi F5 nell'iframe (o ricarica) per vedere le modifiche dopo ogni risposta dell'agent.

Configurazione

Chat con l'editor agentico

Conversazione

Messaggio

es. 'Cambia il colore primario in rosso pompeiano' oppure 'Aggiungi un glossario per i termini occorrenti più di 3 volte'

Invia Pulisci

Upload file di progetto