



La ricerca sull'utilizzo dell'IA per la stima del rischio di contenzioso

AIDA: AI based Dispute Assistant (Tuscany Health Ecosystem, THE)

Paolo Nesi, DISIT Lab, UNIFI (paolo.nesi@unifi.it)

DISIT Lab <https://www.disit.org>

**I nuovi orizzonti della
responsabilità sanitaria**

14 Ottobre 2024

Aula Magna

Nuovo Ingresso Careggi

Obiettivo: valutazione dei rischio

- **Analisi del testo** dei documenti relativi al contenzioso, comprensione del linguaggio naturale,
- Per **rispondere a domande** che potrebbero essere quelle da porre ad un esperto che *studiando il caso* è in grado di *rispondere in relazione al contesto* della singola vertenza.
 - Per esempio
 - Vi sono altre strutture coinvolte?
 - Che tipo di rischio si ha?
 - Potrebbe essere conveniente....?
- **Tecniche:**
 - *Natural Language Processing, BERT, XAI, ...*
 - *LLM, Large Language Models: GPT*

Benefici Attesi

rispondere a domande come un esperto che *studiando il caso* è in grado di fornire suggerimenti

- **referiti** ad affermazioni localizzate nei documenti e fornire spiegazione
- con una **riduzione dei tempi** di analisi di una vertenza,
 - dal punto vista computazionale è instancabile
- **non influenzate** da precedenti documenti,
- **non affetti** da allucinazioni (deduzioni inventate)
- **Con minore varianza** dovuta alle variazioni di contesto, al quale gli umani sono sensibili.

→ *Decision Support System*

Come siamo arrivati al contesto attuale dell'AI (1/3)

- **Per molti anni l'approccio dell'IA alla comprensione del testo si è basata su modelli di NLP, **Natural Language Processing****
 - Tramite analisi lessico-grammaticali, modelli statistici, etc.
 - Per l'identificazione di pattern (di frasi) con sistemi a regole grammaticali, fondate su conoscenza codificata delle regole del linguaggio.
 - Per esempio, codificate tramite ontologie, tassonomie (e.g., SKOS):
 - Per riconoscere intensioni, contesti, luoghi, disambiguare, etc.
 - Sono nati i primi Virtual Assistant:
 - *Watson, Siri, Cortana, Alexa, Paval (UNIFI), etc.*
- **Recentemente sono arrivati modelli di apprendimento: Deep Learning, come Recurrent NN, RNN:**
 - Questi con un numero elevato di strati di Neuroni possono apprendere sequenze complesse di simboli/parole, codificate, etc.

La rivoluzione dei Transformers (2017) (2/3)

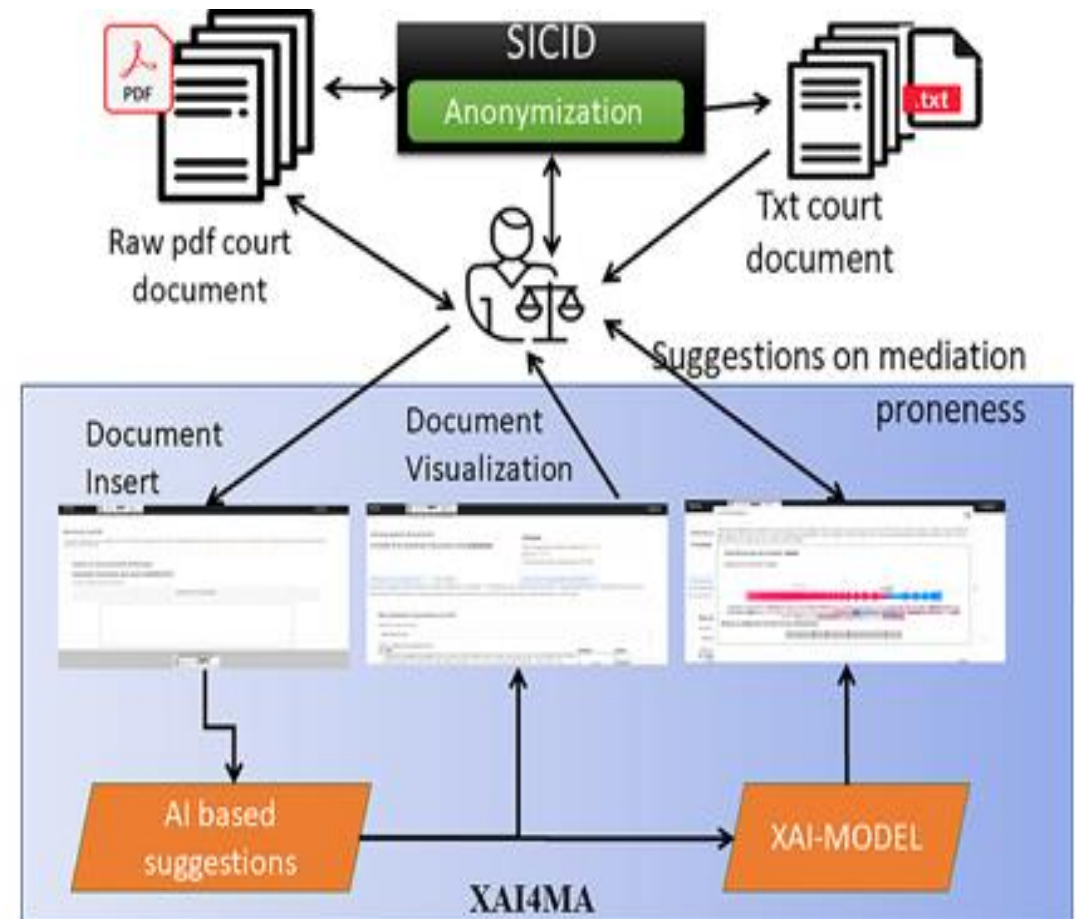
- Hanno sostituito le RNN andando a introdurre il meccanismo di «Attention» per pesare la rilevanza delle parole nelle frasi
- **BERT (Bidirectional Encoder Representations from Transformers) (2018)**,
 - BERT: uno dei primi modelli basati su Transformers a ottenere risultati straordinari per i classici compiti di NLP:
 - riconoscimento, classificazione, disambiguazione, etc.
- **GPT (Generative Pretrained Transformer):** introdotto da OpenAI
 - tra i primi a passare a modelli di auto-regressione per generare testo.
 - **Generazione di risposte in linguaggio naturale**
 - La generazione di risposte, *una sorta di predizione a partire da una sequenza data di parole, il prompt, la richiesta.*

Da BERT a LLM (3/3)

- I LLM, Large Language Model sono l'evoluzione dei **GPT** (e.g., chat **GPT-4**), e si distinguono per la loro scala in termini di:
 - **Numero di parametri:** i LLM hanno bilioni di parametri, per catturare informazioni molto più complesse e dettagliate sul linguaggio.
 - **Dimensione del dataset:** i LLM sono addestrati su dataset enormi, spesso contenenti centinaia di Gbyte-TByte di testo provenienti da varie fonti (libri, articoli, pagine web).
 - **Capacità zero-shot e few-shot:** i LLM, grazie alla loro scala, possono eseguire compiti complessi:
 - zero: produzione di risposte sulla base della loro conoscenza di base
 - few: produzione di risposte sulla base di conoscenza aggiuntiva
 - **Alcune architetture di LLM continuano a imparare** ma magari non è sempre cosa buona, anche perché si rischia di virare il modello, e dimenticare...

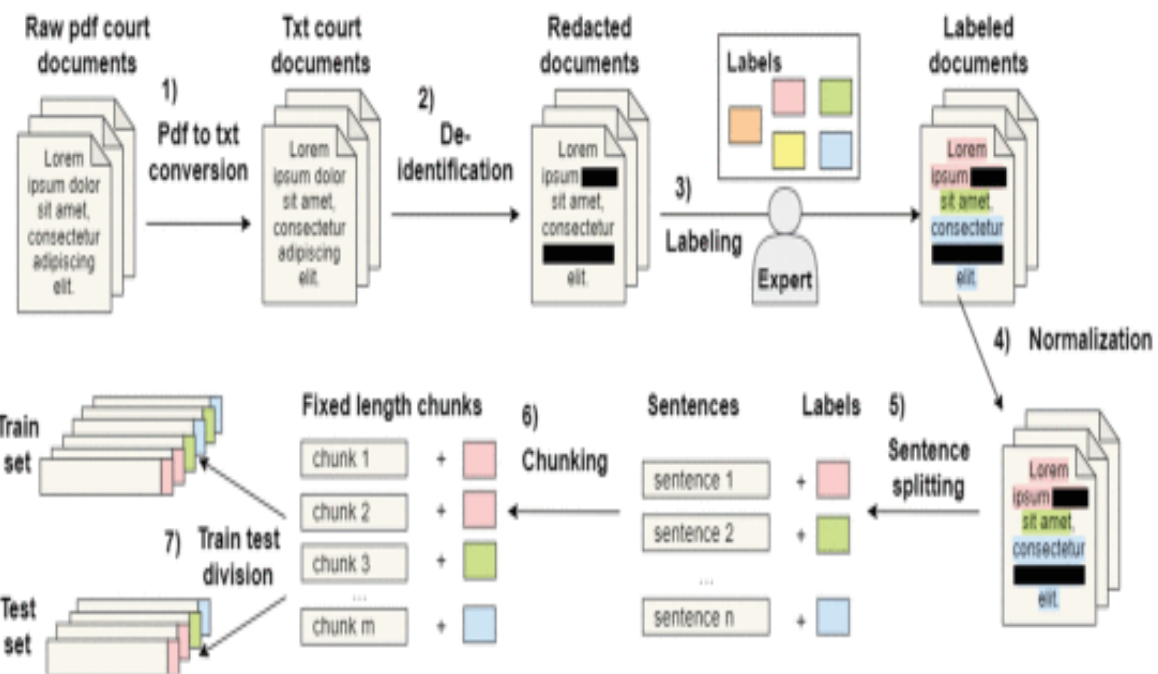
Giustizia Semplice/Agile (PON)

- Come DISIT lab abbiamo sviluppato soluzioni per la valutazione della propensione alla mediazione (in collaborazione con il dip. di Giurisprudenza di UNIFI e del Tribunale di Firenze)
 - Fornito come Assistente accessibile da una pagina web.
- E. Collini, P. Nesi, C. Raffaelli, F. Scandiffio, "**Explainable Artificial Intelligence for Agile Mediation Propensity Assessment**", IEEE Access, pp.1-7, 2024, Print ISSN: 2169-3536, Online ISSN: 2169-3536, DOI: 10.1109/ACCESS.2024.3375766
<https://ieeexplore.ieee.org/document/10464317>



Modello di Comprensione

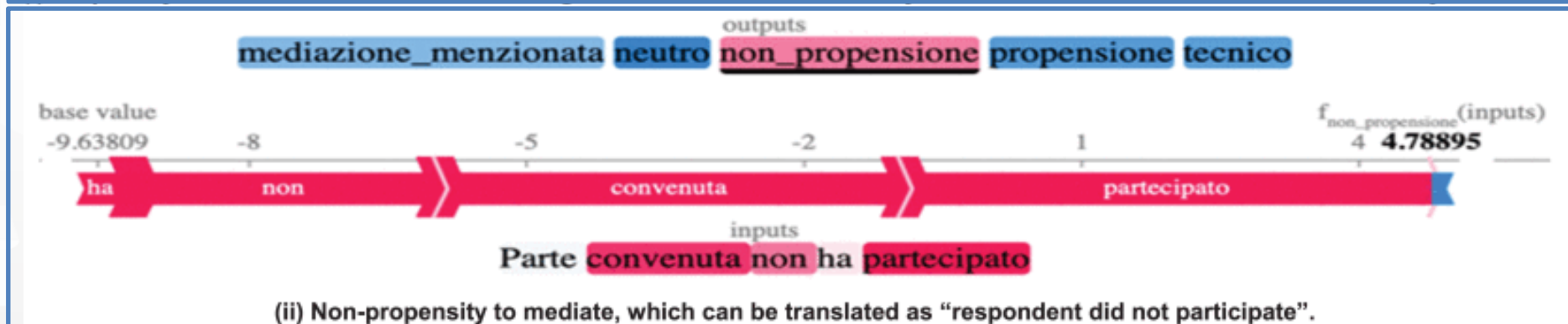
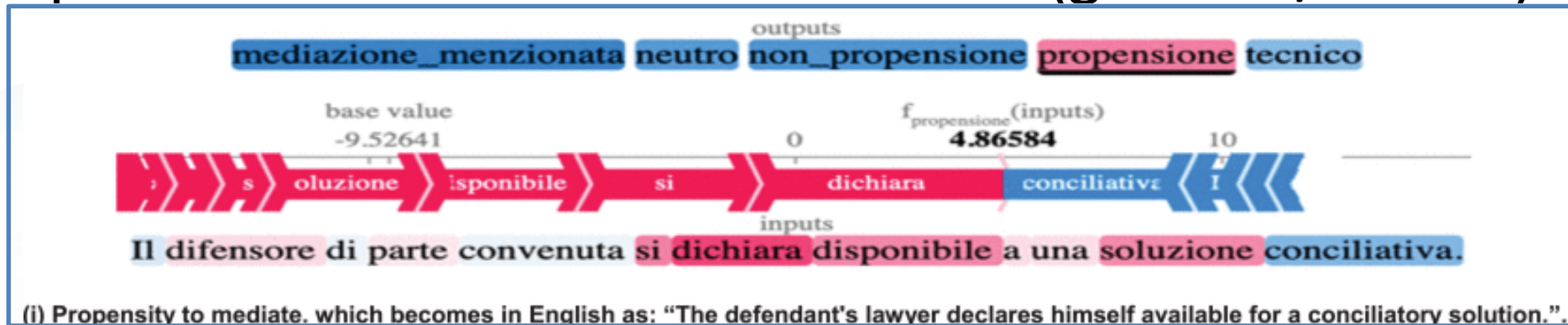
- Collezione dei documenti di un caso
 - Preelaborazione
 - Estrazione delle frasi
- Classificazione delle frasi con tecniche BERT, Transformer



classes	Precision	Recall	F1-score
Propensity to Mediate	1.000	0.923	0.956
Not propensity to Mediate	0.500	1.000	0.667
Neutral	0.959	0.989	0.974
Technician involved	1.000	0.915	0.956
Mentioned mediation	0.944	0.864	0.903
Global	0.958	0.950	0.952

Explainable AI, global/local

- In passo ulteriore è fornire non solo una classificazione ma anche una spiegazione della classificazione delle singole frasi.
- Per questo si utilizzano tecniche di XAI (global e/o local)



AIDA: AI based Dispute Assistant (Tuscany Health Ecosystem)

- **Obiettivo: valutazione dei rischio**
 - **Analisi del testo** dei documenti relativi al contenzioso, comprensione del linguaggio naturale,
 - Per **rispondere a domande** che potrebbero essere quelle da porre ad un esperto che **studiando il caso** è in grado di **rispondere in relazione al contesto** della singola vertenza.
 - Per esempio
 - Vi sono altre strutture coinvolte?
 - Che tipo di rischio si ha?
 - Potrebbe essere conveniente....?
 - **Tecniche:**
 - *Natural Language Processing, BERT, XAI, ...*
 - *LLM, Large Language Models: GPT/RAG Model*

Approccio BERT

- **Obiettivo:**
 - Classificazione di pattern in frasi per la comprensione dei documenti
- **Creazione di un data set di training** significativo e bilanciato
- **Identificazione di possibili modelli BERT per l'italiano** dai quali partire per fare un fine tuning:
 - **Training/addestramento** per tenere conto del contesto legale specifico
 - Fine Tuning con **~4000 frasi, 12 classi** (del data set di training)
- **Valutazione di 8 modelli BERT, pre-addestrati per l'italiano**
 - **per insegnarli a rispondere in modo corretto sulla classificazione.**
 - **BERT Fine Tuned per gli aspetti legali**
 - Totale **circa 240 processi di fine-tuning** su modelli e parametri diversi.
 - Valutazione e verifiche su ogni modello prodotto per **identificare il migliore in modo oggettivo e sistematico.**

Modello BERT Fine Tuned migliore:

Accuratezza sulle singole classi:

AEDOP	• assenza_errore_difetto_organizzativo_procedura:	92.0%
CD	• carenza_documentale:	84.6%
COIN	• coinvolgimento_altre_strutture:	94.4%
DAP	• danno_alla_persona:	96.0%
DM	• danno_morte:	88.5%
DCI	• difetto_consenso_informato:	80.9%
EDOP	• errore_difetto_organizzativo_procedura:	90.6%
NEUTRA	• neutra:	90.7%
PCI	• presenza_consenso_informato:	95.8%
VE	• valore_economico:	99.9%

- **F1 multiclasse: 93,99%**

Esplorazione dei risultati

Premendo il tasto Esplora nell'ultima colonna di una riga è possibile visualizzare una descrizione XAI (explainable AI) della classificazione. Questa rappresenta graficamente la confidenza di appartenenza di un documento ad una classe.

Un singolo documento viene suddiviso in porzioni di testo disgiunte, per ognuna delle quali viene determinata una confidenza.

E' possibile interagire con le valutazioni del modello, in modo da poter visualizzare quali segmenti di testo hanno influenzato maggiormente la classificazione.

NOME DOCUMENTO	AEDOP	CD	COIN	DAP	DM	DCI	EDOP	NEUTRA	PCI	VE	ESPLORA RISULTATI
Internamento San Luca	0.11	0.04	0.49	0.08	0.04	0.02	0.08	0.12	0.03	0.01	Esplora

Visualizzazione dei risultati, AIDA zero

Permette il caricamento dei documenti (PDF/TXT) di ogni vertenza e li elabora fornendo la classificazione delle frasi, e dei documenti

La tabella **(A)** mostra con quanta confidenza è stata assegnata una classe ad ogni documento. Ad ogni riga corrisponde un documento, ad ogni colonna la confidenza su quella classe. *(nella prossima slide la legenda)*. I valori oltre il 25% di confidenza sono marcati in verde.

La tabella **(B)** mostra con quanta confidenza è stata classificata l'intera vertenza.

Risultati per: vertenza Maria Rossi

Premi sul bottone per visualizzare la legenda

[Visualizza legenda](#)

(A)

NOME DOCUMENTO	AEDOP	CD	COIN	DAP	DM	DCI	EDOP	NEUTRA	PCI	VE	ESPLORA RISULTATI
Internamento San Luca	0.11	0.04	0.49	0.08	0.04	0.02	0.08	0.12	0.03	0.01	Esplora
MariaRossiOcchiTOPDF	0.04	0.02	0.12	0.17	0.01	0.01	0.03	0.02	0.01	0.01	Esplora

Classificazione vertenza

VERTENZA	AEDOP	CD	COIN	DAP	DM	DCI	EDOP	NEUTRA	PCI	VE
vertenza Maria Rossi	0.08	0.03	0.30	0.12	0.02	0.01	0.06	0.07	0.02	0.01

(B)

Local
Explainability
of the AI
MODEL

Valutazione	Classificazione
<p>outputs</p> <p>assenza_errore_difetto_organizzativo_procedura convolgimento_altre_strutture danno_alla_persona danno_morte difetto_consenso_informato errore_difetto_organizzativo_procedura neutra presenza_consenso_informato valore_economico</p> <p>base value 0.0364479 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 f_convolgimento_altre_strutture (input) 0.893235</p> <p>inputs</p> <p>La paziente è stata ricoverata in data 14/05/2017 presso il policlinico Ospedale San Luca, Via Guglielmo Lippi Francesconi, 556, 55100 Lucca LU. Sono stati condotti esami di visita oculistica quali</p> <p>✓ Convalida Annota</p>	<p>convolgimento_altre_strutture</p> <p>annotazione: Non ancora fornita</p>
<p>outputs</p> <p>assenza_errore_difetto_organizzativo_procedura carenza_documentale convolgimento_altre_strutture danno_alla_persona danno_morte difetto_consenso_informato errore_difetto_organizzativo_procedura neutra presenza_consenso_informato valore_economico</p> <p>base value f_assenza_errore_difetto_organizzativo_procedura (inputs) 0.0174445 0.1 0.208247 0.3 0.4 0.5 0.6 0.7 0.8</p> <p>inputs</p> <p>Topografia Corneale, OCT, Biometria e fluorangiografia.</p> <p>✓ Convalida Annota</p>	<p>neutra</p> <p>annotazione: Non ancora fornita</p>

Possibile valutare: Convalidare o Correggere una classificazione

www.snap4city.org says
Hai convalidato la classificazione della frase

OK

assenza_errore_difetto_organizzativo_procedura_carenza_documentale **conivolgimento_altre_strutture**
danno_alla_persona danno_morte_difetto_consenso_informato_errore_difetto_organizzativo_procedura_neutra
presenza_consenso_informato_valore_economico

base value 0.0364479 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.893235
f_{conivolgimento_altre_strutture}(input)

inputs
La paziente è stata **ricoverata** in data 14/05/2017 **presso il policlinico Ospedale San Luca, Via Guglielmo Lippi Francesconi, 556, 55100 Lucca LU. Sono stati condotti esami di visita oculistica quali**

Classificazione

conivolgimento_altre_strutture

annotazione: Non ancora fornita

✓ Convalida ✎ Annota

Al momento

- **E' accessibile una versione di AIDA che viene usata via WEB da personale per una validazione e valutazione**
 - Completare la validazione con documenti mai visti di vertenze di varia tipologia
 - Passare ad una valutazione approfondita con gli esperti di Medicina Legale, anche su vertenze complete
- **Stiamo completando una versione basata su tecnologia GPT/LLM che permetterà la formulazione diretta di domande nel contesto dei documenti di ogni vertenza che è stata fornita al sistema.**