# AI Experiences @ Italian Senate

## A concise summary

Carlo Marchetti, PhD
Head of Information Systems Development

JURIX
The Foundation for Legal Knowledge Based Systems

# AI @ Work...

- “Narrow” AI
  - Speech to text - reporting
  - Cybersecurity
  - Translation
    - Mainly products!

- AI in dev projects
  - Amendments
    - editing, ordering, similarity, correction, explanation/correlation*
  - Drafting
    - XML Editor for reviews, leg. references and features*, “testo a fronte”*
  - UIP: feature extraction (e.g. statements in reports), document classification
  - Virtual assistants / chatbots:
    - website assistant, “semantic” search assistant

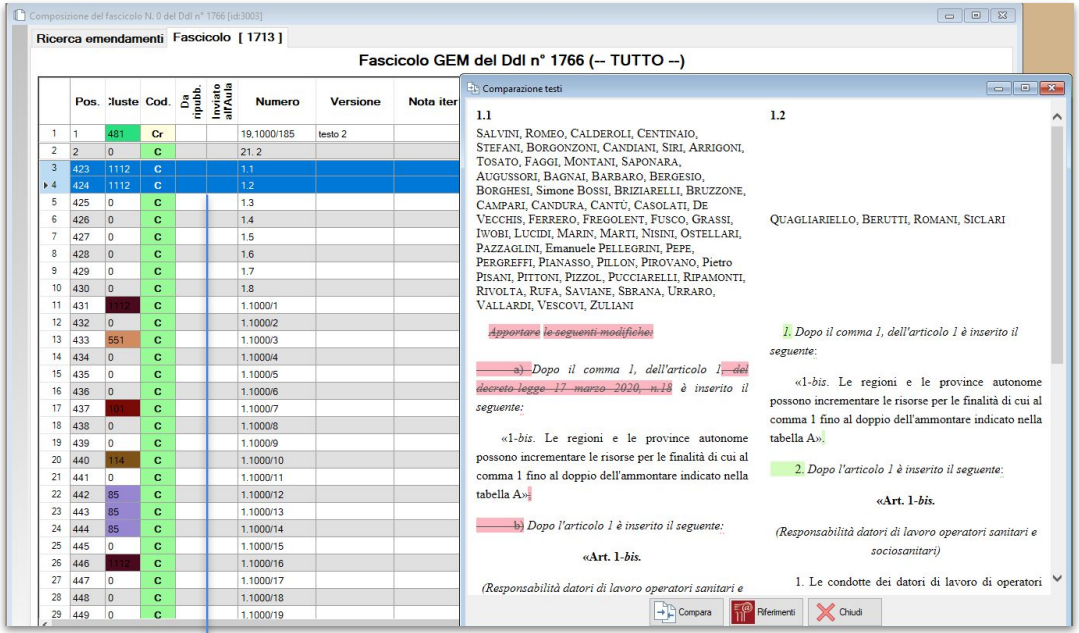
# Amendments & AI

- GEM is a tool for managing amendments in digital format

- Offices access functionalities for:
  - importing amendments from sources
  - ordering a set of amendments
    - by identifying the point in the bill where the amendment has some effects
  - find similar amendments
    - to avoid repeating the same modification in a bill twice or more
  - correct/modify/compare/check effects

- Algorithms based mainly on open source NLP libraries
  - text analysis, vectorization, cosine distance

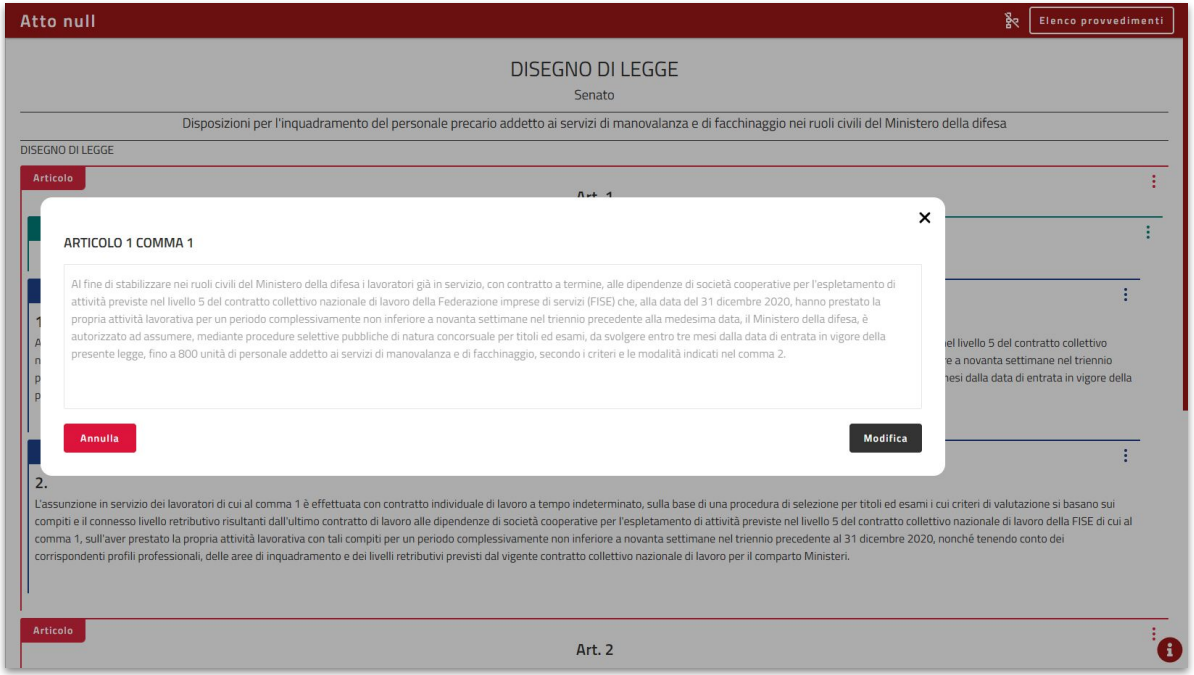
- genAI used only in experimental ways
  - check/suggest corrections/ apply to bill




The system finds similar amendments and diffs them

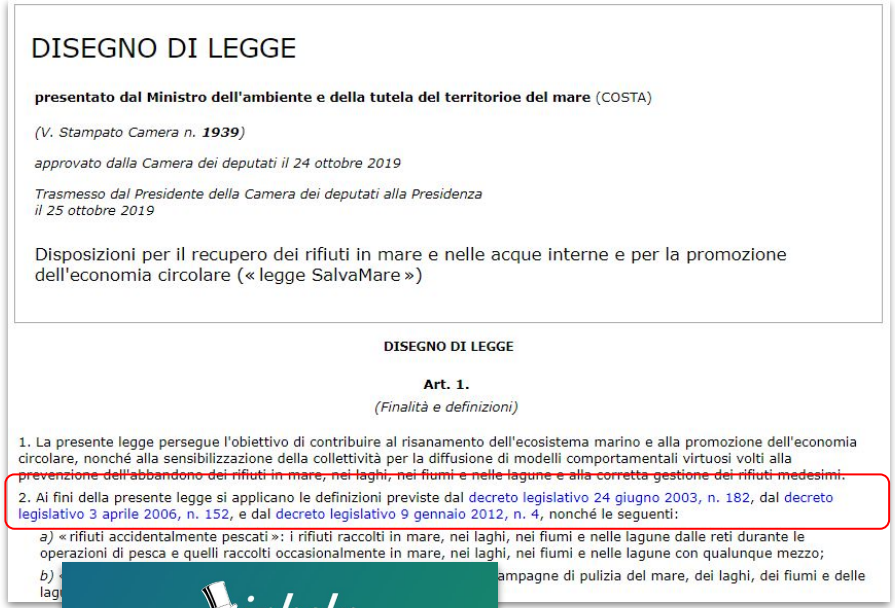
# “Amendments machine” (Machina)

- Project started within the “Polo informatico parlamentare”, i.e. jointly with Chamber of Deputies
- Editor for drafting amendments using a track-change approach
- Amendments are output automatically, according to syntactic rules that allow for immediate and deterministic processing
- Advantages for users and for Offices
- Currently under experimentation in Committees



# Improving Drafting & reviews: finding references

- Offices spend time reviewing acts
- Correctness of legislative references/citations is a relevant and time-consuming activity
- Together with CNR, we developed a tool that analyzes whatever text (.txt, .doc, .htm, .xml) and outputs a text in which legislative references are:
  - identified
  - marked up in xml
  - linked to a URN that can be translated in a URL
- Integration with Normattiva&other leg. Docbases allows for improved reviews and speed up



DISEGNO DI LEGGE

**presentato dal Ministro dell'ambiente e della tutela del territorioe del mare** (COSTA)

*(V. Stampato Camera n. **1939**)*

*approvato dalla Camera dei deputati il 24 ottobre 2019*

*Trasmesso dal Presidente della Camera dei deputati alla Presidenza il 25 ottobre 2019*

Disposizioni per il recupero dei rifiuti in mare e nelle acque interne e per la promozione dell'economia circolare (« legge SalvaMare »)

**DISEGNO DI LEGGE**

**Art. 1.**

*(Finalità e definizioni)*

1. La presente legge persegue l'obiettivo di contribuire al risanamento dell'ecosistema marino e alla promozione dell'economia circolare, nonché alla sensibilizzazione della collettività per la diffusione di modelli comportamentali virtuosi volti alla prevenzione dell'abbandono dei rifiuti in mare, nei laghi, nei fiumi e nelle lagune e alla corretta gestione dei rifiuti medesimi.

2. Ai fini della presente legge si applicano le definizioni previste dal decreto legislativo 24 giugno 2003, n. 182, dal decreto legislativo 3 aprile 2006, n. 152, e dal decreto legislativo 9 gennaio 2012, n. 4, nonché le seguenti:

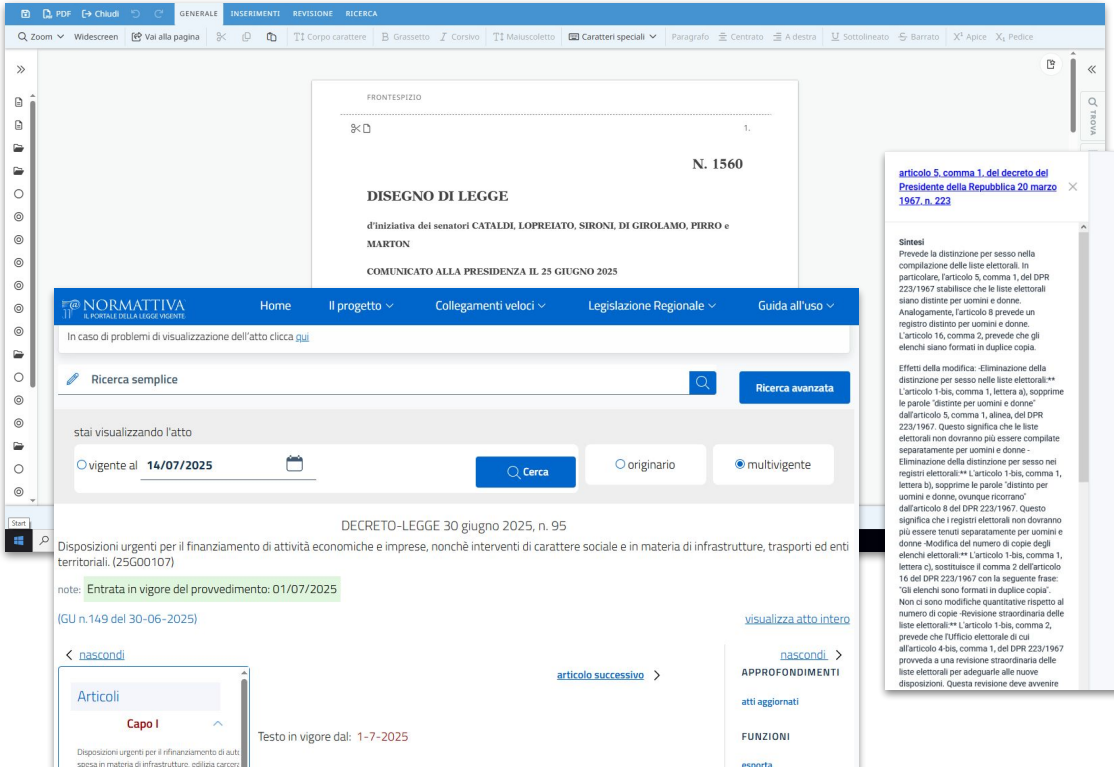
*a)* « rifiuti accidentalmente pescati »: i rifiuti raccolti in mare, nei laghi, nei fiumi e nelle lagune dalle reti durante le operazioni di pesca e quelli raccolti occasionalmente in mare, nei laghi, nei fiumi e nelle lagune con qualunque mezzo;

*b)* ... ampagne di pulizia del mare, dei laghi, dei fiumi e delle lag...

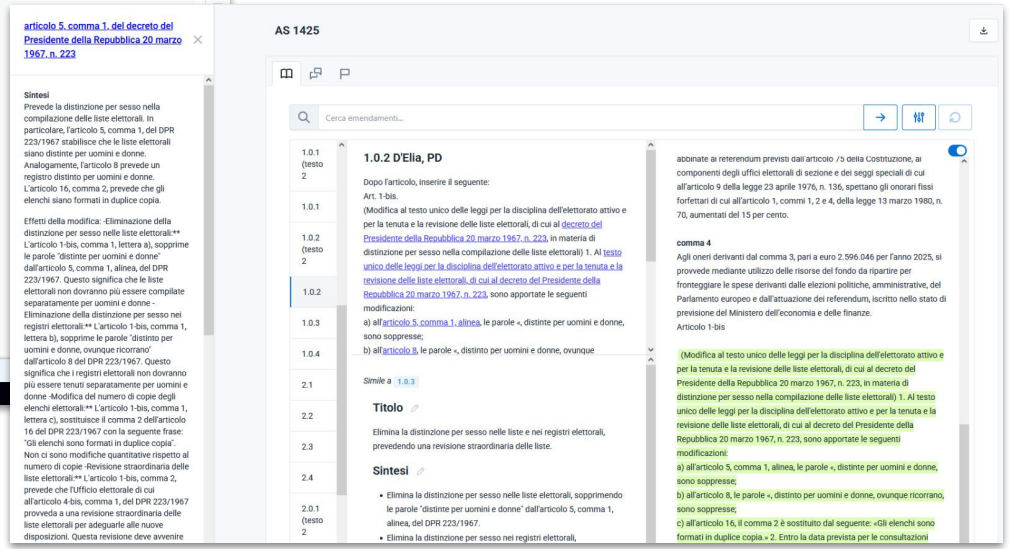
The tool is available as open source at: http://ittig.github.io/Linkoln/

# Integrations and genAI evolutions

Integrations: editors, normattiva etc...




explanation/correlation using genAI (e.g., Asimov)

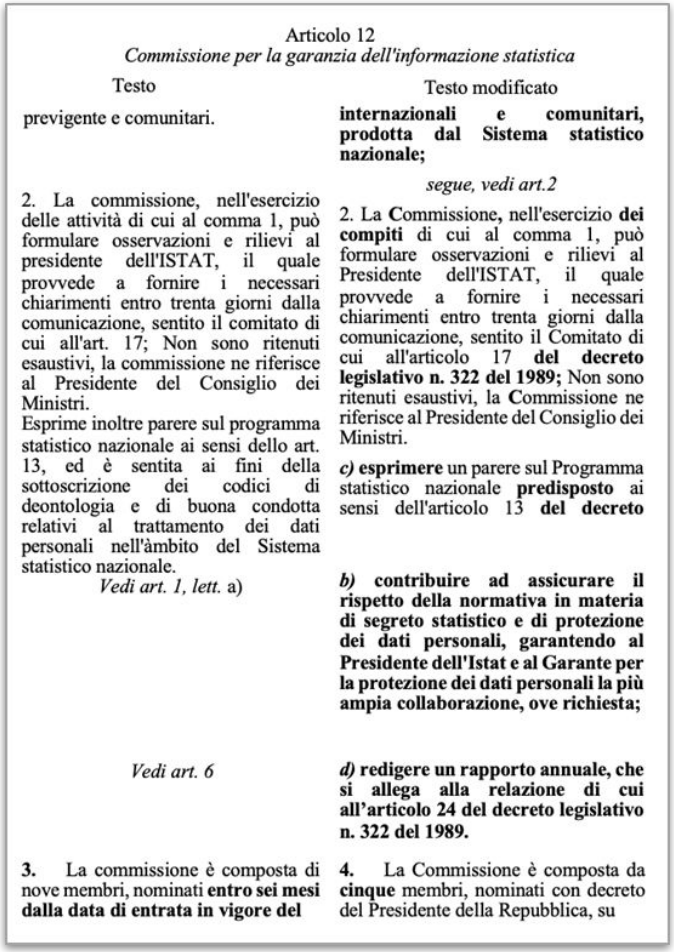


# “Testo a fronte”

- 2009: deterministic algorithm for diffing XML bills (with Univ. of Bologna)
  - requires structured text in any moment of the workflow

- Different scenario:
  - experimenting genAI-based tool for applying amendments to raw texts

  - complex set of rules > complex prompt > complex testing



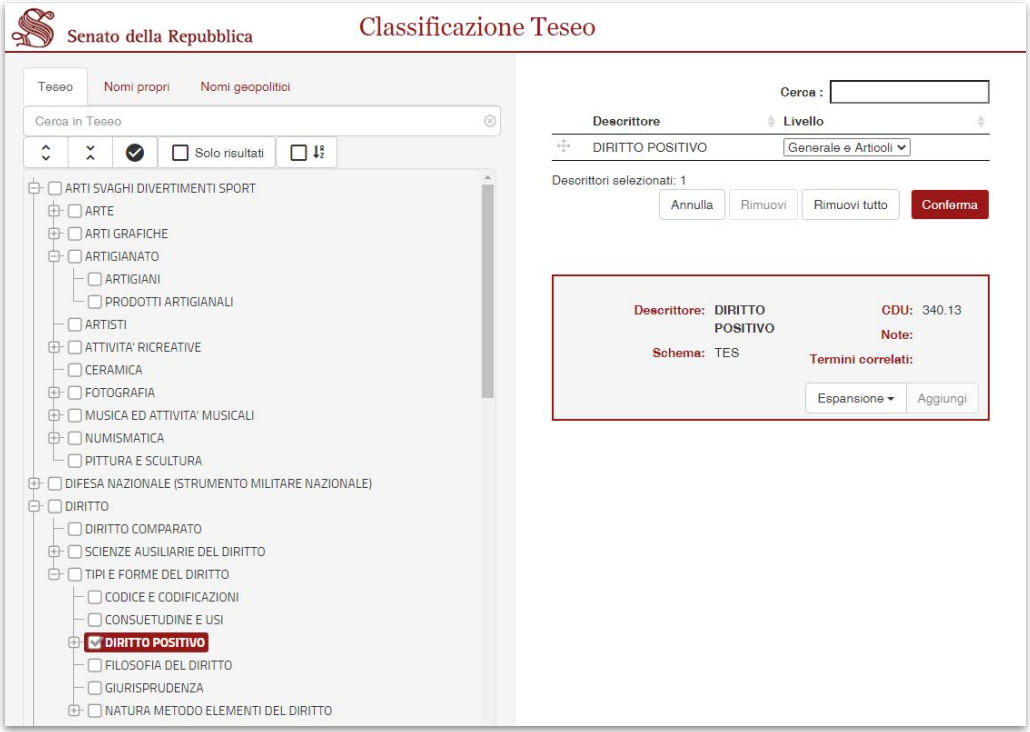
Articolo 12

*Commissione per la garanzia dell'informazione statistica*

| Testo | Testo modificato |
| --- | --- |
| previgente e comunitari. | **internazionali e comunitari, prodotta dal Sistema statistico nazionale;** |
| | *segue, vedi art.2* |
| 2. La commissione, nell'esercizio delle attività di cui al comma 1, può formulare osservazioni e rilievi al presidente dell'ISTAT, il quale provvede a fornire i necessari chiarimenti entro trenta giorni dalla comunicazione, sentito il comitato di cui all'art. 17; Non sono ritenuti esaustivi, la commissione ne riferisce al Presidente del Consiglio dei Ministri. Esprime inoltre parere sul programma statistico nazionale ai sensi dello art. 13, ed è sentita ai fini della sottoscrizione dei codici di deontologia e di buona condotta relativi al trattamento dei dati personali nell'àmbito del Sistema statistico nazionale. *Vedi art. 1, lett.* a) | 2. La **C**ommissione**,** nell'esercizio **dei compiti** di cui al comma 1, può formulare osservazioni e rilievi al Presidente dell'ISTAT, il quale provvede a fornire i necessari chiarimenti entro trenta giorni dalla comunicazione, sentito il Comitato di cui all'articolo 17 **del decreto legislativo n. 322 del 1989;** Non sono ritenuti esaustivi, la **C**ommissione ne riferisce al Presidente del Consiglio dei Ministri. *c)* **esprimere** un parere sul Programma statistico nazionale **predisposto** ai sensi dell'articolo 13 **del decreto** |
| | ***b)* contribuire ad assicurare il rispetto della normativa in materia di segreto statistico e di protezione dei dati personali, garantendo al Presidente dell'Istat e al Garante per la protezione dei dati personali la più ampia collaborazione, ove richiesta;** |
| *Vedi art. 6* | ***d)* redigere un rapporto annuale, che si allega alla relazione di cui all'articolo 24 del decreto legislativo n. 322 del 1989.** |
| **3.** La commissione è composta di nove membri, nominati **entro sei mesi dalla data di entrata in vigore del** | **4.** La Commissione è composta da **cinque** membri, nominati con decreto del Presidente della Repubblica, su |

# Bill classification & ML

- Every article of each bill is classified according to a Thesaurus comprising more than 3.500 labels
- This is a time-consuming activity performed since the early 90s
- Therefore - together with Univ. of Roma 3 - we trained a AI system to learn the classification criteria using our bulk
  - we initially used Comprehend by AWS and then we moved to a machine learning BERT-like RNN
- 80% of correct suggestions + product for supporting semi-automatic bill classification



# Feature extraction & video indexing

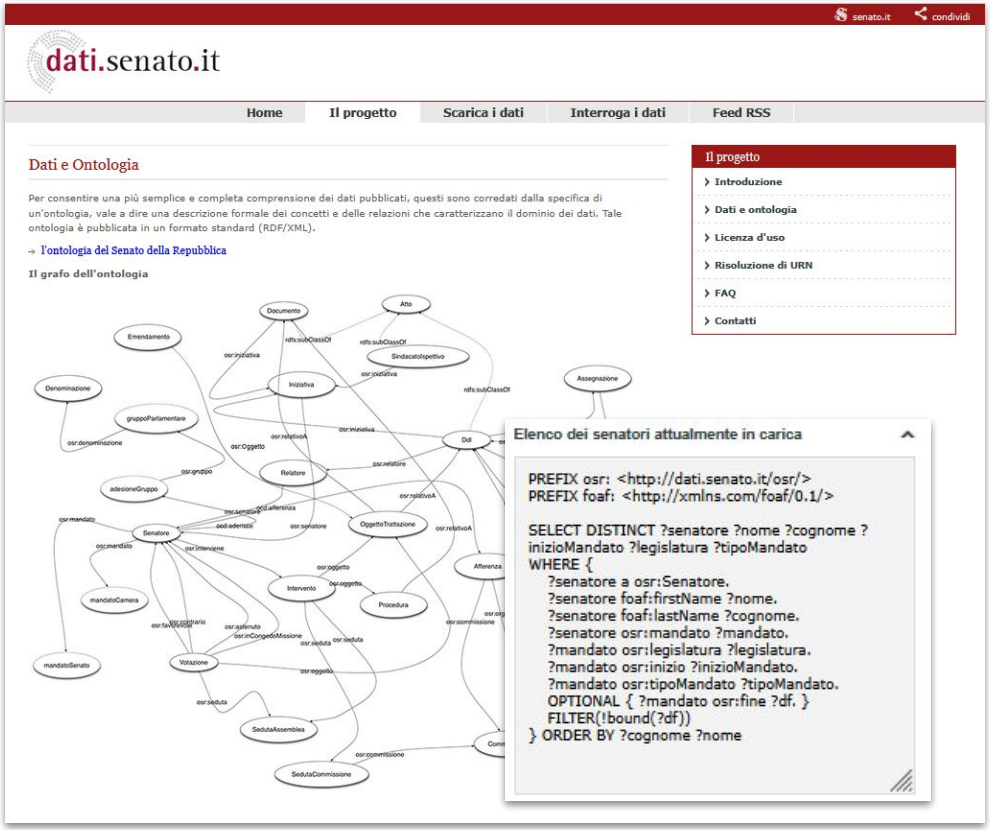
- Extracting features from “raw” text is used to boost productivity and service levels
- The structure of XML reports can be “aligned” with speech2text using AI, achieving synchronization among audio, transcription and official reports
- This allows, afterall, to synchronize video and reports
- Algorithms of sequence alignment developed with Univ. of Roma3

# Open data and open documents

- Since 2012, the Senate publishes millions of elementary data in RDF format (triples) on [dati.senato.it](dati.senato.it)

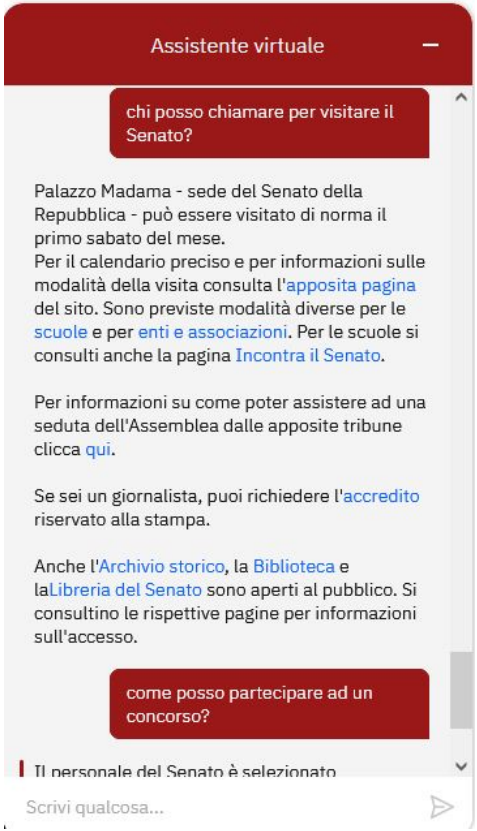
- querying RDF is not an easy task: we receive many support request from users

- testing genAI to produce SPARQL query from natural language + ontology



# Virtual Assistants - Chatbots- Semantic search

- Complexity of information (e.g. on websites) sometimes makes life of users very hard
  - it's not easy to find answers to simple questions: where's a bill, what's a Senator doing, how attend to sittings, etc.
- Chatbots are a simple way of providing a first-level user orientation
- No genAI: Training can be done w/o any specific IT knowledge, starting from a simple knowledge base, e.g. FAQs
  - drawback: high maintenance cost
- System online on FAQ pages
  - based on Intents.
- Embeddings are used internally to integrate assistants with semantic search capabilities



# Concluding remarks

- AI is a very large set of tools. Some have known risks (bias, statistical outcomes, hallucinations, knowledge-cutoff, etc)
  - evaluating costs/risks/benefits is not an easy task.

- AI solutions & humans: the “human-in-the-loop/command” patterns offer guarantees.
  - What is the cost?

- genAI:
  - sometimes the cost of testing and reviewing results is very high.

- Issues worth considering when using AI:
  - AI products are sometimes “black boxes”: they work, somehow, but it’s hard/impossible to express service levels in a clear way: test, test, test…
  - … and genAI is not a panacea!: even when the box is transparent it’s impossible to understand what happens inside!
  - steep learning curve and high resources consumption (experts & money & time) to put solutions in place.
  - super-fast pace of innovation: the market is evolving rapidly, and it could happen to invest in something that gets “old” in a few months.

# References


AGNOLONI, T. – MARCHETTI, C. – BATTISTONI, R. – BRIOTTI, G., Clustering Similar Amendments at the Italian Senate, in Proceedings of LREC 2022, Workshop ParlaCLARIN III, link: http://www.lrec-conf.org/proceedings/lrec2022/workshops/ParlaCLARINIII/pdf/2022.parlaclariniii-1.7.pdf

BACCI, L. – AGNOLONI, T. – MARCHETTI, C. – BATTISTONI, R., Improving Public Access to Legislation Through Legal Citations Detection: The Linkoln Project at the Italian Senate, disponibile al link: https://ebooks.iospress.nl/volumearticle/51783

BERTILLO, D. – DE DONATO, A. – MARCHETTI, C. – MERIALDO, P., Enhancing Accessibility of Parliamentary Video Streams: AI-Based Automatic Indexing Using Verbatim Reports, in Proceedings of LIRAI 2023: First Workshop on Legal Information Retrieval meets Artificial Intelligence, Roma, 4-8 settembre 2023, pp. 31-40, link: https://ceur-ws.org/Vol-3594/paper2.pdf

CIANCARINI, P. – DI IORIO, A. – MARCHETTI, C. – SCHIRINZI, M. – VITALI, F., Bridging the gap between tracking and detecting changes in XML, in Software: Practice and Experience, 46 (2), 2016, pp. 227-250.

DE ANGELIS, A. – DI CICCO, V. – LALLE, G. – MARCHETTI, C. – MERIALDO, P., Multi-Label Classification of Bills from the Italian Senate, in CEUR Workshop Proceedings, CEUR-WS, 2022, p. 39 ss., link: https://ceur-ws.org/Vol-3285/paper3.pdf

SAJEVA A., IANNUCCI S., MARCHETTI C., MERIALDO P., TORLONE R., Clustering Amendments with Semantic Embeddings. SEBD 2024: 312-320, link: https://sebd2024.unica.it/papers/paper58.pdf

DI PORTO, V. – MARCHETTI, C., L'intelligenza artificiale  in Parlamento tra politica e amministrazione, in Rassegna Parlamentare, n. 2, maggio-agosto 2024.

DI PORTO, V. – MARCHETTI, C., Aggiornamenti in tema di intelligenza artificiale  in Parlamento, in Rassegna Parlamentare, n. 3, settembre-dicembre 2024.

# Thank you!

carlo.marchetti@senato.it