ALMA MATER STUDIORUM UNIVERSITY OF BOLOGNA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING – DISI

Natural Language Processing for Legal Domains with Small and Large Language Models in Low-Resource Regimes

Summarization, Question Answering, Retrieval, Classification and Chatbots

Prof. Gianluca Moro

DISI, University of Bologna - Cesena campus

gianluca.moro@unibo.it

Text Mining and NLP Group at DISI-Cesena

LED BY

Gianluca Moro
Associate Professor

Giacomo Frisoni
PostDoc

Luca Ragazzi
PostDoc

Paolo Italiani
PhD Student

Lorenzo Molfetta
PhD Student

Alessio Cocchieri
PhD Student

+International research collaborations.
Group supported by competitive projects and companies.

Natural Language Processing, Generation, Understanding

Large Language Models

Knowledge-Enhanced NLP

Representation Learning

Graph Machine Learning

Explainability

Reasoning

General-domain and Commonsense

Law, Medicine, etc.

Conferences: AAAI, ICLR, ACL, EMNLP, COLING, IJCAI, ECAI, etc.
Journals: Computational Linguistics, Neurocomputing, IEEE Access, AI and Law, Bioinformatics, etc.
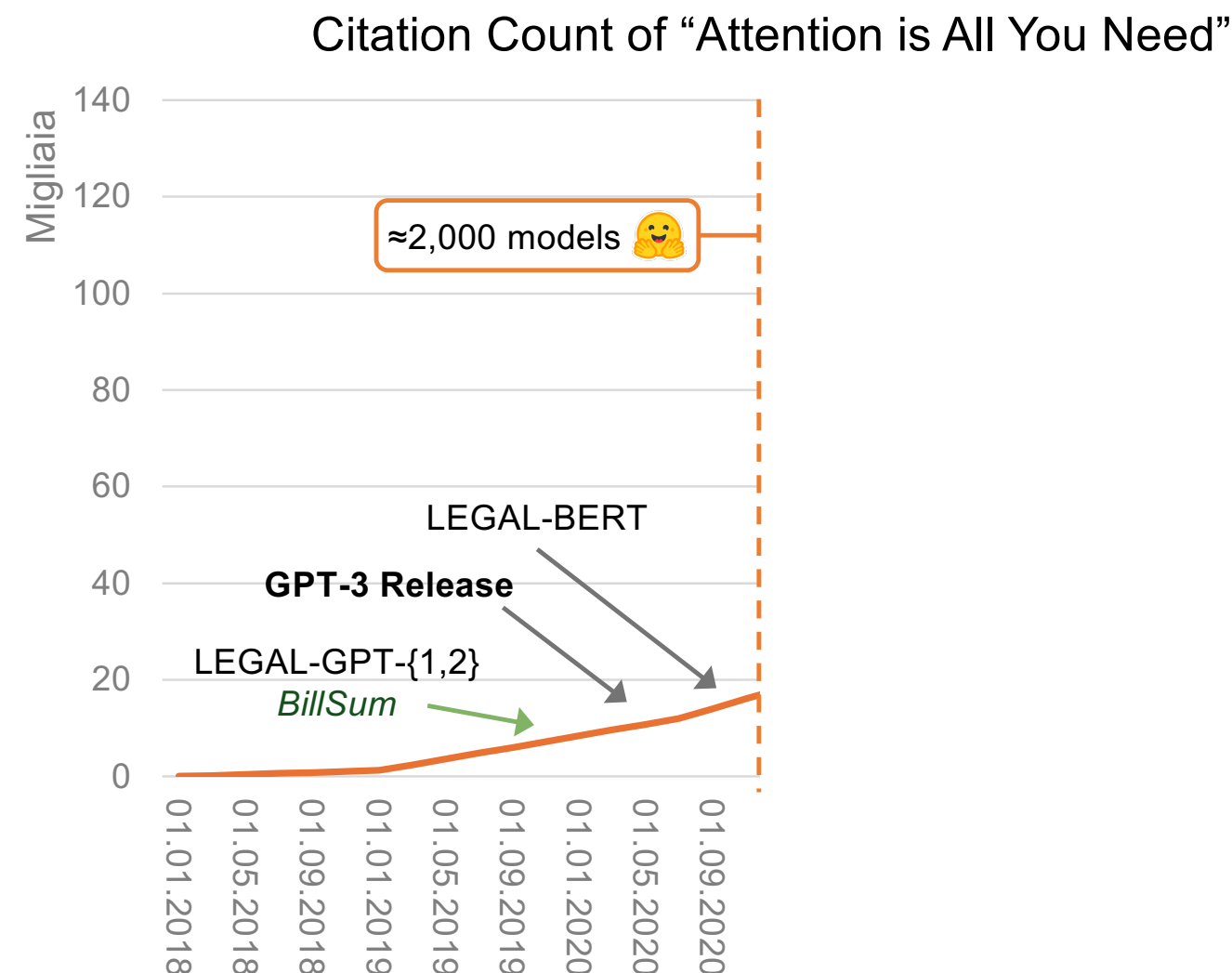
+30 publications in the last 3 years

Legal-Focused NLP Growth (2017-2020) – Legal-BERT

- The citation curve of Transformers is a good proxy for illustrating the rapid evolution of the field and contextualizing key Legal NLP milestones

⏩ **2017-2020**

- *Jun 2017:* **Transformer** paper
- *Oct 2018:* **pre-training** and **fine-tuning** popularized by BERT
- Transformer variants (architectural modification, pre-training, and applications); e.g., GPT, BART, T5
- First legal pre-trained language models
- *May 2020:* GPT-3, a dense decoder-only Transformer (175B parameters)
  - **Prompts and scaling laws**
  - Adaptation to new tasks based on the context provided in the prompt (**in-context learning**)

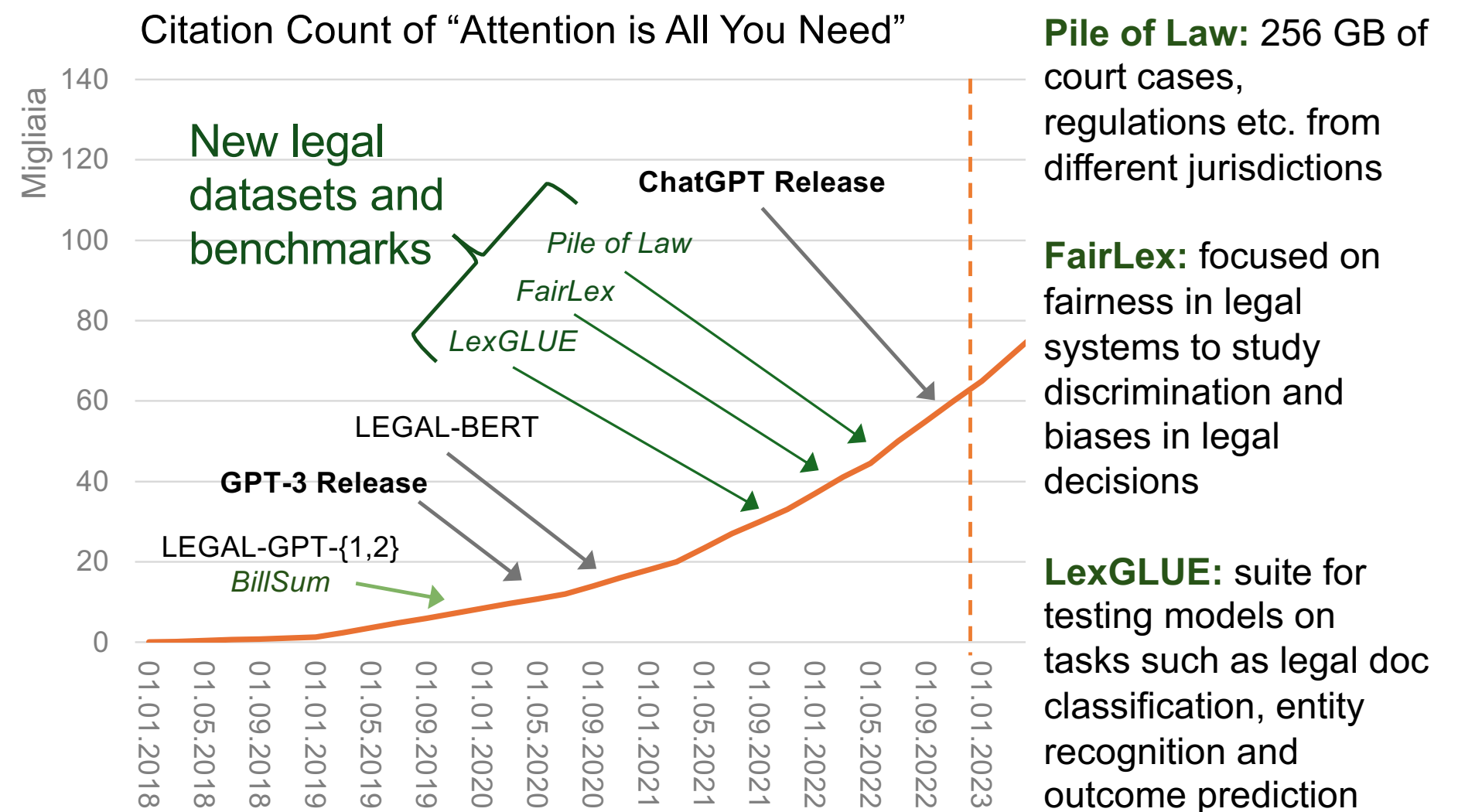Citation Count of "Attention is All You Need"

Legal-Focused NLP Growth (2021-2022) – Legal Datasets

- The citation curve of Transformers is a good proxy for illustrating the rapid evolution of the field and contextualizing key Legal NLP milestones

⏩ **2021-2022**

- *Jul 2021:* Already 16 Transformer-based models on **data modalities other than text** (e.g., vision, audio)
- *Nov 2022:* ChatGPT launch
  - The fastest growing web platform ever; 100M monthly active users in 2 months only
- **LLMs** started taking a center stage in **Generative AI**
- Mass popularization and innovation akin to the early days of iOS and Android apps
- Active release of legal corpora and multi-task datasets

Citation Count of "Attention is All You Need"

New legal datasets and benchmarks

ChatGPT Release

Pile of Law

FairLex

LexGLUE

LEGAL-BERT

GPT-3 Release

LEGAL-GPT-{1,2}

BillSum

Migliaia

Pile of Law: 256 GB of court cases, regulations etc. from different jurisdictions

FairLex: focused on fairness in legal systems to study discrimination and biases in legal decisions

LexGLUE: suite for testing models on tasks such as legal doc classification, entity recognition and outcome prediction
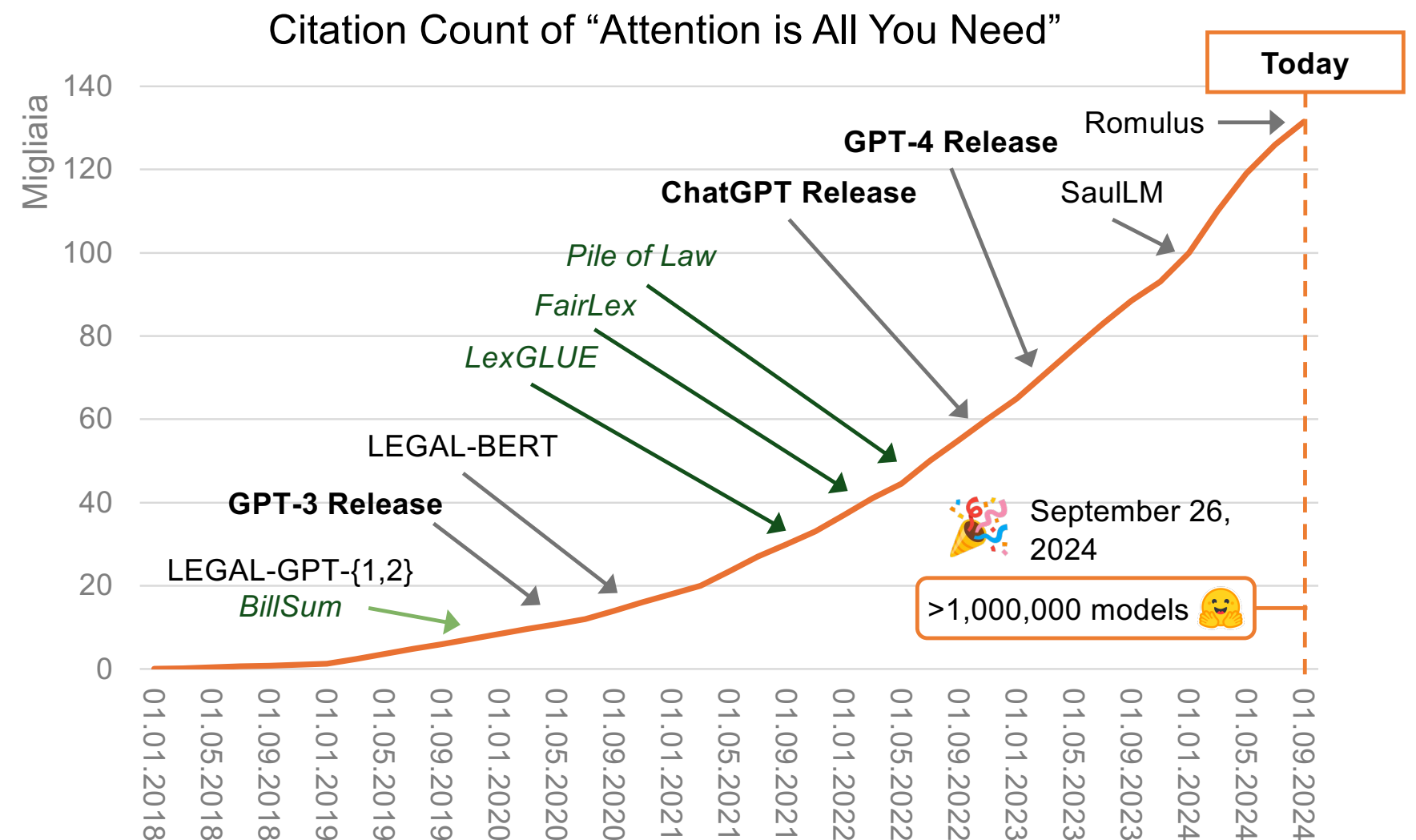
Legal-Focused NLP Growth (2023-2024) – Legal LLMs

- The citation curve of Transformers is a good proxy for illustrating the rapid evolution of the field and contextualizing key Legal NLP milestones

⏩ **2023-2024**

- *Feb 2023:* **Parameter-efficient Fine-tuning** (HuggingFace library)
- *Mar 2023:* GPT-4, 1.8T parameters
  - **Emergent abilities**
- **Foundation models**: the same model for many tasks and modalities
- More focus on **data quality**
- Language models on edge devices
- **Autonomous LLM agents and tools**
- Mixture-of-Experts and Transformer successors, e.g., Mamba, xLSTM
- **First legal LLMs** with continuous pre-training and instruction fine-tuning
  - **SaulLM (March): USA, Europe**
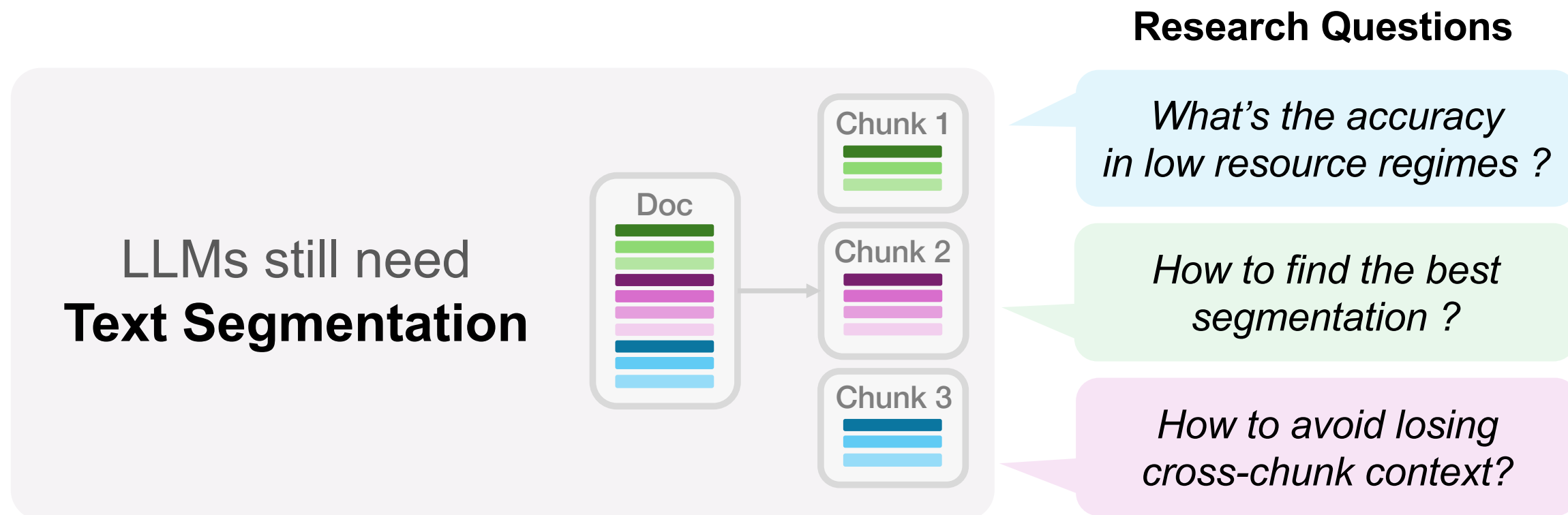 - **Romulus (September): French**

Citation Count of "Attention is All You Need"

Migliaia

Today

GPT-4 Release — Romulus

ChatGPT Release — SaulLM

Pile of Law

FairLex

LexGLUE

LEGAL-BERT

GPT-3 Release

LEGAL-GPT-{1,2}

BillSum

🎉 September 26, 2024

>1,000,000 models 🤗

Legal Document Summarization

Document

SENTENZA 13-22 DICEMBRE 1989

LA CORTE COSTITUZIONALE

composta dai signori: Presidente: dott. Francesco SAJA; Giudici: prof. Giovanni CONSO, prof. Ettore GALLO, dott. Aldo CORASANITI, prof. Giuseppe BORZELLINO, dott. Francesco GRECO, prof. Gabriele PESCATORE, avv. Ugo SPAGNOLI, prof. Francesco Paolo CASAVOLA, prof. Antonio BALDASSARRE, prof. Vincenzo CAIANIELLO, avv. Mauro FERRI, prof. Luigi MENGONI, prof. Enzo CHELI;

ha pronunciato la seguente

SENTENZA

nel giudizio di legittimità costituzionale dell'art. 27, ultimo comma, della legge 9 ottobre 1970 n. 740 (Ordinamento delle categorie di personale sanitario addetto agli istituti di prevenzione e pena non appartenenti ai ruoli organici dell'Amministrazione penitenziaria), promosso con ordinanza emessa il 22 novembre 1988 dal T.A.R. per la Sicilia, sezione staccata di Catania, sul ricorso proposto da Guarnera Francesco contro il Ministero di Grazia e Giustizia, iscritta al n. 332 del registro ordinanze 1989 e pubblicata nella Gazzetta Ufficiale della Repubblica n. 27, prima serie speciale, dell'anno 1989;

Ritenuto in fatto

Il Tribunale amministrativo regionale per la Sicilia - sezione staccata di Catania, ha sollevato questione incidentale di legittimità costituzionale dell'art. 27, ultimo comma, della legge 9 ottobre 1970, n. 740, che, in caso di sospensione cautelare dal servizio, rende obbligatoria, per i medici incaricati che prestano servizio presso gli istituti di prevenzione e pena, la mancata corresponsione degli assegni alimentari, senza alcuna valutazione del caso concreto, e ciò per preteso contrasto con gli artt. 3 e 36 della Costituzione. *[...]*

Considerato in diritto

1. Il T.A.R. della Sicilia, sezione staccata di Catania, dubita della legittimità costituzionale dell'art. 27 ultimo comma della legge 9 ottobre 1970 n. 740 (Ordinamento delle categorie di personale sanitario addetto agli istituti di prevenzione e pena non appartenenti ai ruoli organici dell'Amministrazione penitenziaria). Tale disposizione prevede che "durante il periodo della sospensione cautelare" - regolata dai precedenti commi dello stesso art. 27 - "al medico incaricato non compete alcun assegno". Essa sarebbe perciò in contrasto con l'art. 3 della Costituzione in quanto darebbe luogo ad una ingiustificata disparità di trattamento rispetto agli impiegati civili dello Stato, cui l'art. 82 del d.P.R. 10 gennaio 1957 n. 3 concede in caso di sospensione un assegno alimentare, nonché con l'art. 36 della Costituzione. *[...]*

PER QUESTI MOTIVI
LA CORTE COSTITUZIONALE

Dichiara non fondata la questione di legittimità costituzionale dell'art. 27 ultimo Dichiara comma della legge 9 ottobre 1970 n. 740 (Ordinamento delle categorie di personale sanitario addetto agli istituti di prevenzione e pena non appartenenti ai ruoli organici dell'Amministrazione penitenziaria) sollevata, in riferimento agli artt. 3 e 36 della Costituzione, dal T.A.R. della Sicilia, sezione staccata di Catania, con l'ordinanza in epigrafe.

Così deciso in Roma, nella sede della Corte costituzionale, Palazzo della Consulta, il 13 dicembre 1989.
Il Presidente: SAJA
Il redattore: FERRI
Il cancelliere: MINELLI
Depositata in cancelleria il 22 dicembre 1989.
Il direttore della cancelleria: MINELLI

(Sections: EPIGRAPH, TEXT, DECISION)

Summary

Massima n. 15219

Diversamente dagli impiegati civili dello Stato, i medici incaricati che prestano servizio presso gli istituti di prevenzione e pena possono esercitare liberamente la professione ed assumere altri impieghi o incarichi, sicche' la disposizione che esclude la corresponsione di alcun assegno durante la sospensione cautelare non appare ne' iniqua ne' irragionevole. (Non fondatezza - in riferimento agli artt. 3 e 36 Cost. - della questione di costituzionalita' dell'art. 27, ultimo comma, L. 9 ottobre 1970 n. 740, nella parte in cui vieta la corresponsione degli assegni alimentari ai medici sospesi cautelarmente dal servizio).

LLMs Struggle with Long Legal Documents

- **Context Limits:** LLMs typically handle 4K to 16K input tokens (e.g., *LLaMA-3*), which restricts how much text they can consider at once.

- **Lost in the Middle:** LLMs, even with 100K token contexts (e.g., *Claude-3*), drop in QA [1] and summarization [2] performances with key info in long mid-input
  - Fail to catch the different importance of local context in central sections - *Vanishing attention*
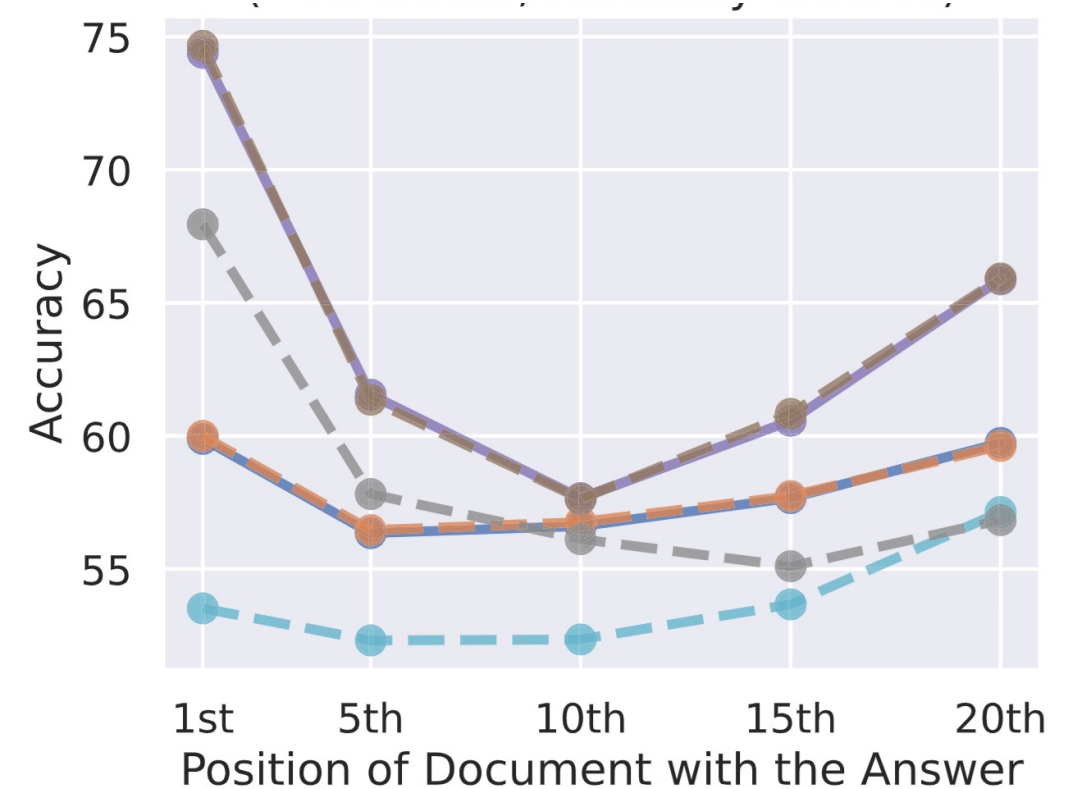
Research Questions

LLMs still need
Text Segmentation

Doc

Chunk 1

What's the accuracy in low resource regimes ?

Chunk 2

How to find the best segmentation ?

Chunk 3

How to avoid losing cross-chunk context?

LLMs Get Lost When The Context Becomes Too Long

Knowledge Chunking

LLM performance decreases as the
input context grows longer
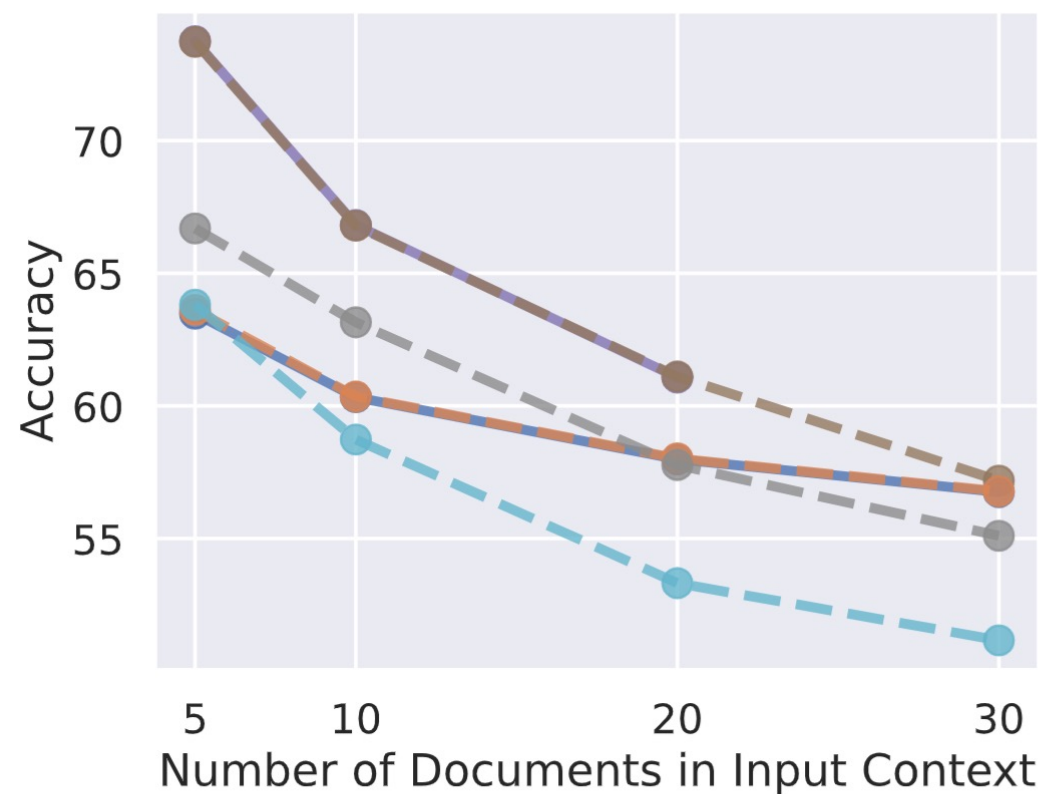
U-shaped LLM performance curve as the location of
relevant information in the input context varies

Source: Multi-document QA
example from Liu et al. (2024)

Legend:
- claude-1.3
- claude-1.3-100k
- gpt-3.5-turbo-0613
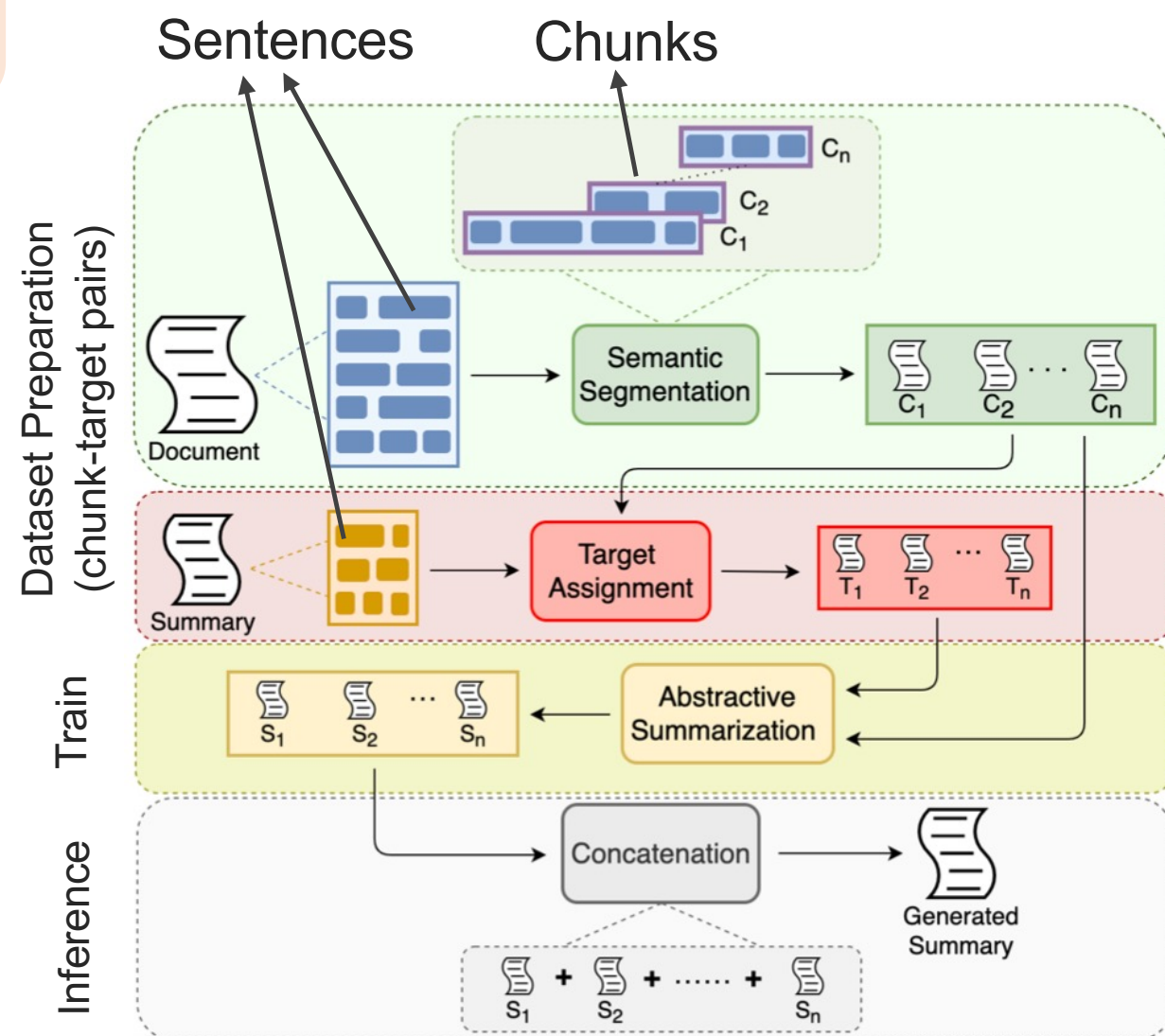- gpt-3.5-turbo-16k-0613
- mpt-30b-instruct
- longchat-13b-16k

Semantic Self-Segmentation for Abstractive Summarization of Long Documents in Low-Resource Regimes

Moro G., Ragazzi L. `AAAI 2022` **GGS Ranking: A++**

Previous approaches were based on input truncation or section-based segmentation and not for legal domains

Sentences   Chunks

🏛 **1st step: LEGAL-BERT** trained with metric learning to identify if 2 sentences are in the same section, on a general-purpose text set

Semantically aggregate sentences into **chunks** of **preferred min-max** size, using **LEGAL-BERT, based on available GPU memory capacity.**

Resolve hardware issues by creating small inputs

Chunk target labeling by assigning summary sentences to most similar chunk (with ROUGE-1)

Resolve data scarcity by transforming 1 training instance in multiple ones

Se3 – SOTA Results!

...with a **quadratic Transformer**

Two legal datasets

Input size

System (*MaxLen*)	8K tok per doc	1.6K tok per doc
	AustLII R1 / R2 / RL	**BillSum** R1 / R2 / RL
Baselines *Truncation-based*		
BART$_{BASE}$ (*1024*)	33.51/23.92/27.88	54.42/35.81/41.98
BART$_{BASE}$ (*512*)	26.61/17.67/21.79	49.84/30.67/37.73
Baselines w/ Se3 - triplet *Segmentation-based*		
BART$_{BASE}$ (*1024*)	**59.04/52.46/53.67**	57.31/37.85/43.78
BART$_{BASE}$ (*512*)	<u>53.14/46.44/47.38</u>	55.65/35.73/40.99
Baselines w/ Se3 - contrastive		
BART$_{BASE}$ (*1024*)	57.96/50.92/52.49	**57.66/38.20/44.11**
BART$_{BASE}$ (*512*)	52.66/45.71/46.66	<u>55.96/35.82/41.27</u>

Different metric losses

Huge improvement!
52.46 vs. 23.92 of ROUGE-2

Se3 – SOTA Results!

Consistently higher for AustLII
despite the baseline input is longer

...with a **quadratic Transformer**

Two legal datasets

Input size

System (*MaxLen*)	8K tok per doc AustLII R1 / R2 / RL	1.6K tok per doc BillSum R1 / R2 / RL
Baselines *Truncation-based*		
BART$_{BASE}$ (*1024*)	33.51/23.92/27.88	54.42/35.81/41.98
BART$_{BASE}$ (*512*)	26.61/17.67/21.79	49.84/30.67/37.73
Baselines w/ Se3 - triplet *Segmentation-based*		
BART$_{BASE}$ (*1024*)	59.04/52.46/53.67	57.31/37.85/43.78
BART$_{BASE}$ (*512*)	53.14/46.44/47.38	55.65/35.73/40.99
Baselines w/ Se3 - contrastive		
BART$_{BASE}$ (*1024*)	57.96/50.92/52.49	**57.66/38.20/44.11**
BART$_{BASE}$ (*512*)	52.66/45.71/46.66	55.96/35.82/41.27

Different metric losses

Huge improvement!
52.46 vs. 23.92 of ROUGE-2

...and with a **linear Transformer?**

	8K tok per doc	1.6K tok per doc
System (*MaxLen*)	AustLII R1 / R2 / RL	BillSum R1 / R2 / RL
Baselines		
LED$_{BASE}$ (*4096*)	50.27/39.85/42.04	58.83/39.83/45.71
LED$_{BASE}$ (*2048*)	42.76/32.20/35.71	58.38/39.37/45.09
Baselines w/ Se3 - triplet		
LED$_{BASE}$ (*4096*)	57.89/48.96/50.28	58.51/39.71/45.66
LED$_{BASE}$ (*2048*)	**60.03/53.03/54.57**	58.38/39.53/45.48
Baselines w/ Se3 - contrastive		
LED$_{BASE}$ (*4096*)	57.82/49.06/50.50	**59.18/40.18/46.04**

Se3 – SOTA Results!

Consistently higher for AustLII despite the baseline input is longer

...with a **quadratic Transformer**

Two legal datasets

Input size

System (*MaxLen*)	AustLII 8K tok per doc R1 / R2 / RL	BillSum 1.6K tok per doc R1 / R2 / RL
Baselines *Truncation-based*		
BART$_{BASE}$ (*1024*)	33.51/23.92/27.88	54.42/35.81/41.98
BART$_{BASE}$ (*512*)	26.61/17.67/21.79	49.84/30.67/37.73
Baselines w/ Se3 - triplet *Segmentation-based*		
BART$_{BASE}$ (*1024*)	**59.04/52.46/53.67**	57.31/37.85/43.78
BART$_{BASE}$ (*512*)	<u>53.14/46.44/47.38</u>	55.65/35.73/40.99
Baselines w/ Se3 - contrastive		
BART$_{BASE}$ (*1024*)	57.96/50.92/52.49	**57.66/38.20/44.11**
BART$_{BASE}$ (*512*)	52.66/45.71/46.66	<u>55.96/35.82/41.27</u>

Different metric losses

...and with a **linear Transformer?**

System (*MaxLen*)	AustLII 8K tok per doc R1 / R2 / RL	BillSum 1.6K tok per doc R1 / R2 / RL
Baselines		
LED$_{BASE}$ (*4096*)	50.27/39.85/42.04	58.83/39.83/45.71
LED$_{BASE}$ (*2048*)	42.76/32.20/35.71	58.38/39.37/45.09
Baselines w/ Se3 - triplet		
LED$_{BASE}$ (*4096*)	<u>57.89/48.96/50.28</u>	58.51/39.71/45.66
LED$_{BASE}$ (*2048*)	**60.03/53.03/54.57**	58.38/39.53/45.48
Baselines w/ Se3 - contrastive		
LED$_{BASE}$ (*4096*)	57.82/49.06/50.50	**59.18/40.18/46.04**

in **data-scarcity** scenario: 10, 100 samples

System (*MaxLen*)	BillSum (10) R1 / R2 / RL	BillSum (100) R1 / R2 / RL
Baselines		
LED$_{BASE}$	41.10/21.15/27.93	47.68/26.98/32.43
Solutions w/ Se3		
LED$_{BASE}$ (*512*)	**46.94**/23.04/**29.29**	**50.45/27.73**/33.74
LED$_{BASE}$ (*256*)	46.22/**24.32**/29.16	48.13/27.16/31.89

2x lower GPU RAM consumption w.r.t. LED's 4096 baseline

Align-Then-Abstract Representation Learning for Low-Resource Summarization

Moro G., Ragazzi L. Neurocomputing 2023 **Journal Ranking: Q1**

What about learning how to segment a document as well?

Minimize cosine distance

$$\mathcal{L}_{align}^{\eta} = 1 - \frac{\bar{X}^e \cdot \bar{Y}^e}{|\bar{X}^e||\bar{Y}^e|}$$

$$\mathcal{L} = \mathcal{L}_{align}^{\eta} + \mathcal{L}_{summ}^{Y}$$

Cross-entropy loss

$$\mathcal{L}_{summ}^{Y} = -\frac{1}{|\bar{Y}|} \sum_{t=1}^{|\bar{Y}|} \log p(y_t | y_{1:t-1}, \bar{X})$$
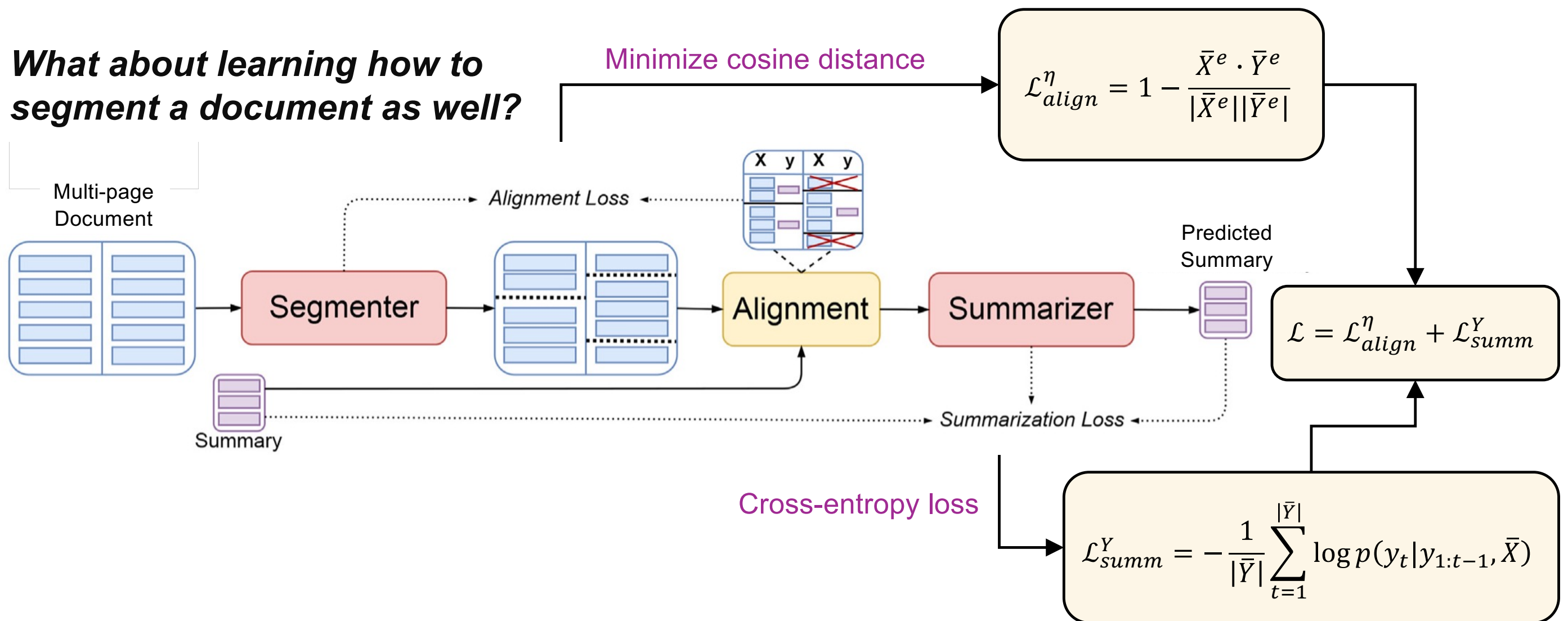
Athena – Improvement over Se3

We register better ROUGE scores w.r.t. Se3 in few-shot learning!

Model	BillSum (10)				BillSum (100)			
	$R\text{-}1_{f1}$	$R\text{-}2_{f1}$	$R\text{-}L_{f1}$	$\mathcal{R}$	$R\text{-}1_{f1}$	$R\text{-}2_{f1}$	$R\text{-}L_{f1}$	$\mathcal{R}$
SE3 (1024)	44.37	21.17	27.57	30.74	47.85	26.67	33.36	35.68
SE3 (512)	46.58	22.03	28.23	31.93	49.88	26.84	33.33	36.34
SE3 (256)	46.50	23.24	28.54	32.44	48.17	26.55	31.51	35.11
SE3 (128)	41.48	22.73	26.37	30.00	42.42	25.42	28.98	32.10
ATHENA (dynamic)	**47.57**	**24.14**	**30.35**	**33.69**	**51.59**	**29.36**	**35.04**	**38.32**

Dynamic size because Athena learns where to segment

"But few-shot performance may depend on the subset used for training…"

		Subsets				
Dataset	Instances	#1	#2	#3	#4	#5
BILLSUM	10	47.57/24.14/30.35	47.88/24.85/30.81	48.57/25.40/31.07	47.36/23.77/29.56	47.54/24.37/30.29
	100	51.59/29.36/35.04	51.04/28.96/34.87	50.66/28.38/34.42	51.23/28.98/34.69	51.09/28.65/34.76

We obtain consistent scores regardless of the training subsets!

Question Answering

According to the plaintiffs, what specific Fourth Amendment right did the TSA and DEA violate when they seized and confiscated their money without reasonable suspicion of crime?

Enhancing Legal Question Answering with Data Generation and Knowledge Distillation from Large Language Models

Italiani P., Ragazzi L., Moro G. AI and Law   Journal Ranking: Q1

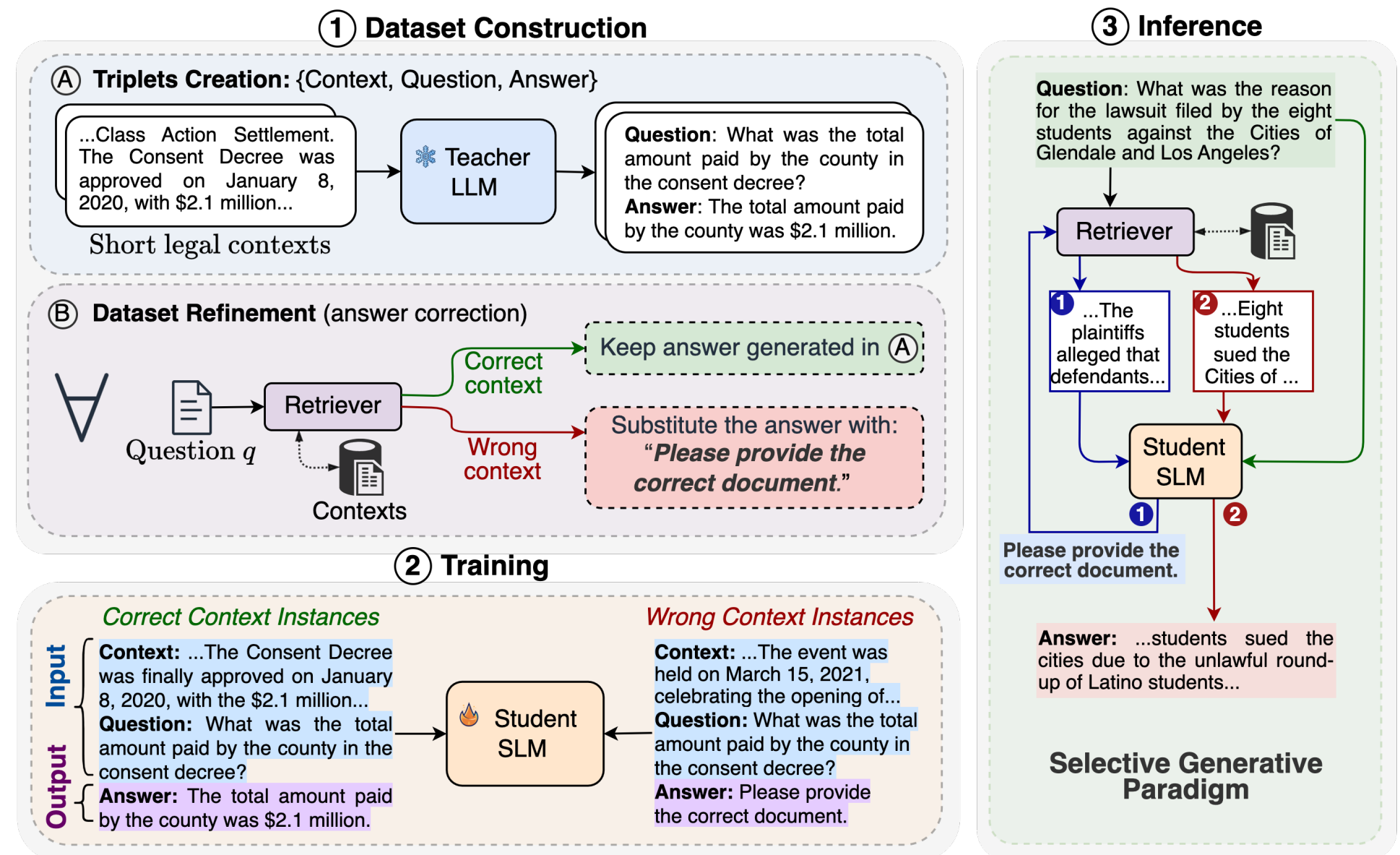A new distillation method for Legal Question Answering (LQA):

1.A A frozen LLM (LLaMA-2-13B) produces artificial Question-Answer pairs from legal cases

1.B QA dataset refined adding instances with wrong contexts to questions

2. Training a Student LM (Small LM) with this refined dataset

3. Inference: The SLM trained to answer questions can also directly evaluate the pertinence of the retrieved document and request to retriever a new one as needed.

- SLMs when equipped with SGP improve retrieval performances by up to 9%

① Dataset Construction

Ⓐ Triplets Creation: {Context, Question, Answer}

...Class Action Settlement. The Consent Decree was approved on January 8, 2020, with $2.1 million...

→ ❄ Teacher LLM →

Question: What was the total amount paid by the county in the consent decree?
Answer: The total amount paid by the county was $2.1 million.

Short legal contexts

Ⓑ Dataset Refinement (answer correction)

∀ Question q → Retriever → Contexts

Correct context → Keep answer generated in Ⓐ

Wrong context → Substitute the answer with: "*Please provide the correct document.*"

② Training

Correct Context Instances

Input
Context: ...The Consent Decree was finally approved on January 8, 2020, with the $2.1 million...
Question: What was the total amount paid by the county in the consent decree?
Output
Answer: The total amount paid by the county was $2.1 million.

→ 🔥 Student SLM

Wrong Context Instances

Context: ...The event was held on March 15, 2021, celebrating the opening of...
Question: What was the total amount paid by the county in the consent decree?
Answer: Please provide the correct document.

③ Inference

Question: What was the reason for the lawsuit filed by the eight students against the Cities of Glendale and Los Angeles?

→ Retriever → 📚

① ...The plaintiffs alleged that defendants...
② ...Eight students sued the Cities of ...

→ Student SLM

① Please provide the correct document.

Answer: ...students sued the cities due to the unlawful round-up of Latino students...

Selective Generative Paradigm

Enhancing Legal Question Answering with Data Generation and Knowledge Distillation from Large Language Models

Italiani P., Ragazzi L., Moro G.   AI and Law   **Journal Ranking: Q1**
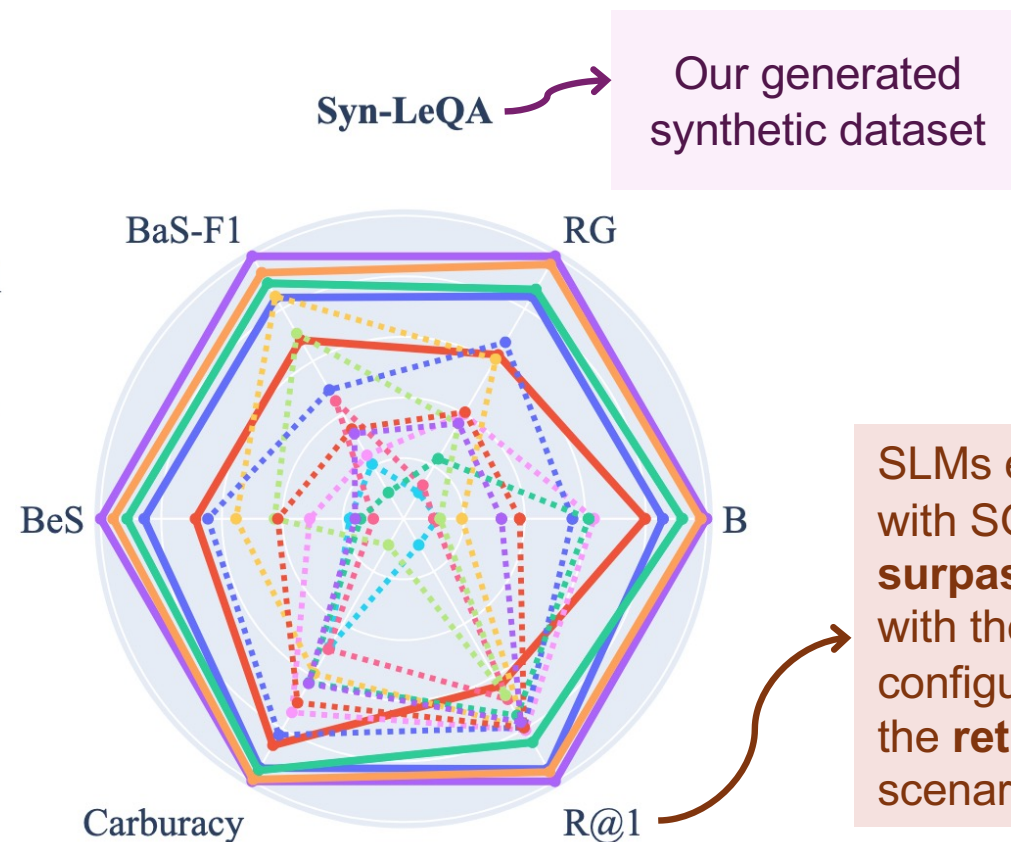
Fine-tuned SLMs **outperform** zero-shot LLMs on Syn-LeQA, both on **syntactic** (RG = avg. ROUGE-1\2\L, B = avg. BLEU-1\2\3\4) and **semantic** (BeS = BERT-Score, BaS-F1 = BLEU-aligned F1 Score) metrics, despite the substantial difference in **resources** required.

Syn-LeQA → Our generated synthetic dataset

SLM   LLM

Bubble size corresponds to the **number of parameters** in each model

Our retrieval enhanced fine-tuned **SLMs**
- BART-base
- FLAN-T5-small
- FLAN-T5-base
- FLAN-T5-large
- BART-large

Retrieval enhanced Zero-shot **LLMs**
- Mistral-7B
- Orca-2-13B
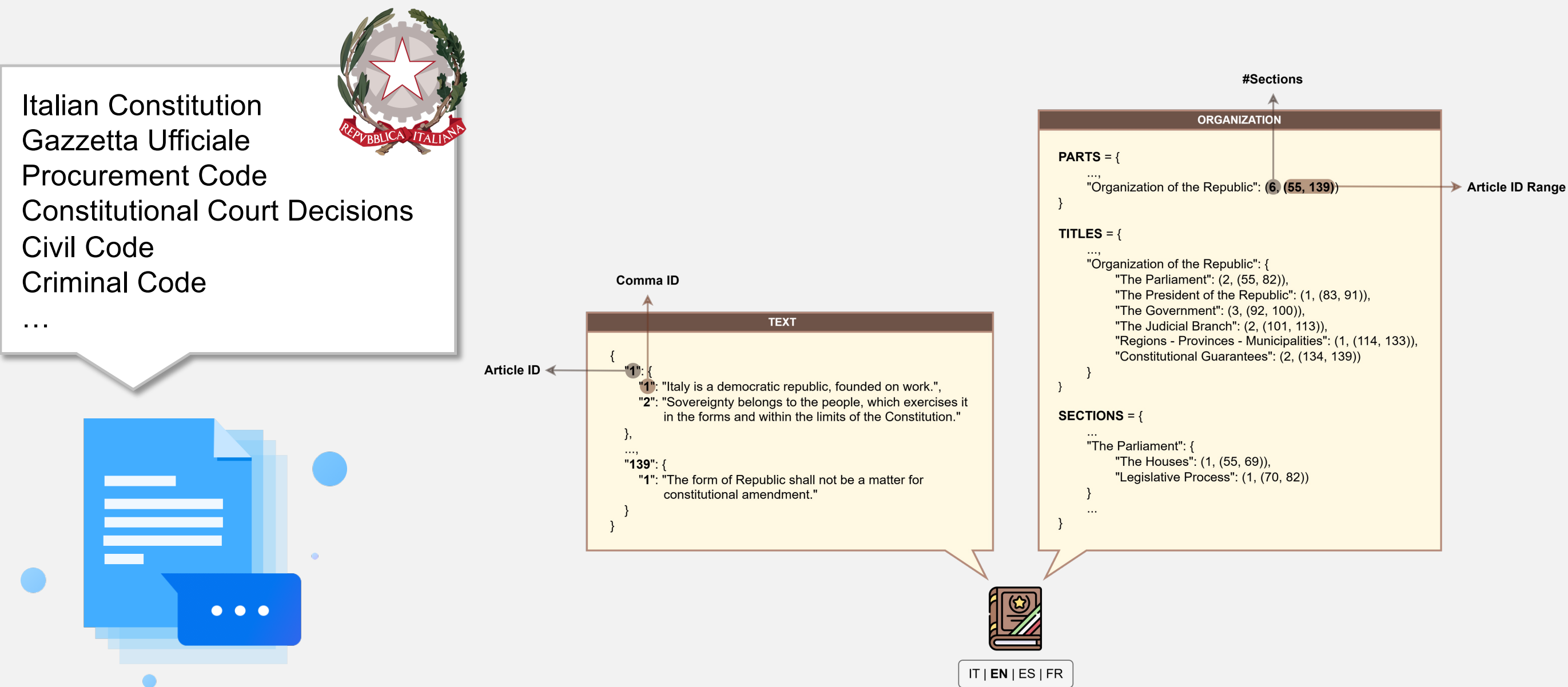- Notus-7B
- Llama-2-7B
- Zephyr-7B
- Starling-7B

BaS-F1   RG   B   R@1   Carburacy   BeS

SLMs equipped with SGP **surpass** LLMs with the same configuration on the **retrieval** scenario (R@1)

RG

Avg. response time in sec per question with Nvidia RTX 3090
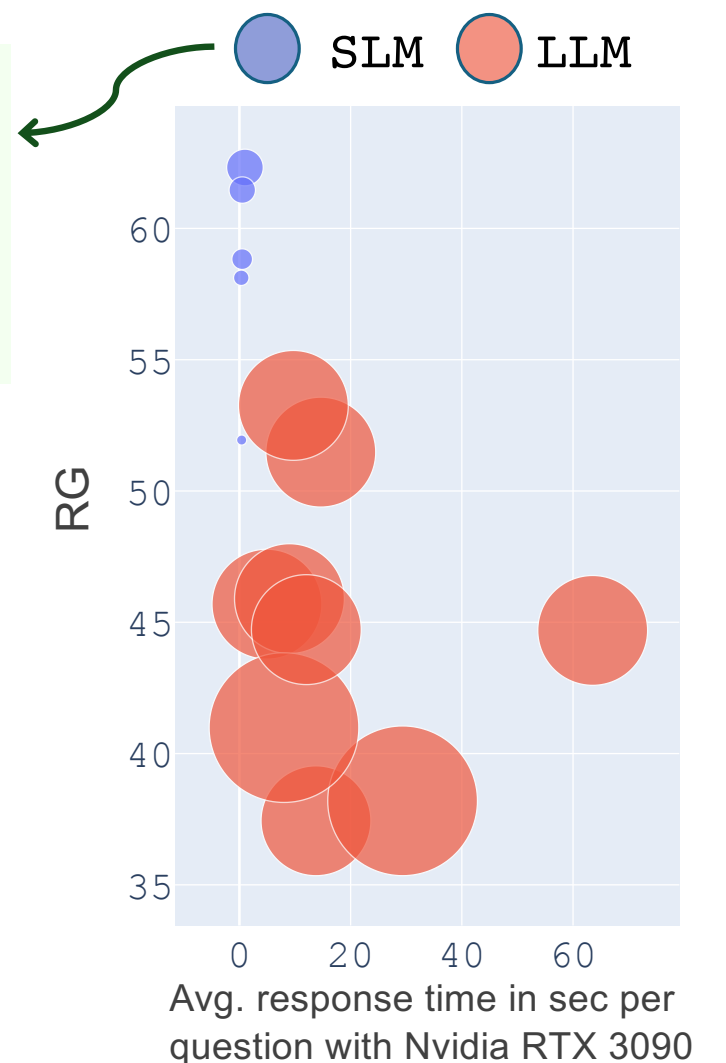
Creation of new Italian & Multi-Lingual Legal Datasets and Benchmarks

Italian Constitution
Gazzetta Ufficiale
Procurement Code
Constitutional Court Decisions
Civil Code
Criminal Code
…

#Sections

ORGANIZATION

PARTS = {
    …,
    "Organization of the Republic": (6, (55, 139))
}

→ **Article ID Range**

TITLES = {
    …,
    "Organization of the Republic": {
        "The Parliament": (2, (55, 82)),
        "The President of the Republic": (1, (83, 91)),
        "The Government": (3, (92, 100)),
        "The Judicial Branch": (2, (101, 113)),
        "Regions - Provinces - Municipalities": (1, (114, 133)),
        "Constitutional Guarantees": (2, (134, 139))
    }
}

SECTIONS = {
    …
    "The Parliament": {
        "The Houses": (1, (55, 69)),
        "Legislative Process": (1, (70, 82))
    }
    …
}

Comma ID

TEXT

{
    "1": {
        "1": "Italy is a democratic republic, founded on work.",
        "2": "Sovereignty belongs to the people, which exercises it
             in the forms and within the limits of the Constitution."
    },
    …,
    "139": {
        "1": "The form of Republic shall not be a matter for
             constitutional amendment."
    }
}

Article ID

IT | **EN** | ES | FR

COMMA: A Multi-Task and Multi-Lingual Dataset of Constitutional Court Verdicts

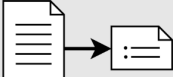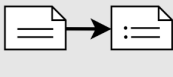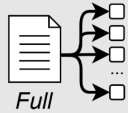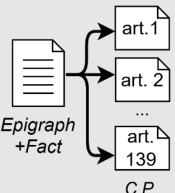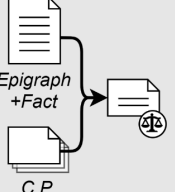Ragazzi L., Frisoni G., Moro G., Italiani P., Molfetta L., Folin V.

- **Collecting +14,000 Italian verdicts** from 1956 to 2022 by scraping, parsing and structuring the text data for 7 tasks
  - Translated in 3 other languages
  - Civil law system

Extension of **"LAWSUIT: a LArge expert Written SUmmarization dataset of ITalian constitutional court verdicts".**
AI and Law, 2024.
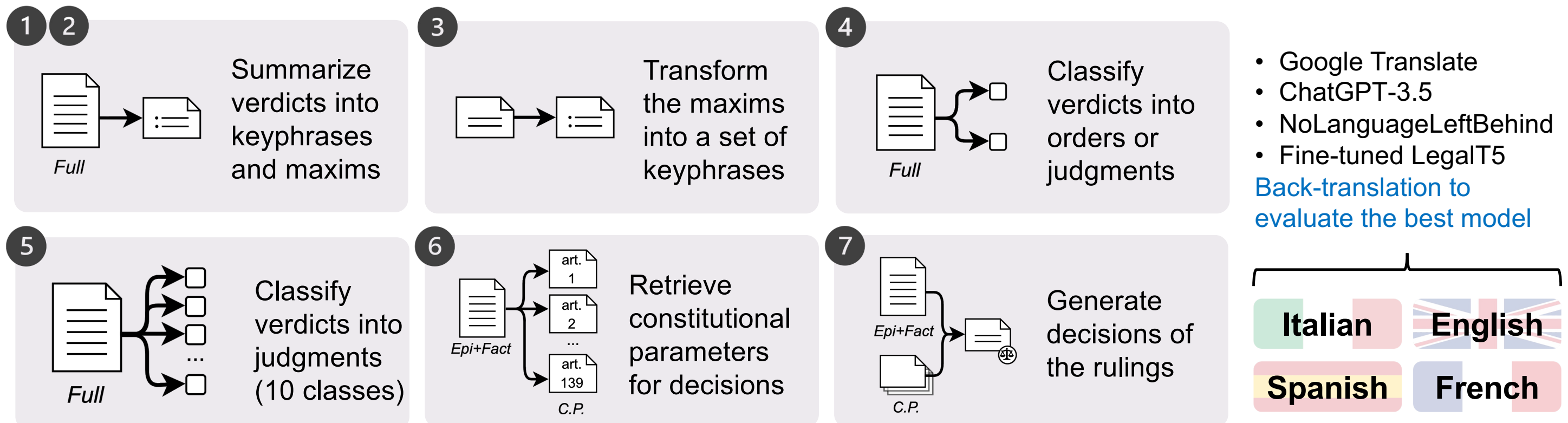
(1)(2) Summarize verdicts into keyphrases and maxims

(3) Transform the maxims into a set of keyphrases

(4) Classify verdicts into orders or judgments

(5) Classify verdicts into judgments (10 classes)

(6) Retrieve constitutional parameters for decisions

(7) Generate decisions of the rulings

- Google Translate
- ChatGPT-3.5
- NoLanguageLeftBehind
- Fine-tuned LegalT5

Back-translation to evaluate the best model

Italian **English**
Spanish **French**

COMMA: A Multi-Task and Multi-Lingual Dataset of Constitutional Court Verdicts

Ragazzi L., Frisoni G., Moro G., Italiani P., Molfetta L., Folin V.

ABSTRACTIVE SUMMARIZATION		
1 **DOC→SUMM** *Full*	In the judgment of constitutional legitimacy of letter b) of number 1) of note II-bis of art. 1 of Part One of the Tariff attached to Presidential Decree 26 April 1986, no. 131 (Approval of the consolidated text of the provisions concerning the registration tax) *[...]* BELIEVED IN FACT. *[...]* CONSIDERED IN LAW. *[...]* The Constitutional Court declares the manifest inadmissibility of the question *[...]* raised, with reference to the articles 3 and 53 of the Constitution and the principles of reasonableness, rationality, and non-contradiction, by the Provincial Tax Commission of Milan with the ordinance indicated in the epigraph. Thus decided in Rome, in the seat of the Constitutional Court, Palazzo della Consulta, on 22 June 2011.	The question of constitutional legitimacy of letter b) of number 1) of note II-bis of art. 1 of Part One of the Tariff attached to Presidential Decree 26 April 1986, no. 131-as referred to in the first period of number 21) of Part II of Table A attached to the Presidential Decree 26 October 1972, no. 633-is manifestly inadmissible given, in relation to the aforementioned art. 53 of the Constitution, the absolute lack of motivation on the non-manifest groundlessness, and, with reference to art. 3 of the Constitution and the principles of reasonableness, rationality and non-contradiction, the incomplete description of the case object of the main proceedings with consequent insufficient motivation on the relevance.
2 **DOC→BULLET POINTS** *Full*	In the judgment of constitutional legitimacy of letter b) of number 1) of note II-bis of art. 1 of Part One of the Tariff attached to Presidential Decree 26 April 1986, no. 131 (Approval of the consolidated text of the provisions concerning the registration tax) *[...]* BELIEVED IN FACT. *[...]* CONSIDERED IN LAW. *[...]* The Constitutional Court declares the manifest inadmissibility of the question *[...]* raised, with reference to the articles 3 and 53 of the Constitution and the principles of reasonableness, rationality, and non-contradiction, by the Provincial Tax Commission of Milan with the ordinance indicated in the epigraph. Thus decided in Rome, in the seat of the Constitutional Court, Palazzo della Consulta, on 22 June 2011.	• taxes • tax breaks • tariff facilitation for the purchase of the "first home" • *[...]* • absolute lack of motivation on the non-manifest groundlessness and insufficient motivation on the relevance • manifest inadmissibility of the question
3 **SUMM→BULLET POINTS**	The question of constitutional legitimacy of letter b) of number 1) of note II-bis of art. 1 of Part One of the Tariff attached to Presidential Decree 26 April 1986, no. 131-as referred to in the first period of number 21) of Part II of Table A attached to the Presidential Decree 26 October 1972, no. 633-is manifestly inadmissible given, in relation to the aforementioned art. 53 of the Constitution, the absolute lack of motivation on the non-manifest groundlessness, and, with reference to art. 3 of the Constitution and the principles of reasonableness, rationality and non-contradiction, the incomplete description of the case object of the main proceedings with consequent insufficient motivation on the relevance.	• taxes • tax breaks • tariff facilitation for the purchase of the "first home" • *[...]* • absolute lack of motivation on the non-manifest groundlessness and insufficient motivation on the relevance • manifest inadmissibility of the question

COMMA: A Multi-Task and Multi-Lingual Dataset of Constitutional Court Verdicts

Ragazzi L., Frisoni G., Moro G., Italiani P., Molfetta L., Folin V.

DOCUMENT CLASSIFICATION			
4	**RULING TYPE** *Full*	In the judgment of constitutional legitimacy of letter b) of number 1) of note II-bis of art. 1 of Part One of the Tariff attached to Presidential Decree 26 April 1986, no. 131 (Approval of the consolidated text of the provisions concerning the registration tax) [...] BELIEVED IN FACT. [...] CONSIDERED IN LAW. [...] The Constitutional Court declares the manifest inadmissibility of the question [...] raised, with reference to the articles 3 and 53 of the Constitution and the principles of reasonableness, rationality, and non-contradiction, by the Provincial Tax Commission of Milan with the ordinance indicated in the epigraph. Thus decided in Rome, in the seat of the Constitutional Court, Palazzo della Consulta, on 22 June 2011.	ORDER
5	**JUDGMENT TYPE**	"	JUDGMENT OF CONSTITUTIONAL LEGITIMACY IN AN INCIDENTAL WAY
INFORMATION RETRIEVAL			
6	**ARTICLE RETRIEVAL** *Epigraph +Fact* *C.P.*	In the joint judgments of constitutional legitimacy of art. 14 of the law of the Lombardy Region 19 August 1974, n. 48 (Rules for the regulation of waste water discharges) and art. 25, first paragraph, of law 10 May 1976, n. 319 (Regulations for the protection of water from pollution) promoted with orders issued on 26 May, 14 May, 23 April, 5 April (n. 2) and 12 May 1982 (n. 2) by the Court of Cassation in the proceedings on appeals proposed by [...] BELIEVED IN FACT. [...]	art. 25 co. 2 art. 27 co. 1 art. 117
TEXT GENERATION			
7	**DECISION GENERATION** *Epigraph +Fact* *C.P.*	In the judgment of constitutional legitimacy of letter b) of number 1) of note II-bis of art. 1 of Part One of the Tariff attached to Presidential Decree 26 April 1986, no. 131 (Approval of the consolidated text of the provisions concerning the registration tax) [...] BELIEVED IN FACT. [...] [Constitution Article 3 Text] [Constitution Article 53 Text]	The Constitutional Court declares the manifest inadmissibility of the question [...] raised, with reference to the articles 3 and 53 of the Constitution and the principles of reasonableness, rationality, and non-contradiction, by the Provincial Tax Commission of Milan with the ordinance indicated in the epigraph. Thus decided in Rome, in the seat of the Constitutional Court, Palazzo della Consulta, on 22 June 2011.

Projects

Artificial Intelligence for Public Administrations Connected

- Developing AI solutions for Italian public services' ecosystems: start-ups, SME, and larger firms of the gov tech sector
  - Project awarded with the European Seal of Excellence

- Scientific Research and Technologies on efficient Large Language Models, Deep Neural Networks, Graph Machine Learning and Generative AI
  - For chatbots, document and data processing, data analytics, knowledge extraction and injection, information search, regulatory compliance, assistance in preparing legal acts
  - For fiscal analysis, fraud detection, financial prediction and planning
  - For vehicular traffic analysis and predictions, AI for food and sustainable agriculture

- Funded by the Italian Ministry of Enterprises & Made in Italy
  - € 5,722,000; Approved in March 2023 – ends in April 2026

- Consortium: Università Bocconi (coordinator), Istituto per la Finanza e l'Economia Locale, Università di Bologna, Università di Milano Bicocca, Maggioli Group, Tempo Srl

Other Funded Projects on Legal Domains

- 1x Research Scholarship (2023)
  - **Information Extraction** on Legal Documents (LLMs and metrics)
- 1x PhD Scholarship (2024-2027)
  - Knowledge-Enhanced NLP Methods and Application to **Compliance Text Generation**
  - Create documents that are compliant with schemes and laws in force

- 2x PhD Scholarships (2023-2026; 2024-2027)
  - Knowledge-Enhanced Conversational Agents for the Procurement Code

- 4x PhD Scholarships (2020-2023; 2022-2025; 2024-2027)
  - Design and Development of Advanced AI Solutions on NLP and LLMs

Procurement Code: Conversational Agent for Legal Q&A

Goals

- Provision of **reference articles** to verify the faithfulness of the predicted answers
- Reduction of **hallucinations** and **outdated information** in output generations
- Creation of an LLM **sustainable** for the **low computational resources** of companies

Problems

- **Lack of datasets** for fine-tuning a conversational agent in the procurement code domain
- Missing benchmarks and test sets **for the conversational application context**
- Absence of Italian legal encoders for retrieval and foundational LLMs
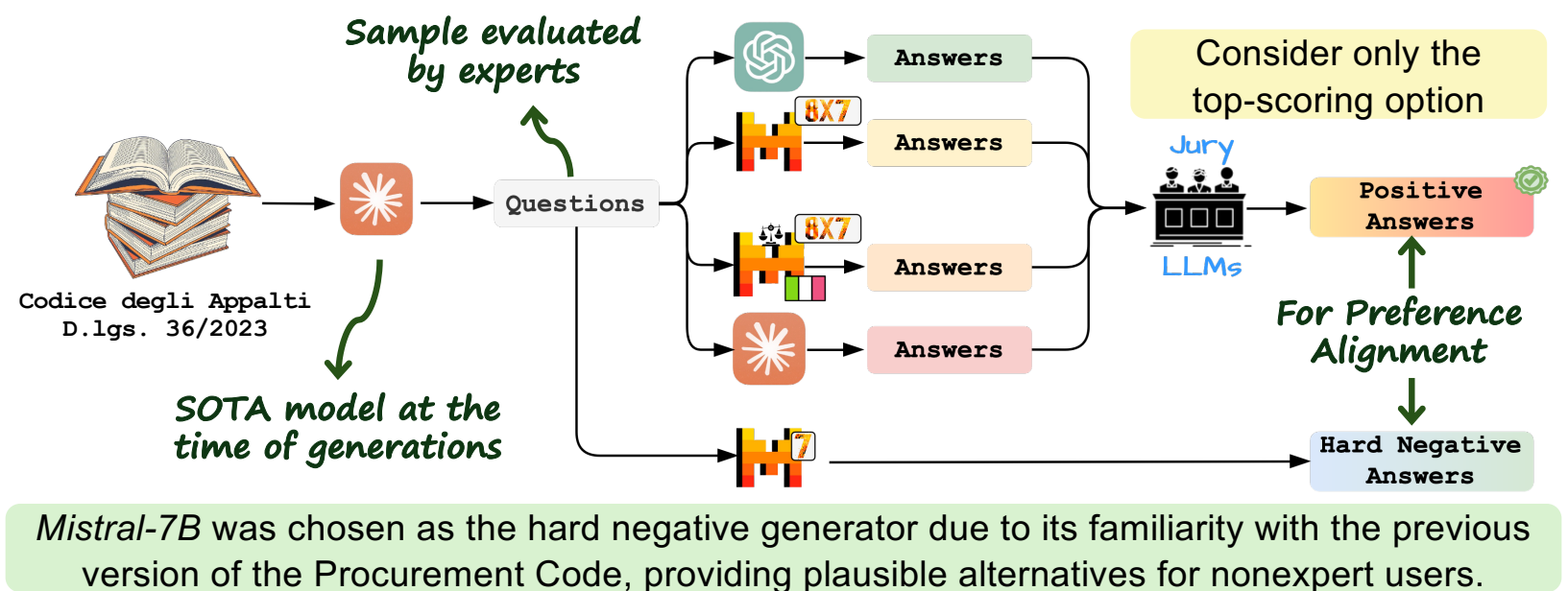
Our contributions

- We synthetically generated a **Q&A corpus** augmented with articles to support the answer prediction task
- We finetuned a **retrieval model** to navigate the Procurement Code and select commas relevant to the user queries
- We trained a **foundational legal LLM** and instruction fine-tuned it for conversational interactions

Data Generation

- Claude3-Opus was prompted with each article separately and asked to **generate** different **questions** to cover all topics in the input context
- Three different SOTA LLMs, in addition to our aligned *Mixtral_7x8B* model, were used to **generate answers based on the input question and the reference article**
- We used "**LLMs as a Jury**" to automatically evaluate generations

Codice degli Appalti D.lgs. 36/2023

Sample evaluated by experts

SOTA model at the time of generations

Questions

8X7

8X7

7

Answers

Answers

Answers

Answers

Consider only the top-scoring option

Jury

LLMs

Positive Answers

For Preference Alignment

Hard Negative Answers

Mistral-7B was chosen as the hard negative generator due to its familiarity with the previous version of the Procurement Code, providing plausible alternatives for nonexpert users.
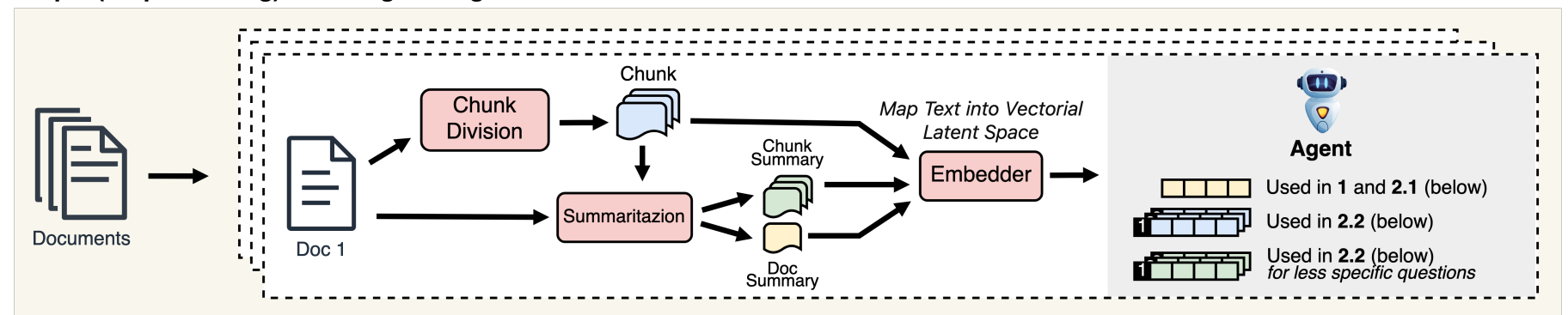
# ParlaMentis: Chatbot for the Chamber of Deputies

- We implemented a retrieval-enhanced conversational agent with **reasoning capabilities** (via ReAct) to assist parliamentarians in researching information and drafting bills

- We **pre-trained** an LLM **on parliamentarian acts** and **fine-tuned it on COMMA** to obtain a model capable of engaging in legal discourses while preserving a conversational nature

- Our architecture assigns each document an **agent** that is responsible for the fetching of its chunk and for communicating with the generative LLM that must generate the final answer

- We segment documents into chunks to isolate independent information
- Documents and chunks are automatically **summarized** and **embedded** to guide retrieval

- Agents are **summoned** and utilized to address some parts of the question
- The core LLM is **iteratively** prompted to reason on the previous steps and formulate a more complete answer

Step 1 (Preprocessing): Creating one Agent for each document

Step 2 (REACT): A pool of Agents to formulate the answer is retrieved based on similarity with the input question
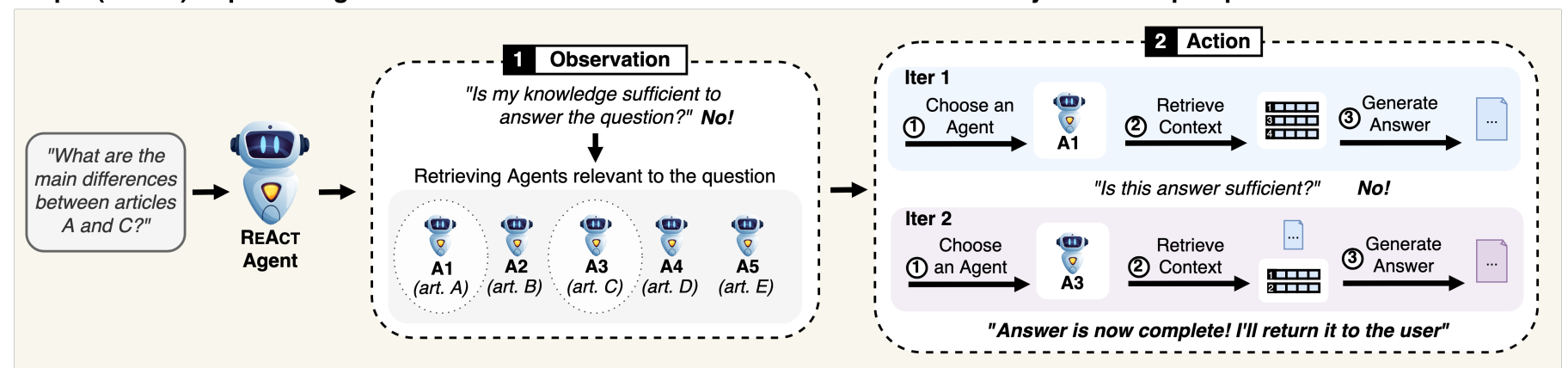
Our Publications

Publications – i

Recent works: 34 since 2020 on LMs, LLMs and AI research (+120 in total)

Knowledge Graph Learning

Frisoni G., Moro G. Phenomena Explanation from Text: Unsupervised Learning of Interpretable and Statistically Significant Knowledge. **CCIS 2020**.

Frisoni G., Moro G., Carbonaro A. Learning Interpretable and Statistically Significant Knowledge from Unlabeled Corpora of Social Text Messages: A Novel Methodology of Descriptive Text Mining. **DATA 2020**. 🏅

Frisoni G., Moro G., Carbonaro A. Unsupervised Descriptive Text Mining for Knowledge Graph Learning. **KDIR 2020**.

Frisoni G., Moro G., Carbonaro A. Towards Rare Disease Knowledge Graph Learning from Social Posts of Patients. **Rii Forum 2020**.

Semantic Parsing

Frisoni G., Moro G., Carbonaro A. A Survey on Event Extraction for Natural Language Understanding: Riding the Biomedical Literature Wave. **IEEE Access 2021**.

Frisoni G., Moro G., Balzani L. Text-to-Text Extraction and Verbalization of Biomedical Event Graphs. **COLING 2022**.

Graph Representation Learning

Frisoni G., Moro G., Carbonaro A., Carlassare G. Unsupervised Event Graph Similarity Learning on Biomedical Literature. **Sensors 2021**.

Ferrari I., Frisoni G., Italiani P., Moro G., Sartori C. Comprehensive Analysis of Knowledge Graph Embedding Techniques Benchmarked on Link Prediction. **Electronics (Graph ML SI) 2022**.

Graph Injection

Frisoni G., Italiani P., Boschi F., Moro G. Enhancing Biomedical Scientific Reviews Summarization with Graph-based Factual Evidence Extracted from Papers. **DATA 2022**. 🏅

Frisoni G., Italiani P., Moro G., Bartolini I., Boschetti MA., Carbonaro A. Graph-Enhanced Biomedical Abstractive Summarization Via Factual Evidence Extraction. **SN Computer Science 2022**.

Frisoni G., Italiani P., Moro G., Salvatori S. Cogito Ergo Summ: Abstractive Summarization of Biomedical Papers via Semantic Parsing Graphs and Consistency Rewards. **AAAI 2023**.

Moro G., Ragazzi L., Valgimigli L. Graph-Based Abstractive Summarization of Extracted Essential Knowledge for Low-Resource Scenarios. **ECAI 2023**.

Moro G., Ragazzi L., Valgimigli L., Vincenzi F., Freddi D. Revelio: Interpretable Long-Form Question-Answering. **ICLR 2024**.

Ragazzi L., Moro G., Valgimigli L., Fiorani R. Cross-Document Distillation via Graph-based Summarization of Extracted Essential Knowledge. **IEEE Transactions on Audio, Speech and Language Processing 2024**.

Publications – ii

Knowledge Distillation

Italiani P., Ragazzi L., Moro G. Enhancing Legal Question Answering with Data Generation and Knowledge Distillation from Large Language Models. **Artificial Intelligence and Law 2024**.

Semantic Text Segmentation

Moro G., Ragazzi L. Semantic Self-Segmentation for Abstractive Summarization of Long Documents in Low-Resource Regimes. **AAAI 2022**.

Moro G., Ragazzi L. Align-then-Abstract Representation Learning for Low-Resource Summarization. **Neurocomputing 2023**.

Moro G., Ragazzi L., Valgimigli L., Frisoni G., Sartori C., Marfia G. Efficient Memory-Enhanced Transformer for Long-Document Summarization in Low-Resource Regimes. **Sensors 2023**.

Retrieval-Enhanced LMs

Moro G., Ragazzi L., Valgimigli L. Discriminative Marginalized Probabilistic Neural Method for Multi-Document Summarization of Medical Literature. **ACL 2022**.

Frisoni G., Mizutani M., Moro G., Valgimigli L. BioReader: a Retrieval-Enhanced Text-to-Text Transformer for Biomedical Literature. **EMNLP 2022**.

Moro G., Ragazzi L., Valgimigli L., Molfetta L. Retrieve-and-Rank End-to-End Summarization of Biomedical Studies. **SISAP 2023**.

Frisoni G., Cocchieri A., Presepi A., Moro G. To Generate or To Retrieve? On the Effectiveness of Artificial Contexts for Medical Open-Domain Question Answering. **ACL 2024**.

Datasets and Benchmarks

Moro G., Ragazzi L., Valgimigli L. Carburacy: Summarization Models Tuning and Comparison in Eco-Sustainable Regimes with a Novel Carbon-Aware Accuracy. **AAAI 2023**.

Ragazzi L., Moro G., Guidi S., Frisoni G. LAWSUIT: a LArge expert Written SUmmarization dataset of ITalian constitutional court verdicts. **Artificial Intelligence and Law**.

Differentiable Sampling

Italiani P., Frisoni G., Moro G., Carbonaro A., Sartori C. Evidence, my Dear Watson: Abstractive Dialogue Summarization on Learnable Relevant Utterances. **Neurocomputing 2023**.

Ragazzi L., Italiani P., Moro G., Panni M. What Are You Token About? Differentiable Perturbed Top-k Token Selection for Scientific Document Summarization. **ACL 2024**.

Publications – iii

Self-Supervised Retrieval

Moro G., Salvatori S., Frisoni G Efficient Self-Supervised Metric Information Retrieval: A Bibliography Based Method Applied to COVID Literature. **Sensors 2021**.

Moro G., Valgimigli L. Efficient Self-Supervised Metric Information Retrieval: a Bibliography Based Method Applied to COVID Literature. **Sensors 2021**.

Moro G., Salvatori S., Frisoni G. Efficient Text-Image Semantic Search: A Multi-Modal Vision-Language Approach for Fashion Retrieval. **Neurocomputing 2022**.

Moro G., Salvatori S., Frisoni G. Deep Vision-Language Model for Efficient Multi-Modal Similarity Search in Fashion Retrieval. **SISAP 2022**.

Moro G., Valgimigli L., Rossi A., Casadei C., Montefiori A. Self-supervised Information Retrieval Trained from Self-Generated Sets of Queries and Relevant Documents. **SISAP 2022**.

Others

Moro G., Di Luca F., Dardari D., Frisoni G. Human Being Detection from UWB NLOS Signals: Accuracy and Generality of Advanced Machine Learning Models. **Sensors 2022**.

Moro G., Piscaglia N., Ragazzi L., Italiani P. Multi-Language Transfer Learning for Low-Resource Legal Case Summarization. **Artificial Intelligence and Law 2023**.

Bussotti J. F., Ragazzi L., Frisoni G., Moro G., Papotti P. Unknown Claims: Generation of Fact-Checking Training Examples from Unstructured and Structured Data. **EMNLP 2024**.