

Bibliografia

Questa bibliografia è organizzata per fasi della pipeline MERL-T.

Premessa Generale: Sistemi RAG, Agenti LLM e Applicazioni Legali

- [Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. \(2020\). Retrieval-augmented generation for knowledge-intensive NLP tasks.](#) *Advances in Neural Information Processing Systems*, 33, 9459-9474. (Fondamentale per il concetto di RAG)
 - [Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. \(2024\). Retrieval-Augmented Generation for Large Language Models: A Survey.](#) *arXiv preprint arXiv:2312.10997*. (Survey completo e recente sullo stato dell'arte del RAG)
 - [Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., ... & Qiu, M. \(2023\). The Rise and Potential of Large Language Model Based Agents: A Survey.](#) *arXiv preprint arXiv:2309.07864*. (Survey sugli agenti LLM, rilevante per l'approccio MoE)
 - [Zhong, H., Xiao, C., Tu, C., Li, T., Liu, Z., & Sun, M. \(2020\). How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence.](#) *arXiv preprint arXiv:2004.12158*. (Panoramica sull'applicazione dell'IA nel dominio legale)
 - [Katz, D. M., Bommarito II, M. J., & Blackman, J. \(2017\). A general approach for predicting the behavior of the Supreme Court of the United States.](#) *PLoS one*, 12(4), e0174698. (Esempio storico di applicazione di AI/ML al diritto, sebbene non con LLM)
-

1. Pre-processing e Named Entity Recognition (NER) Giuridico

- 1.1. Normalizzazione testo:
 - (Riferimenti standard su NLP pre-processing: [Manning, C. D., & Schütze, H. \(1999\). Foundations of statistical natural language processing. MIT press.](#) - Capitoli pertinenti sulla tokenizzazione e normalizzazione)
 - **Specificità giuridica:** Potrebbe non esserci un paper esclusivo sulla normalizzazione *giuridica*, ma si applicano principi generali. La sfida è definire le regole specifiche per abbreviazioni e riferimenti canonici (es. URN).
- 1.2. Named Entity Recognition (NER) giuridico:
 - [Li, S., Zhao, Z., Hu, R., Li, W., Liu, T., & Du, X. \(2018\). LegalNER: A manually annotated corpus for named entity recognition in the legal domain.](#) *Proceedings of the 27th International Conference on Computational Linguistics*. (Esempio di corpus annotato per NER legale)
 - [Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. \(2020\). LEGAL-BERT: The Muppets straight out of Law School. Findings of the Association for Computational Linguistics: EMNLP 2020](#), 2898-2904. (Modello BERT pre-addestrato su testi legali, utile come base per un NER fine-tuned)

- [**Leitner, E., Rehm, G., & Moreno-Schneider, J. \(2019\). Fine-grained named entity recognition in legal documents.**](#) *Proceedings of the Natural Legal Language Processing Workshop 2019. (Focus sul NER fine-grained nel dominio legale)*
 - [**Angelidis, S., Chalkidis, I., & Kormpa, K. \(2021\). Named entity recognition and relation extraction in the legal domain: a survey.**](#) *Artificial Intelligence and Law, 29(4), 459-507. (Survey specifica su NER e RE in ambito legale)*
-

2. Routing della query (Router MoE Addestrabile)

- [**Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. \(2017\). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer.**](#) *arXiv preprint arXiv:1701.06538. (Paper fondamentale sull'architettura Mixture-of-Experts (MoE))*
 - [**Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. \(1991\). Adaptive mixtures of local experts.**](#) *Neural computation, 3(1), 79-87. (Lavoro pionieristico sulle mixture of experts)*
 - [**Zhou, Y., Mianaji, P., Lee, R., Narayan, A., yfikaci, O., Schuurmans, D., & Jiang, J. \(2022\). Mixture-of-Experts with Expert Choice Routing.**](#) *Advances in Neural Information Processing Systems, 35, 7103-7114. (Evoluzione del routing MoE)*
 - [**Singhal, A., Tu, T., Lee, L., Toutanova, K., & Stoyanov, V. \(2022\). Large Language Models can be Lazy Learners: Analyze Shortcuts in In-Context Learning.**](#) *Findings of the Association for Computational Linguistics: ACL 2023. (Rilevante per capire come gli LLM scelgono strategie, applicabile al routing)*
 - **RLCF (Reinforcement Learning from Community Feedback):** Non esiste (al momento della ricerca) un paper canonico specifico chiamato "RLCF". Tuttavia, il concetto è una derivazione di:
 - [**Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... & Lowe, R. \(2022\). Training language models to follow instructions with human feedback.**](#) *Advances in Neural Information Processing Systems, 35, 27730-27744. (Paper fondamentale su RLHF - Reinforcement Learning from Human Feedback)*
 - L'idea di "Community Feedback" è un'estensione/adattamento di RLHF, dove il feedback proviene da una comunità di esperti (giuridici in questo caso) invece che da annotatori generici. La validazione si baserà sui principi di RLHF, adattati al contesto comunitario.
-

3. Context Augmentation (Recupero Informazioni)

- **Generale GraphRAG:**
 - [**Guo, B., Zheng, Z., Xiong, C., Yu, Q., Hu, W., Wen, J. R., & Xu, J. \(2024\). LightRAG: A Scalable and Lightweight Framework for Enhancing Large Language Models via Graph-Based Retrieval-**](#)

- [**Augmented Generation**](#). arXiv preprint arXiv:2410.05779.
(Riferimento diretto menzionato nella pipeline per GraphRAG)

 - [**Esmeili, E., He, X., Kang, W. C., & McAuley, J. \(2023\). KnowledGPT: Enhancing Large Language Models with Retrieval and Storage.**](#) arXiv preprint arXiv:2312.08682. *(Approccio che integra KG con LLM)*
- **3.1. Recupero da database vettoriale (ANN):**
 - [**Johnson, J., Douze, M., & Jégou, H. \(2019\). Billion-scale similarity search with GPUs.**](#) IEEE Transactions on Big Data, 7(3), 535-547. *(Describe FAISS, una libreria comune per ANN)*
 - [**Wang, L., Wang, Z., Wu, W., & Zhang, X. \(2020\). Milvus: A Purpose-Built Vector Data Management System.**](#) Proceedings of the 2021 International Conference on Management of Data. *(Describe Milvus)*
 - [**Reimers, N., & Gurevych, I. \(2019\). Sentence-BERT: Sentence embeddings using Siamese BERT-networks.**](#) arXiv preprint arXiv:1908.10084. *(Tecnica comune per generare embedding di qualità per la ricerca semantica)*
- **3.2. Recupero da Knowledge Graph (KG):**
 - [**Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., Melo, G. D., Gutierrez, C., ... & Zimmermann, A. \(2021\). Knowledge graphs.**](#) ACM Computing Surveys (CSUR), 54(4), 1-37. *(Survey completa sui Knowledge Graph)*
 - [**Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. \(2013\). Translating embeddings for modeling multi-relational data.**](#) Advances in neural information processing systems, 26. *(Tecnica fondamentale per gli embedding nei KG)*
 - **Specificità giuridica dei KG:**
 - [**Casellas, N. \(2011\). Legal ontology engineering: methodologies, modelling trends, and the ontology of professional judicial knowledge.**](#) Springer Science & Business Media. *(Testo di riferimento sull'ingegneria delle ontologie legali)*
 - [**Francesconi, E. \(2014\). Legal knowledge graphs for the Semantic Web.**](#) Semantic Web, 5(2), 87-90. *(Focus sui KG specifici per il diritto)*
- **3.3. Recupero dinamico via API:**
 - *(Non richiede validazione scientifica per l'atto della chiamata API in sé, ma la qualità e affidabilità delle API (VisualLexAPI, API Sentenze) sono cruciali e potrebbero essere oggetto di valutazione esterna o reportistica).*

4. Costruzione dei prompt per LLM

- [**Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. \(2020\). Language models are few-shot learners.**](#) Advances in neural information processing systems, 33, 1877-1901. *(Paper su GPT-3 che ha reso popolare il prompting)*

- [**Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. \(2022\). Chain-of-thought prompting elicits reasoning in large language models.**](#) *Advances in Neural Information Processing Systems*, 35, 24824-24837. (*Tecnica di prompting per migliorare il ragionamento*)
 - [**Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. \(2022\). Large language models are zero-shot reasoners.**](#) *Advances in Neural Information Processing Systems*, 35, 22199-22213. (*Un'altra tecnica di prompting fondamentale*)
 - [**Gao, T., Yao, Y., & Chen, D. \(2021\). SimCSE: Simple Contrastive Learning of Sentence Embeddings.**](#) *arXiv preprint arXiv:2104.08821*. (*Rilevante se gli embedding influenzano la selezione del contesto inserito nel prompt*)
 - [**Liu, Y., Iter, D., Xu, Y., Wang, S., Xu, R., & Zhu, C. \(2021\). What Makes Good In-Context Examples for GPT-3?**](#) *arXiv preprint arXiv:2101.06804*. (*Analisi su come strutturare i prompt con esempi*)
-

5. Inferenza LLM (Moduli Esperti)

- **Fine-tuning per domini specifici:**
 - [**Howard, J., & Ruder, S. \(2018\). Universal language model fine-tuning for text classification.**](#) *arXiv preprint arXiv:1801.06146*. (*ULMFiT, tecnica pionieristica di fine-tuning*)
 - [**Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. \(2020\). Don't stop pretraining: Adapt language models to domains and tasks.**](#) *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. (*Dimostra l'importanza dell'adattamento al dominio*)
 - [**Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. \(2020\). LEGAL-BERT: The Muppets straight out of Law School.**](#) (*Citato anche per NER, ma fondamentale per il fine-tuning su dati legali*)
 - **Modelli ottimizzati per RAG:**
 - La ricerca su modelli *specificamente* ottimizzati per RAG è in corso. Spesso si utilizzano LLM generici con buone capacità di seguire istruzioni e utilizzare il contesto fornito. La validazione qui dipende dalle performance del modello scelto sul task specifico.
 - [**Yu, W., Zhang, Z., Chen, Z., Zhang, H., Meng, F., Wang, X., ... & Zhao, T. \(2023\). Chain-of-Note: Enhancing Robustness in Retrieval-Augmented Language Models.**](#) *arXiv preprint arXiv:2311.09210*. (*Esempio di tecnica per migliorare la robustezza dei modelli RAG*)
 - **Utilizzo di LoRA (Low-Rank Adaptation):**
 - [**Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. \(2021\). LoRA: Low-Rank Adaptation of Large Language Models.**](#) *arXiv preprint arXiv:2106.09685*. (*Tecnica efficiente per il fine-tuning*)
-

6. Sintesi e combinazione risposte (Sintetizzatore MoE Addestrabile)

- **Riferimenti MoE:** Vedi Sezione 2 (Routing). L'applicazione qui è sulla sintesi anziché sul routing, ma i principi architetturali sono simili.
 - **Riferimenti RLCF:** Vedi Sezione 2 (Routing). L'addestramento del sintetizzatore si basa sugli stessi principi di apprendimento dal feedback della comunità di esperti.
 - **Combinazione di output da diversi modelli/fonti:**
 - [Wang, B., Xu, C., & Ma, T. \(2021\). Understanding and improving knowledge distillation. Proceedings of the 38th International Conference on Machine Learning.](#) (Concetti di distillazione e combinazione di modelli, anche se non specifici per MoE/sintesi RAG)
 - [Du, N., Dai, A. M., Liu, X., Lee, H., & Le, Q. V. \(2022\). GLaM: Efficient Scaling of Language Models with Mixture-of-Experts.](#) arXiv preprint arXiv:2112.06905. (Esempio di architettura MoE su larga scala, utile per capire come i gate combinano gli output)
-

7. Post-processing e output finale

- **7.2. Tracking e citazione fonti:**
 - [Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. \(2024\). Retrieval-Augmented Generation for Large Language Models: A Survey.](#) (La sezione sulla valutazione dei sistemi RAG spesso discute la fedeltà e l'attribuzione delle fonti)
 - [Nakano, R., Hilton, J., Balaji, S., Wu, J., Ouyang, L., Kim, T., ... & Schulman, J. \(2021\). Webgpt: Browser-assisted question-answering with human feedback.](#) arXiv preprint arXiv:2112.09332. (Esempio di sistema che pone enfasi sulla citazione delle fonti)
 - [Menick, J., Trebacz, M., Mikulik, V., Rae, J. W., Elsen, E., & Simonyan, K. \(2022\). Teaching language models to support answers with verified quotes.](#) arXiv preprint arXiv:2203.11147. (Focus specifico sul supporto delle risposte con citazioni verificate)
- **7.3. Filtri di sicurezza/qualità:**
 - [Gehman, S., Gururangan, S., Sap, M., Choi, Y., & Smith, N. A. \(2020\). RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models.](#) Findings of the Association for Computational Linguistics: EMNLP 2020. (Valutazione della tossicità)
 - [Perez, F., He, J., He, H., Pathak, N., & Sap, M. \(2022\). Ignore the Noise: Robust Conditional Diffusion Models via Diffusion Guidance.](#) arXiv preprint arXiv:2211.10121. (Potenzialmente rilevante per filtri basati su modelli)
 - La validazione qui spesso si basa su dataset specifici per bias, tossicità, e adesione a linee guida (es. Constitutional AI: [Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., ... & Clark, J. \(2022\). Constitutional AI: Harmlessness from AI Feedback.](#) arXiv preprint arXiv:2212.08073.)