D1.2 - Report on 2nd Year Management and Project Activities

Project: NEREO - Neural Information Retrieval and NLP Systems
Grant Agreement: PRIN 2022
Deliverable ID: D1.2
Work Package: WP1 - Management and Project Activities
Due Date: M24
Lead Beneficiary: UNIPI

---

1. Executive Summary

This report serves as the final management deliverable for the NEREO project, covering the activities of the second year (M13-M24). The primary focus of Y2 was the **convergence** of the two research lines (Upstream and Cascading) into a unified prototype, the **validation** of results through rigorous evaluation, and the **communication** of these results to the wider community. We successfully executed the project closure procedures, including final financial audits, and hosted a Technology Transfer Event that attracted stakeholders from industry and academia.

2. Comprehensive Management Activities

2.1 Project Closure and Administration (Task 1.2)

Management in the final year ensured strict adherence to MUR guidelines for project termination.

- **Financial Consolidation:** We coordinated the collection of financial statements from all Research Units (UNIPI, UNIROMA1). This involved a detailed review of all expenditure (personnel, equipment, travel) to match the approved budget. A final "Financial Activity Report" was generated and uploaded to the ministerial portal.
- **Audit Preparation:** To ensure compliance, we organized a pre-audit internal review. All supporting documents (timesheets, meeting minutes, public procurement records for GPU purchases) were archived and digitized for the final audit.
- **IPR & Licensing:** We established a clear Intellectual Property Rights framework for the project's outputs. The codebases (DanteLLM, Robust-RecSys) were released under **Apache 2.0**, ensuring they are "Open by Default" (per DMP) while being permissive enough for industrial adoption.

2.2 Coordination and Integration (Task 1.1)

The complexity of cascading systems required tight coordination between partners.

- **Integration Workshops:** We held two major physical meetings (Rome M18, Pisa M22). These "hackathon" sessions were dedicated to the technical integration of the Upstream (WP2) and Downstream (WP3) components. Specifically, the "E2Rank" module (UNIPI) was successfully piped into the "QPP-RA" aggregator (UNIROMA1).
- **Advisory Board Review:** The External Advisory Board was convened at M23 to review the "Prototype Evaluation" (D4.3). Their feedback was instrumental in refining the final "Use Case" definitions.

2.3 Dissemination and Technology Transfer (Task 4.2)

Related Event: NEREO Tech Day (Rome, M24)

To maximize impact (Objective O4), we moved beyond academic papers to active technology transfer.

- **NEREO Tech Day:** We organized a full-day public workshop titled *"The Future of Neural Search in Italy"*. The event featured:
  - **Demos:** Live demonstration of DanteLLM and the RAG pipeline.
  - **Industry Panel:** A roundtable with attendees from Italian tech companies (e.g., Translated, Seeweb) discussing the adoption of Italian LLMs.
  - **Attendance:** Over 50 participants attended (physically and virtually), meeting our Dissemination KPI.
- **Open Science:** We ensured 100% Open Access for the 20+ papers published during the project, using a mix of Green (arXiv) and Gold (Open Access Journals) routes.

---

3. Project Activities Summary (Year 2)

Year 2 was characterized by high-intensity research and development.

3.1 WP2: Finalizing Upstream Systems

The team delivered the "final" versions of the upstream components:

- **Efficiency:** The release of **E2Rank** solved the latency bottleneck, allowing the cascading system to operate in near real-time.
- **Semantic Precision:** The **Eclipse** project provided the interpretability layer, allowing developers to debug why a document was retrieved.

3.2 WP3: Finalizing Cascading Systems

The team completed the "downstream" reasoners:

- **Aggregators: QPP-RA** proved that "System-as-a-User" is a viable paradigm, obtaining superior ranking performance by aggregating LLM judgments.
- **Mechanistic Control:** The **Wavelet** discovery gave us the theoretical tools to control how LLMs attend to retrieved context.

---

4. Final Scientific Output

The project concluded with a strong portfolio of publications in 2025, including top-tier venues like **ACL, CVPR, ECIR, and ICTIR**.

2025

Journals

- **Consistent Counterfactual Explanations via Anomaly Control and Data Coherence.**
  Maria Movin, Federico Siciliano, Rui Ferreira, Fabrizio Silvestri, and Gabriele Tolomei.
  IEEE Trans. Artif. Intell., 6(4), 794-804.
  [DOI: 10.1109/TAI.2024.3496616](DOI: 10.1109/TAI.2024.3496616)

- **Projection-Displacement-Based Query Performance Prediction for Embedded Space of Dense Retrievers.**
  Suchana Datta, Guglielmo Faggioli, Nicola Ferro, Debasis Ganguly, Cristina Ioana Muntean, Raffaele Perego, and Nicola Tonellotto.

ACM Trans. Inf. Syst., 44(1), Article 7.

Conferences

- **Beyond Position: the emergence of wavelet-like properties in Transformers.**
  Valeria Ruscio, Umberto Nanni, and Fabrizio Silvestri.
  ACL 2025.
  URL

- **Task Singular Vectors: Reducing Task Interference in Model Merging.**
  Antonio Andrea Gargiulo, Donato Crisostomi, Maria Sofia Bucarelli, Simone Scardapane, Fabrizio Silvestri, and Emanuele Rodolà.
  CVPR 2025.
  DOI: 10.1109/CVPR52734.2025.01742

- **E2Rank: Efficient and Effective Layer-Wise Reranking.**
  Cesare Campagnano, Antonio Mallia, Jack Pertschuk, and Fabrizio Silvestri.
  ECIR 2025.
  DOI: 10.1007/978-3-031-88714-7_41

- **QPP-RA: Aggregating Large Language Model Rankings.**
  Filippo Betello, Matteo Russo, Paul Dütting, Stefano Leonardi, and Fabrizio Silvestri.
  ICTIR 2025.
  DOI: 10.1145/3731120.3744575

- **Eclipse: Contrastive Dimension Importance Estimation with Pseudo-Irrelevance Feedback for Dense Retrieval.**
  Giulio D'Erasmo, Giovanni Trappolini, Fabrizio Silvestri, and Nicola Tonellotto.
  ICTIR 2025.
  DOI: 10.1145/3731120.3744579

- **Are Convolutional Sequential Recommender Systems Still Competitive? Introducing New Models and Insights.**
  Federico Siciliano, Antonio Purificato, Filippo Betello, Nicola Tonellotto, and Fabrizio Silvestri.
  IJCNN 2025.
  DOI: 10.1109/IJCNN64981.2025.11229036

- **A Theoretical Analysis of Recommendation Loss Functions under Negative Sampling.**
  Giulia Di Teodoro, Federico Siciliano, Nicola Tonellotto, and Fabrizio Silvestri.
  IJCNN 2025.
  DOI: 10.1109/IJCNN64981.2025.11228603

- **Neural Prioritisation for Web Crawling.**
  Ophir Frieder, Nicola Ferro, Joel Mackenzie, Marc Najork, Edie Rasmussen, and Nicola Tonellotto.
  ICTIR 2025.
  DOI: 10.1145/3731120.3744597

- **Efficient Recommendation with Millions of Items by Dynamic Pruning of Sub-Item Embeddings.**
  Nicola Tonellotto, Aleksandr Petrov, and Craig Macdonald.
  SIGIR 2025.
  DOI: 10.1145/3726302.3729963

- **A Reproducibility Study of PLAID.**
  Sean MacAvaney and Nicola Tonellotto.
  arXiv preprint.
  URL: arXiv:2404.14989

2024

Conferences

- **Dimension Importance Estimation for Dense Information Retrieval.**
  Giulio D'Erasmo and Nicola Tonellotto.
  SIGIR 2024.
  [DOI: 10.1145/3626772.3657691](DOI: 10.1145/3626772.3657691)

- **Faster Learned Sparse Retrieval with Block-Max Pruning.**
  Antonio Mallia, Torsten Suel, and Nicola Tonellotto.
  SIGIR 2024.
  [DOI: 10.1145/3626772.3657906](DOI: 10.1145/3626772.3657906)

5. Conclusion

The NEREO project has successfully concluded. We have delivered a novel theoretical framework for cascading systems, a suite of robust open-source tools, and a thriving community around Italian Neural Search. The management activities ensured this was achieved on time and within budget.