

POZNAN UNIVERSITY OF TECHNOLOGY

FACULTY OF COMPUTING AND TELECOMMUNICATION
Institute of Computing Science

Doctoral dissertation

# FROM LITERATURE TO MODELS: SEMI-AUTOMATIC CONSTRUCTION OF BIOMEDICAL PETRI NET-BASED MODELS

Adam Aron Rynkiewicz

Supervisor
prof. dr hab. inż. Piotr Formanowicz

POZNAŃ 2026

Podziękowania

*Chciałbym serdecznie podziękować mojemu promotorowi, prof. dr. hab. inż. Piotrowi Formanowiczowi, za nieocenione wsparcie, cierpliwość i wszystkie wskazówki, których udzielał mi podczas pracy nad doktoratem. Dziękuję za bycie mentorem, na którego pomoc i doświadczenie zawsze mogłem liczyć.*

*Szczególne podziękowania kieruję do mojej Mamy, której poświęcenie, wsparcie i wychowanie miały ogromny wpływ na to, kim jestem i dokąd udało mi się dojść.*

*Chciałbym również podziękować dr Raulowi Palmie za wyrozumiałość i elastyczność, które pozwoliły mi skutecznie łączyć obowiązki zawodowe z pracą nad doktoratem. Jego wsparcie w tym zakresie znacząco ułatwiło mi realizację obu tych zobowiązań.*

*Na końcu chciałbym podziękować mojej żonie za nieustanne wsparcie, wyrozumiałość i obecność przez cały okres pracy nad doktoratem. Jej pomoc i wiara we mnie były dla mnie niezwykle ważne.*

## Abstract

This dissertation investigates methods for transforming unstructured biomedical literature into structured, computationally actionable knowledge that can support the construction of Petri net-based models. The work is motivated by the rapidly growing volume of biomedical publications and the increasing difficulty of integrating fragmented scientific findings into coherent representations of complex biological systems. Modern systems biology requires models that capture interactions among multiple molecular, cellular, and physiological processes, yet constructing such models manually remains highly time-consuming and dependent on extensive expert knowledge. Petri nets provide an attractive modeling formalism in this context because they combine an intuitive graphical representation of biological processes with rigorous mathematical semantics and can be applied even when detailed quantitative parameters are unavailable.

The central objective of this work was therefore to develop computational methods that reduce the effort required to extract biologically meaningful entities and relations from scientific literature while preserving the precision, interpretability, and traceability required for scientific modeling. Rather than aiming for complete automation, the research adopts a semi-automated, human-in-the-loop approach in which artificial intelligence supports domain experts by identifying and organizing candidate knowledge that can subsequently be validated and incorporated into formal models.

The methodological contributions cover several stages of biomedical information extraction. First, the research investigates Named Entity Recognition, including improvements to an existing model and the development of a new model family for multi-type, span-aware entity extraction. These studies demonstrate that biomedical entity recognition can be made computationally efficient and scalable while retaining high predictive performance. At the same time, the results show that accurate identification of entities alone is insufficient for reconstructing coherent biological processes, highlighting the need for subsequent stages of information extraction and interpretation.

The work then addresses Entity Linking through the development of a new model and a modular framework combining contextual semantic representations, candidate retrieval, disambiguation, and reranking. While these approaches achieve competitive linking performance with relatively modest training resources, their application to biological modeling reveals an important conceptual limitation: many biologically relevant entities are inherently context-dependent and cannot always be mapped to fixed database identifiers without losing information relevant to the modeled process. Consequently, strict normalization may in some cases reduce, rather than improve, the structural correctness of computational models.

A further contribution concerns the transition from traditional multi-stage information extraction pipelines based on Named Entity Recognition, Entity Linking, and Relation Extraction toward unified, relation-centered extraction enabled by Large Language Models. Reasoning-capable models make it possible to identify entities and relations jointly, thereby improving coherence and reducing error propagation between separate pipeline components. These findings motivated the development of a platform supporting the semi-automatic extraction of biologically meaningful interactions and their incorporation into Petri net-based models. Evaluation showed that a substantial portion of relevant interactions can be recovered directly from biomedical text, while also revealing limitations associated with implicit knowledge, information contained in figures, document segmentation, and long-range dependencies.

To improve the reliability of the extraction process, the dissertation further introduces an evidence-grounded annotation approach that preserves document context and improves traceability. This methodology is extended by incorporating verification across multiple language models.

Together, these developments shift the extraction process toward a precision-oriented regime in which fewer relations are produced, but individual candidates are better supported by textual evidence and are therefore more suitable for expert validation and scientific modeling.

The research was developed and evaluated in the biomedical domain of high-density lipoprotein metabolism and function, including processes related to lipid transport, inflammation, and cardiovascular biology. A Petri net model describing HDL-related processes affected by local inflammation and oxidative stress was also constructed.

Overall, the dissertation demonstrates both the potential and the current limitations of artificial intelligence for knowledge extraction and biological model reconstruction. Its main conclusion is that fully automated reconstruction of complete biological pathways from scientific literature alone is neither presently feasible nor necessarily desirable. A more effective strategy is to use artificial intelligence to generate reliable, evidence-grounded candidate relations that accelerate model development while preserving expert control and scientific rigor. Future work should extend this approach by incorporating additional information sources and improving the representation of implicit knowledge and long-range contextual dependencies.

## Streszczenie

Niniejsza rozprawa doktorska dotyczy metod przekształcania nieustrukturyzowanej literatury biomedycznej w ustrukturyzowaną wiedzę, wspierającą budowę modeli opartych na sieciach Petriego. Motywacją do podjęcia tych badań jest szybko rosnąca liczba publikacji biomedycznych oraz wynikające z tego coraz większe trudności w integrowaniu rozproszonych wyników badań w spójne reprezentacje złożonych systemów biologicznych. Współczesna biologia systemowa wymaga modeli zdolnych do odwzorowania interakcji pomiędzy wieloma procesami molekularnymi, komórkowymi i fizjologicznymi, jednak ich ręczne konstruowanie pozostaje procesem bardzo czasochłonnym i wymagającym rozległej wiedzy eksperckiej. Sieci Petriego stanowią w tym kontekście atrakcyjny formalizm modelowania, ponieważ łączą intuicyjną reprezentację graficzną procesów biologicznych z rygorystyczną semantyką matematyczną, a jednocześnie mogą być stosowane również wtedy, gdy szczegółowe parametry ilościowe nie są dostępne.

Głównym celem pracy było opracowanie metod obliczeniowych ograniczających nakład pracy konieczny do wydobywania z literatury naukowej biologicznie istotnych obiektów i relacji, przy jednoczesnym zachowaniu precyzji, interpretowalności i możliwości śledzenia pochodzenia informacji, wymaganych w modelowaniu naukowym. Zamiast dążyć do pełnej automatyzacji, przyjęto podejście półautomatyczne, oparte na współpracy człowieka z systemem, w którym sztuczna inteligencja wspiera ekspertów dziedzinowych poprzez identyfikowanie i porządkowanie potencjalnie istotnej wiedzy, która następnie może zostać zweryfikowana i włączona do modeli formalnych.

Wkład metodyczny obejmuje kilka etapów ekstrakcji informacji biomedycznej. W pierwszej kolejności analizowano rozpoznawanie nazwanych obiektów, w tym ulepszenia istniejącego modelu oraz opracowanie nowej rodziny modeli umożliwiających wielotypową ekstrakcję obiektów z uwzględnieniem ich zakresów tekstowych. Badania te pokazują, że rozpoznawanie obiektów biomedycznych może być realizowane w sposób wydajny obliczeniowo i skalowalny przy zachowaniu wysokiej jakości predykcji. Jednocześnie wyniki wskazują, że sama poprawna identyfikacja obiektów nie jest wystarczająca do odtwarzania spójnych procesów biologicznych, co podkreśla potrzebę stosowania kolejnych etapów ekstrakcji i interpretacji informacji.

W dalszej części pracy podjęto zagadnienie normalizacji obiektów poprzez opracowanie nowego modelu oraz modularnego środowiska łączącego kontekstowe reprezentacje semantyczne, wyszukiwanie kandydatów, ujednoznacznianie oraz ranking. Chociaż zaproponowane rozwiązania osiągają konkurencyjną skuteczność przy relatywnie niewielkich zasobach treningowych, ich zastosowanie do modelowania biologicznego ujawnia istotne ograniczenie koncepcyjne: wiele biologicznie istotnych obiektów ma charakter silnie zależny od kontekstu i nie zawsze może być odwzorowane na stałe identyfikatory baz danych bez utraty informacji ważnych z punktu widzenia modelowanego procesu. W konsekwencji ścisła normalizacja może w niektórych przypadkach pogarszać, zamiast poprawiać, poprawność strukturalną modeli obliczeniowych.

Kolejny istotny element pracy dotyczy przejścia od tradycyjnych, wieloetapowych potoków ekstrakcji informacji, opartych na rozpoznawaniu obiektów, normalizacji obiektów i ekstrakcji relacji, do zintegrowanej ekstrakcji skoncentrowanej na relacjach, umożliwionej przez duże modele językowe. Modele zdolne do rozumowania umożliwiają jednoczesną identyfikację obiektów i relacji, co sprzyja zachowaniu większej spójności oraz ogranicza propagację błędów pomiędzy kolejnymi komponentami potoku. Wyniki te stały się podstawą do opracowania platformy wspierającej półautomatyczną ekstrakcję biologicznie istotnych interakcji oraz ich włączanie do modeli opartych na sieciach Petriego. Ewaluacja wykazała, że znaczną część istotnych interakcji można wydobyć bezpośrednio z tekstów biomedycznych, a jednocześnie ujawniła ograniczenia związane z wiedzą niejawną, informacjami zawartymi na rycinach, segmentacją dokumentów oraz zależności-

5:2832327401

ami nielokalnymi.

W celu zwiększenia wiarygodności procesu ekstrakcji w rozprawie zaproponowano również podejście do adnotacji oparte na dowodach, które zachowuje kontekst dokumentu i poprawia możliwość śledzenia źródeł informacji. Metodologia ta została następnie rozszerzona o mechanizmy weryfikacji wykorzystujące wiele modeli językowych. Łącznie rozwiązania te przesuwają proces ekstrakcji w kierunku podejścia zorientowanego na precyzję, w którym generowanych jest mniej relacji, lecz poszczególni kandydaci są lepiej uzasadnieni materiałem źródłowym, a tym samym bardziej odpowiedni do walidacji i wykorzystania w modelowaniu.

Badania opracowano i oceniono w domenie biomedycznej związanej z metabolizmem i funkcją lipoprotein o wysokiej gęstości, obejmującej między innymi procesy transportu lipidów, zapalenia oraz biologii układu sercowo-naczyniowego. Opracowano również model sieci Petriego opisujący procesy związane z HDL pod wpływem lokalnego zapalenia oraz stresu oksydacyjnego.

Podsumowując, rozprawa pokazuje zarówno potencjał, jak i aktualne ograniczenia sztucznej inteligencji w zakresie ekstrakcji wiedzy i rekonstrukcji modeli biologicznych. Głównym wnioskiem jest stwierdzenie, że w pełni automatyczna rekonstrukcja kompletnych szlaków biologicznych wyłącznie na podstawie literatury naukowej nie jest obecnie ani w pełni wykonalna, ani koniecznie pożądana. Bardziej efektywną strategią jest wykorzystanie sztucznej inteligencji do generowania wiarygodnych, opartych na dowodach kandydackich relacji, które przyspieszają rozwój modeli, jednocześnie zachowując kontrolę eksperta i rygor naukowy. Dalsze prace powinny koncentrować się na włączaniu dodatkowych źródeł informacji oraz doskonaleniu sposobów reprezentowania wiedzy niejawnej i nielokalnych zależności kontekstowych.

6:1007102078

# Contents


**1 Introduction** **3**

**2 Theoretical Foundations** **7**

2.1 Introduction . . . 7
2.2 Biological foundations - lipoproteins . . . 7
2.3 Petri nets . . . 9
2.3.1 Fundamental concepts . . . 9
2.3.2 Analysis methods . . . 11
2.4 Large Language Models . . . 12
2.4.1 Neural Networks . . . 12
2.4.2 Transformer-based Large Language Models . . . 13
2.4.3 Fine-tuning and Optimization of LLMs . . . 15
2.4.4 Vector Representations and Databases . . . 16
2.5 Named Entity Recognition . . . 17
2.6 Entity Linking . . . 18
2.7 Relation Extraction . . . 19

**3 Literature review** **20**

3.1 Introduction . . . 20
3.2 High-Density Lipoproteins and Inflammation . . . 20
3.2.1 Introduction . . . 20
3.2.2 HDL Structure, Function, and Heterogeneity . . . 20
3.2.3 Paradigm Shift: From HDL Quantity to Functionality . . . 20
3.2.4 HDL as a Modulator of Inflammation . . . 21
3.2.5 Dysfunctional HDL in Inflammatory States . . . 21
3.2.6 Bidirectional Relationship Between HDL and Inflammation . . . 21
3.2.7 Clinical Implications . . . 21
3.2.8 Synthesis . . . 22
3.3 Named Entity Recognition . . . 22
3.3.1 Traditional Discriminative Approaches . . . 22
3.3.2 Zero-Shot and Open-Vocabulary NER . . . 22
3.3.3 LLM-Based Generative Extraction and Distillation . . . 22
3.3.4 Synthesis . . . 23
3.4 Entity Linking Landscape . . . 23
3.4.1 Bi-Encoder and Cross-Encoder Architectures . . . 23
3.4.2 End-to-End Frameworks . . . 24
3.4.3 Synthesis and Limitations . . . 24
3.5 Relation Extraction in Biomedical Text Mining . . . 24




7:4068492855

3.5.1 Evolution of Relation Extraction Methods . . . 24
3.5.2 Document-level Relation Extraction . . . 25
3.5.3 Large Language Models for Relation Extraction . . . 25
3.5.4 Synthesis and Limitations . . . 26
3.6 Petri net-based models automatic creation . . . 26
3.6.1 Petri Nets in Biomedical Process Modeling . . . 26
3.6.2 Knowledge transformation . . . 26
3.6.3 Knowledge synthesis . . . 27
3.6.4 Synthesis and Limitations . . . 27

**4 HDL in inflammation Petri net-based model 29**
4.1 Introduction . . . 29
4.2 Baseline model . . . 29
4.3 Model extension - inflammation . . . 32
4.4 Model extension - treatment . . . 36
4.5 In silico knock-out analysis . . . 39
4.6 Chapter summary . . . 42

**5 Named Entity Recognition 48**
5.1 Introduction . . . 48
5.2 UniversalNER enhancements . . . 49
5.2.1 UniversalNER model family . . . 49
5.2.2 Quantization procedure . . . 50
5.2.3 Evaluation datasets . . . 50
5.2.4 Evaluation results . . . 51
5.3 ner-on model . . . 52
5.4 Chapter summary . . . 54

**6 Entity Linking 55**
6.1 Introduction . . . 55
6.2 Entity Disambiguation . . . 55
6.2.1 Overview . . . 56
6.2.2 Datasets . . . 57
6.2.3 Models . . . 57
6.2.4 Evaluation . . . 58
6.3 FELA – End-to-end Entity Linking . . . 59
6.3.1 FELA overview . . . 59
6.3.2 FELA evaluation . . . 60
6.4 Chapter summary . . . 61

**7 Semi-Automatic Construction of Petri Net-Based Models 63**
7.1 Introduction . . . 63
7.2 Watson – Initial Implementation . . . 64
7.2.1 Watson Overview . . . 64
7.2.2 Datasets . . . 66
Fine-tuning datasets . . . 66
Evaluation datasets . . . 67
Evidence attribution dataset . . . 67

8:7392881928

7.2.3 Fine-Tuned Models . . . 68
7.2.4 Processing pipeline . . . 68
7.2.5 Results . . . 70
Manual evaluation . . . 70
Atherosclerosis . . . 71
HDL metabolism . . . 72
Automatic evaluation . . . 73
Summary and implications . . . 75
7.3 Watson – Improvements . . . 76
7.3.1 Iterative Information Extraction . . . 77
7.3.2 Extractor-critic-consensus . . . 79
7.3.3 Evaluation . . . 80
7.3.4 Summary . . . 83

**8 Summary** **86**

**Bibliography** **89**

9:5665940121

Table 0.1: List of abbreviations.

| **Abbreviation** | **Full term** |
|---|---|
| ABCG1 | ATP-binding Cassette Sub-family G Member 1 |
| ABCA1 | ATP-binding Cassette Transporter A1 |
| ANN | Artificial Neural Network |
| BioNLP | Biomedical Natural Language Processing |
| CD36 | Cluster Of Differentiation 36 |
| CRF | Conditional Random Field |
| CI | Confidence Interval |
| CNN | Convolutional Neural Network |
| DNN | Deep Neural Network |
| DPO | Direct Preference Optimization |
| ED | Entity Disambiguation |
| EL | Entity Linking |
| ECC | Extractor-Critic-Consensus |
| GIN | Global Integrative Network |
| HDL-C | HDL Cholesterol |
| HL | Hepatic Lipase |
| HMM | Hidden Markov Model |
| HDLR | High-density Lipoprotein Receptor |
| INDRA | Integrated Network And Dynamical Reasoning Assembler |
| IFN-$\gamma$ | Interferon Gamma |
| IDL | Intermediate-density Lipoproteins |
| IIE | Iterative Information Extraction |
| KB | Knowledge Base |
| $k$-NN | $k$-Nearest Neighbor |
| LLMs | Large Language Models |
| Lp(a) | Lipoprotein(a) |
| LSTM | Long Short-term Memory |
| LDLR | Low-density Lipoprotein Receptor |
| LDL | Low-density Lipoproteins |
| MUI | Material UI |
| MeSH | Medical Subject Heading |
| MCT | Minimal Common Transition |
| NLRP3 | Nod-like Receptor Protein 3 |
| NER | Named Entity Recognition |
| NLP | Natural Language Processing |
| ORPO | Odds Ratio Preference Optimization |
| PPAR | Peroxisome Proliferator-activated Receptor |
| PN | Petri Net |
| RNN | Recurrent Neural Network |
| RLHF | Reinforcement Learning From Human Feedback |
| RE | Relation Extraction |
| RCT | Reverse Cholesterol Transport |

10:7321912815

| **Abbreviation** | **Full term** |
|---|---|
| SFT | Supervised Fine-tuning |
| TNF-$\alpha$ | Tumor Necrosis Factor Alpha |
| VLDL | Very-low-density Lipoproteins |
| apoA-I | Apolipoprotein A-i |
| BET | Bromodomain And Extraterminal |
| CVD | Cardiovascular Disease |
| CETP | Cholesteryl Ester Transfer Protein |
| cAMP | Cyclic Adenosine Monophosphate |
| FA | Fatty Acid |
| HDL | High-density Lipoprotein |
| LCAT | Lecithin-cholesterol Acyltransferase |
| LPL | Lipoprotein Lipase |
| ODEs | Ordinary Differential Equations |
| PON1 | Paraoxonase-1 |
| PLTP | Phospholipid Transfer Protein |
| rHDL | Reconstituted Hdl |
| RVX-208 | Apabetalone |
| SR-B1 | Scavenger Receptor Class B Type 1 |
| SAA | Serum Amyloid A |
| S1P | Sphingosine-1-phosphate |

11:3623113475

# Chapter 1

# Introduction

In recent years, the volume of scientific output in the biomedical domain has grown at an unprecedented rate, with over one million articles published annually [43]. While this abundance of information reflects rapid scientific progress, it also introduces a fundamental challenge: the knowledge contained within these publications remains largely fragmented and difficult to integrate. Transforming dispersed findings into coherent, actionable insights typically requires extensive manual effort, making the process both time-consuming and prone to inconsistencies. As a result, despite the continuous expansion of biomedical knowledge, its full potential is often underutilized due to limited interoperability and integration.

This problem is particularly evident in modern biology, where the scale and complexity of available data have increased dramatically over the last two decades. Since the beginning of the twenty-first century, advances in high-throughput technologies, including genome sequencing, transcriptomics, proteomics, metabolomics, and other large-scale experimental approaches, have transformed biological research. These technologies have enabled the generation of vast amounts of data describing biological systems at multiple levels of organization. However, they have also revealed that biological phenomena cannot be fully understood by studying individual molecules, genes, proteins, or pathways in isolation. Instead, biological function emerges from complex interactions among many components, often involving feedback loops, nonlinear dependencies, spatial and temporal regulation, and cross-talk between different biological processes.

Consequently, there has been a growing need for systems-level analysis in the biological sciences. Rather than focusing exclusively on isolated components, systems biology aims to investigate biological processes as integrated and dynamic networks. This perspective is especially important in areas where multiple processes interact simultaneously, such as metabolism, inflammation, immune regulation, lipid transport, and disease progression. In such contexts, a purely reductionist approach may be insufficient, because the behavior of the whole system may not be directly predictable from the properties of its individual parts. Systems biology therefore provides a conceptual and computational framework for organizing heterogeneous biological knowledge, formulating hypotheses, and studying how local interactions give rise to global system behavior.

Models play a central role in this framework. They provide formal representations of biological knowledge and make it possible to reason about the structure and dynamics of complex systems. A model can serve several purposes: it may summarize current knowledge, reveal inconsistencies or gaps in understanding, support simulation-based experimentation, generate new hypotheses, or help interpret experimental observations. In biomedical research, models are particularly valuable because they allow researchers to investigate processes that are difficult, expensive, or ethically problematic to study directly in wet-laboratory settings. Computational simulations can



12:1030187167

be repeated under controlled conditions, modified systematically, and used to explore alternative biological scenarios.

At the same time, constructing reliable models of biological systems is a challenging task. Biological systems often involve many interacting entities and processes, and the available information is usually distributed across numerous publications, databases, and experimental reports. Moreover, this information may be incomplete, ambiguous, context-dependent, or even contradictory. As a result, model construction requires not only domain knowledge but also careful interpretation and integration of heterogeneous evidence. These difficulties are further amplified by the rapidly increasing volume of biomedical literature, which makes comprehensive manual analysis increasingly impractical.

Systems biology, a subfield of computational biology, offers a promising framework for addressing this issue. By adopting a holistic, system-level perspective, it enables the study of complex biological processes as interconnected networks rather than isolated components. This systems approach allows for the discovery of interactions and dependencies that may remain hidden under traditional reductionist methodologies. Among the various modeling paradigms employed in systems biology, Petri nets have emerged as a versatile formalism.

Several mathematical and computational approaches can be used to model biological systems. One of the most established classes of methods is based on ordinary differential equations (ODEs), which describe changes in the concentrations or activities of biological entities over time. ODE-based models are powerful and well suited for quantitative simulation when sufficient kinetic parameters and experimental measurements are available [79]. However, they also have important limitations. In many biological contexts, precise quantitative data are difficult or impossible to obtain. Reaction rates, concentrations, binding affinities, and regulatory parameters may be unknown, highly variable between conditions, or dependent on biological context. This makes the construction and calibration of detailed quantitative models difficult, especially for large-scale systems.

For this reason, graph-based and network-based modeling approaches are often used as alternatives or complements to purely quantitative methods. These approaches make it possible to represent entities and interactions without requiring complete numerical parameterization. They are therefore particularly useful when the available knowledge is mainly qualitative or semi-quantitative, as is often the case in biomedical literature. Network models can capture structural relationships between biological components and support analyses of connectivity, causality, reachability, and system organization. Within this broader family of approaches, Petri nets are especially attractive because they combine an intuitive graphical representation with a rigorous mathematical foundation.

Petri nets provide a simple yet powerful framework for computational modeling of dynamic systems. Structurally, they have the form of bipartite graphs consisting of two types of nodes: *places*, which represent states or entities within a system, and *transitions*, which denote interactions or events that modify these states. The directed arcs connecting places and transitions define the flow of information, material, or control through the modeled system. In biological applications, places may correspond, for example, to molecules, proteins, cellular states, biological conditions, or higher-level physiological entities, whereas transitions may represent reactions, transport processes, activation events, inhibition events, or regulatory mechanisms.

An important advantage of Petri nets is that they can represent both the structure and behavior of a system. The graphical structure provides an interpretable view of biological interactions, while the mathematical semantics enable formal analysis and simulation. Depending on the level of detail and the available data, different classes and extensions of Petri nets may be used, including

13:4075994280

stochastic, time, continuous, hybrid, and colored Petri nets. This flexibility makes the formalism applicable across multiple levels of biological organization, ranging from biochemical reactions and cellular processes to organ-level interactions and even population dynamics. Petri nets are therefore well suited to systems biology, where models often need to integrate heterogeneous mechanisms and operate under conditions of incomplete quantitative knowledge.

In addition to their expressive power, Petri nets, as well as other computational modeling approaches, address a practical limitation of experimental biology: the high cost and limited reproducibility of wet-laboratory experiments. Computational simulations, in contrast, can be executed repeatedly at relatively low cost, enabling systematic exploration of hypotheses and scenarios that would be difficult or infeasible to test experimentally. Moreover, formal models can make assumptions explicit, support reproducible reasoning, and provide a basis for comparing alternative mechanistic explanations. These properties are particularly relevant in biomedical domains characterized by complex interactions between molecular, cellular, and physiological processes.

Despite these advantages, a significant bottleneck remains in the application of computational modeling approaches: the time and expertise required to construct accurate and reliable models. Model development typically involves an extensive review of the literature, extraction of relevant entities and relationships, and careful integration of this information into a coherent formal representation. This process is inherently labor-intensive and does not scale well with the rapidly growing volume of scientific publications. The problem is not limited to finding relevant articles. It also includes identifying biological entities, recognizing their synonyms and database identifiers, extracting mechanistic relations, resolving ambiguities, and determining whether a given statement is relevant for a specific modeling context.

This dissertation addresses this challenge by proposing methods to semi-automate the transformation of unstructured biomedical text into structured, computationally actionable models. Importantly, the level of automation is deliberately balanced. While fully automated pipelines based on machine learning and artificial intelligence may be technically feasible, they often introduce errors that can compromise model precision and reliability. In scientific applications, where interpretability and trust are essential, such trade-offs are particularly critical. Therefore, the approach adopted in this work emphasizes semi-automation, combining the efficiency of computational methods with the accuracy ensured by human oversight.

The contributions of this dissertation lie at the intersection of systems biology, Natural Language Processing (NLP), and computational modeling. In particular, it focuses on the extraction and integration of biomedical knowledge related to lipoproteins, and on leveraging this information to construct Petri net-based models. Lipoproteins constitute an important biological application domain because they participate in lipid transport, metabolism, inflammation, and cardiovascular disease-related processes. Their biological roles involve numerous interacting molecular and cellular mechanisms, making them suitable for systems-level analysis. To support the construction of formal models in this domain, key NLP tasks - Named Entity Recognition (NER), Entity Linking (EL), and Relation Extraction (RE) - are addressed as essential components of the proposed methodology.

The remainder of this dissertation is organized as follows. Chapter 2 introduces the theoretical background necessary to understand the presented work. Given the interdisciplinary nature of the research, this chapter begins with an overview of the biological foundations of lipoproteins, which constitute the primary application domain. It then presents the formal definition and properties of Petri nets, with particular attention to their relevance for modeling biological systems. This is followed by an introduction to Large Language Models (LLMs) and their role in modern NLP. Finally, the chapter discusses the fundamental NLP tasks of NER, EL, and RE.

14:4270158508

Chapter 3 provides a comprehensive review of the current state of the art in the relevant domains, including lipoprotein biology, and the aforementioned NLP tasks. Chapter 4 presents a Petri net-based model describing the relationship between high-density lipoproteins (HDLs) and inflammation, together with its structural and knock-out analyses. The results show that control of the HDL metabolic network is concentrated in a small number of regulatory transitions and that these control points change substantially between inflammatory and therapeutic conditions. In the pathological state, network behavior is dominated by impaired HDL maturation, oxidative damage, and disrupted adiponectin regulation, whereas therapeutic interventions shift control toward inflammation suppression, reduction of oxidative stress, and restoration of cholesterol transport. Overall, the analysis demonstrates that HDL functionality emerges from context-dependent network interactions and illustrates the usefulness of Petri net modeling for identifying regulatory bottlenecks and system-wide responses to therapeutic interventions.

Subsequent chapters focus on the methodological contributions of this work. Chapter 5 introduces the approaches developed for NER, including efficient quantized variants of `UniversalNER` and the multi-type, span-aware `ner-on` models. The results show that accurate entity recognition is an important first step, but is insufficient on its own for constructing coherent biological models.

Chapter 6 addresses entity linking through contextual representations and semantic retrieval, culminating in the modular `FELA` pipeline. While the proposed methods achieve competitive performance, their application to biological modeling reveals limitations of strict knowledge-base normalization for state-dependent and context-specific entities. This motivates a shift from entity-centered normalization toward relation-centered representation.

Chapter 7 presents methods for semi-automatic construction of Petri net-based models from biomedical literature. It introduces base and improved relation extraction and verification strategies that emphasize evidence grounding, traceability, and precision, while reducing unsupported relations. The results further suggest that smaller domain-adapted models can perform comparably to larger reasoning models, highlighting the importance of extraction strategy and complementary information sources over model scale alone.

Finally, Chapter 8 summarizes the main findings of the dissertation, discusses its limitations, and outlines directions for future research.

15:1065283929

# Chapter 2

# Theoretical Foundations

## 2.1 Introduction

This chapter establishes the theoretical foundations necessary for understanding the contributions and context of the present dissertation. It provides an overview of key concepts spanning both the biological domain and the computational methodologies used throughout this work.

The chapter is structured as follows. Section 2.2 introduces the biological background of lipoproteins, outlining their structure, function, and relevance in biomedical research. Section 2.3 presents the formal framework of Petri nets, including essential definitions and properties relevant to modeling complex systems. Section 2.4 discusses the fundamental concepts underlying LLMs, with emphasis on their architecture and role in modern NLP tasks.

Subsequent sections focus on key information extraction tasks addressed in this dissertation. Section 2.5 introduces the problem of NER, detailing its objectives and methodological evolution. Section 2.6 presents the EL task, highlighting its connection to NER and its role in grounding textual entities in structured KBs. Finally, Section 2.7 describes the RE problem, emphasizing its challenges and importance in capturing semantic relationships between entities.

Together, these sections provide the necessary interdisciplinary background required to contextualize the methods and results presented in the subsequent chapters.

## 2.2 Biological foundations - lipoproteins

Lipids are essential biomolecules involved in energy storage, membrane structure, and cellular signaling. However, because most lipids are hydrophobic and therefore poorly soluble in aqueous environments such as blood plasma, specialized transport systems are required to allow their circulation throughout the body. This role is fulfilled by lipoproteins, which are heterogeneous macromolecular complexes composed of lipids and proteins that enable the efficient transport of triglycerides, cholesterol, and other lipid species between tissues and organs [35, 41].

Lipoproteins possess a characteristic spherical structure consisting of a hydrophobic core surrounded by an amphipathic surface layer. The core is primarily composed of triglycerides and cholesteryl esters, whereas the outer shell contains phospholipids, unesterified cholesterol, and specific proteins known as apolipoproteins. Apolipoproteins contribute to the structural stability of lipoprotein particles and serve as ligands for cellular receptors and cofactors for enzymes involved in lipid metabolism [35, 18].

Based on their density, size, lipid composition, and metabolic function, circulating lipoproteins are conventionally classified into seven major categories:



16:5777246135

- **Chylomicrons** are the largest and least dense lipoprotein particles. They are synthesized by enterocytes in the small intestine following dietary fat absorption and are particularly enriched in triglycerides. Their primary function is to transport dietary triglycerides, cholesterol, and fat-soluble vitamins from the intestine to peripheral tissues and subsequently to the liver [64, 35].

- **Chylomicron remnants** are generated during the progressive hydrolysis of chylomicron triglycerides by lipoprotein lipase. Compared with their precursors, these particles are smaller and relatively enriched in cholesterol. Chylomicron remnants are rapidly removed from circulation through receptor-mediated uptake by the liver [64].

- **Very-low-density lipoproteins (VLDL)** are synthesized and secreted by the liver to transport endogenously produced triglycerides to peripheral tissues. Although smaller than chylomicrons, VLDL particles remain triglyceride-rich and play a central role in lipid distribution throughout the body [18, 41].

- **Intermediate-density lipoproteins (IDL)** are metabolic intermediates formed during the catabolism of VLDL particles. As triglycerides are progressively removed, IDL particles become relatively enriched in cholesteryl esters. These particles may either be cleared by the liver or undergo further metabolism to form low-density lipoproteins [18].

- **Low-density lipoproteins (LDL)** constitute the principal carriers of cholesterol in human plasma, transporting approximately two-thirds of circulating cholesterol. LDL particles originate primarily from the metabolic conversion of VLDL and IDL. Elevated plasma concentrations of LDL cholesterol are strongly associated with the development of atherosclerotic cardiovascular disease due to the accumulation and retention of LDL particles within the arterial wall [36, 35]. Increased levels of LDL are frequently observed in individuals with obesity, type 2 diabetes mellitus, hypertriglyceridemia, and various chronic inflammatory conditions [41].

- **HDLs** are the smallest and most protein-rich lipoproteins. HDL particles play a critical role in reverse cholesterol transport, a process through which excess cholesterol is removed from peripheral tissues and delivered to the liver for excretion or recycling. Beyond their involvement in cholesterol transport, HDL particles exhibit anti-inflammatory, antioxidant, antithrombotic, and cytoprotective properties that may contribute to their protective effects against atherosclerosis [113, 35, 95, 63, 34, 23, 62].

- **Lipoprotein(a) [Lp(a)]** is a specialized lipoprotein particle structurally similar to LDL. It consists of an LDL-like particle containing apolipoprotein B-100 covalently linked via a disulfide bond to apolipoprotein(a). Elevated plasma concentrations of Lp(a) are largely genetically determined and represent an independent risk factor for atherosclerotic cardiovascular disease [131, 70].

The metabolism of lipoproteins is a highly dynamic and interconnected process involving continuous particle remodeling, enzymatic modification, and receptor-mediated uptake. Disturbances in lipoprotein metabolism can result in dyslipidemia, a major contributor to cardiovascular disease, which remains one of the leading causes of morbidity and mortality worldwide [140, 96]. Consequently, understanding the structure, function, and metabolism of lipoproteins is fundamental for investigating the mechanisms underlying lipid disorders and their clinical consequences [36, 35].

17:1026007040

## 2.3 Petri nets

Petri nets (PNs) are a mathematical formalism with an intuitive graphical representation for modeling and analyzing complex systems. They are particularly useful for systems in which multiple processes may occur concurrently, asynchronously, or under specific structural constraints. Owing to their ability to capture both structural and functional aspects of biological processes, they have become a widely used framework for modeling biochemical reactions, signaling pathways, metabolic networks, regulatory mechanisms, and other biological systems. This section introduces the fundamental concepts of Petri nets and outlines the principal analytical methods employed in this thesis. A major advantage of Petri nets is that they combine an intuitive graphical notation with a rigorous mathematical foundation. The graphical representation supports human interpretation and communication of complex models, while the underlying formalism enables systematic analysis, simulation, and verification. This combination is especially valuable in biological applications, where models must often be understandable to domain experts while remaining sufficiently formal for computational analysis. Furthermore, Petri nets can represent causal dependencies, concurrency, feedback loops, conflicts, and alternative pathways, all of which are frequently encountered in biological systems.

### 2.3.1 Fundamental concepts

Petri nets were originally developed by Karl Adam Petri [101] as a formalism for describing distributed systems. A classical Petri net has the structure of a bipartite directed graph consisting of two disjoint classes of nodes: places and transitions. Formally, a Petri net can be defined as a tuple:

$$Q = (P, T, F, W, M_0)$$

where $P = p_1, p_2, \ldots, p_n$ is a finite set of places, $T = t_1, t_2, \ldots, t_m$ is a finite set of transitions, $F \subseteq (P \times T) \cup (T \times P)$ is a set of arcs, $W \to \mathbb{Z}^+$ is a weight function assigning positive integer weights to arcs, and $M_0 \to \mathbb{N}$ is the initial marking of the net. The sets of places and transitions are disjoint, i.e., $P \cap T = \emptyset$, but are not empty $P \cup T \neq \emptyset$ [94, 68]. In biological applications, places typically represent entities such as genes, proteins, metabolites, lipoprotein particles, cellular states, or other molecular and biological components. Transitions correspond to biological events and processes, including biochemical reactions, molecular transformations, transport mechanisms, degradation, secretion, activation, inhibition, binding, dissociation, or regulatory interactions. Directed arcs define the relationships between these components and determine how the state of the system may change. Arcs from places to transitions indicate the entities or conditions required for a process to occur, whereas arcs from transitions to places indicate the entities or conditions produced or modified by that process.

The dynamic behavior of a Petri net is represented by tokens, which are distributed among places. The arrangement of tokens at a given moment, referred to as a marking, describes the current state of the modeled system. In biological models, the interpretation of tokens depends on the modeling assumptions. In a qualitative model, tokens may represent the presence, availability, or activity of a biological entity. In more quantitative or semi-quantitative variants, they may approximate molecule counts, concentrations, activity levels, or other measurable quantities. This flexibility allows Petri nets to be adapted to different levels of biological detail and to situations in which precise numerical parameters are unavailable.

18:1823418986

System evolution is governed by the transition firing rule. A transition is considered enabled when each place in its preset contains at least as many tokens as specified by the weight of the corresponding incoming arc. Once enabled, the transition may fire, consuming tokens from places in its preset and producing tokens in places in its postset according to the weights of the associated arcs. More formally, firing a transition $t_j$ changes the marking $M$ into a new marking $M'$ according to:

$$M'(p_i) = M(p_i) - W(p_i, t_j) + W(t_j, p_i), \forall p_i \in P$$

where $W(p_i, t_j)$ denotes the weight of the arc from place $p_i$ to transition $t_j$, and $W(t_j, p_i)$ denotes the weight of the arc from transition $t_j$ to place $p_i$ [14]. If a given arc does not exist, its weight is treated as zero. Through successive firing events, the network changes its marking and thereby simulates the dynamics of the underlying biological process [94, 66].

The ability to represent changes in state through discrete firing events makes Petri nets suitable for modeling biological processes in which the occurrence of one event depends on the availability of specific entities or conditions. For example, a biochemical reaction may require the presence of substrates, an enzymatic process may depend on an active enzyme, and a signaling event may occur only when a receptor is activated. Petri nets can express such dependencies explicitly. Moreover, if several transitions are enabled at the same time, they may represent processes that can occur independently or concurrently. This is important in biological systems, where multiple reactions and regulatory events often take place in parallel.

Figure 2.1 presents a simple example illustrating the use of Petri nets to model the chemical reaction $2\,\mathrm{Na} + 2\,\mathrm{H_2O} \longrightarrow 2\,\mathrm{NaOH} + \mathrm{H_2}$. In the initial marking, two tokens are assigned to the places representing sodium (Na) and water ($H_2O$). The transition corresponds to the chemical reaction itself, while the arc weights reflect the reaction stoichiometry. After the transition fires, the resulting marking contains tokens representing the reaction products, sodium hydroxide (NaOH) and hydrogen ($H_2$).

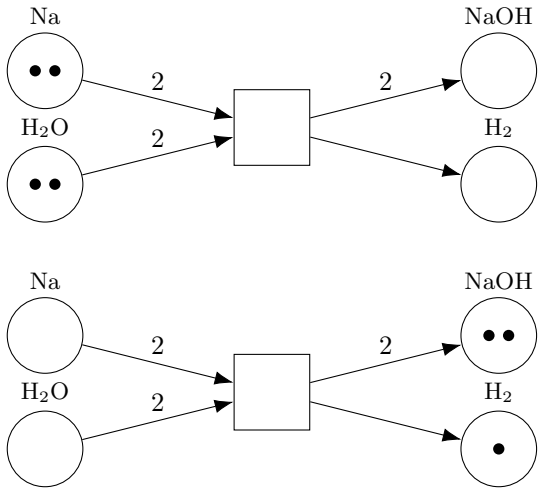



FIGURE 2.1: Petri net representation of the reaction $2\,\mathrm{Na} + 2\,\mathrm{H_2O} \longrightarrow 2\,\mathrm{NaOH} + \mathrm{H_2}$. The upper panel shows the initial marking, where tokens represent the reactants. Arc weights correspond to stoichiometric coefficients. The lower panel depicts the marking obtained after the transition fires, producing the reaction products according to the specified stoichiometry.

Although the example shown in Figure 2.1 is intentionally simple, the same formalism can be applied to highly complex biological systems. The modular architecture of Petri nets enables

19:1033650768

the construction of large-scale models encompassing numerous interacting components, ranging from intracellular signaling cascades to organism-wide physiological processes. Complex models can be built by combining smaller modules that represent specific subprocesses, such as enzymatic reactions, transport mechanisms, regulatory pathways, or disease-related interactions. This modularity is particularly useful in systems biology, where knowledge is often organized around partially understood biological subsystems.

Another important feature of Petri nets is their capacity to represent different types of relationships within a single formal framework. For instance, they can model sequential dependencies, where one process must precede another; branching behavior, where alternative pathways are possible; synchronization, where several conditions must be satisfied before a process can occur; and feedback, where the products of a process influence earlier stages of the system. These properties make Petri nets well suited to the representation of biological mechanisms, which often involve complex regulatory architectures rather than simple linear chains of events.

Several extensions of the classical Petri net formalism have been introduced to increase its expressiveness. In stochastic Petri nets, firing events are associated with probability distributions or rates, allowing random behavior and variability to be represented [37]. Timed Petri nets include explicit temporal information, making it possible to model delays or durations of biological processes [148]. Continuous and hybrid Petri nets allow token values to be treated as continuous quantities, which is useful when modeling concentrations or population-level behavior [4]. Colored Petri nets introduce token types, enabling more compact representation of systems with repeated structures or multiple similar entities [59]. Although the present thesis focuses primarily on Petri net-based modeling in the context of knowledge representation and model construction, these extensions demonstrate the flexibility of the formalism and its applicability to a broad range of biological modeling problems.

### 2.3.2 Analysis methods

The graphical representation of a Petri net provides an intuitive overview of the modeled system. Places are typically depicted as circles, transitions as rectangles, and tokens as dots or numerical values located within places. While this visualization is useful for understanding a network structure, formal analysis requires a mathematical representation of the net.

The incidence matrix provides an algebraic representation of a Petri net and serves as the basis for several analysis methods [94]. For a Petri net with places $P = \{p_1, \ldots, p_n\}$ and transitions $T = \{t_1, \ldots, t_m\}$, the incidence matrix $A \in \mathbb{Z}^{n \times m}$ is defined such that row $i$ corresponds to place $p_i$ and column $j$ corresponds to transition $t_j$. Its element $a_{ij}$ gives the net change in the marking of place $p_i$ resulting from the firing of transition $t_j$. More specifically,

$$a_{ij} = w(t_j, p_i) - w(p_i, t_j),$$

where $w(t_j, p_i)$ denotes the weight of the arc from transition $t_j$ to place $p_i$, representing the number of tokens produced in $p_i$, and $w(p_i, t_j)$ denotes the weight of the arc from place $p_i$ to transition $t_j$, representing the number of tokens consumed from $p_i$.

The incidence matrix forms the basis for the computation of transition invariants ($t$-invariants). A $t$-invariant is a natural numbers vector $x \in \mathbb{N}^m$ satisfying the equation $A \cdot x = 0$. Biologically, individual transitions represent elementary processes, whereas a $t$-invariant corresponds to a set of transitions whose collective execution do not change the state of the modeled system. Consequently, $t$-invariants can be interpreted as balanced subprocesses or functional pathways operating within the system.

20:6110455083

For each $t$-invariant $x$, its support is defined as

$$s(x) = \{t_j : x_j > 0,\ j = 0, 1, 2, \ldots, m\},$$

that is, the set of transitions whose corresponding entries in $x$ are nonzero. The support therefore identifies the transitions involved in the process represented by the invariant.

Analysis of $t$-invariants is widely used to identify subnetworks capable of operating in a balanced steady state, in which the firing of the participating transitions does not change the overall marking of the net. A Petri net is said to be covered by $t$-invariants if every transition belongs to the support of at least one $t$-invariant. Such coverage indicates that every transition can participate in at least one balanced subprocess of the modeled system, although it does not imply that all transitions are active simultaneously or that every biologically relevant behavior is captured by steady-state processes [94, 66].

An additional structural analysis can be performed using Minimal Common Transition (MCT) sets. Two transitions belong to the same MCT set if they occur in the supports of exactly the same $t$-invariants. Equivalently, every $t$-invariant support either contains both transitions or contains neither of them. This relation partitions the set of transitions into mutually disjoint MCT sets. MCT sets containing more than one transition are referred to as non-trivial MCT sets. In biological Petri net-based models, such non-trivial MCT sets frequently correspond to functional modules, as the grouped transitions are structurally coupled and therefore participate together in subprocesses represented by the corresponding $t$-invariants [115].

Beyond structural analyses based on invariants and network topology, Petri net-based models can also be investigated through simulation. One particularly informative approach is knockout analysis, in which selected transitions are disabled to mimic the loss of specific biological functions. The resulting changes in network behavior can then be examined to assess the importance of individual processes and to identify their influence on the overall system dynamics [114]. Such analyses provide valuable insights into the functional organization and robustness of biological networks.

## 2.4 Large Language Models

This section introduces the theoretical and technical foundations underlying LLMs, which is a central component of modern NLP systems. The discussion begins with an overview of artificial neural networks and deep learning paradigms, followed by the transformer architecture that enabled the emergence of LLMs. Particular emphasis is placed on decoder-only transformer models, which dominate current large-scale generative systems. Subsequently, techniques for fine-tuning and optimizing LLMs are outlined. Finally, related concepts, including vector representations and vector databases, are introduced as essential components of contemporary LLM-based systems.

### 2.4.1 Neural Networks

Artificial Neural Networks (ANNs) are computational models inspired by the organization of biological nervous systems. They consist of interconnected computational units called artificial neurons, arranged into layers. Information flows through the network from an input layer, through one or more hidden layers, to an output layer. During training, the parameters of the network are optimized using gradient-based learning algorithms, allowing the model to approximate complex non-linear functions [44, 86, 49].

Each neuron computes a weighted combination of its inputs followed by a non-linear activation function:

21:5240553644

$$a = \sigma\left(\sum_{i=1}^{n} w_i x_i + b\right),$$

where $x_i$ denotes the input features, $w_i$ the learnable weights, $b$ the bias term, and $\sigma$ an activation function such as ReLU or sigmoid.

A Deep Neural Network (DNN) is formed by composing multiple layers of neurons. The computation performed by the $l$-th layer can be expressed as

$$\mathbf{h}^{(l)} = \sigma\left(W^{(l)} h^{(l-1)} + b^{(l)}\right),$$

where $h^{(0)} = x$ represents the input vector, $W^{(l)}$ is the weight matrix, $b^{(l)}$ the bias vector, and $h^{(l)}$ the output representation of layer $l$. By stacking many such transformations, deep neural networks learn increasingly abstract representations of the input data.

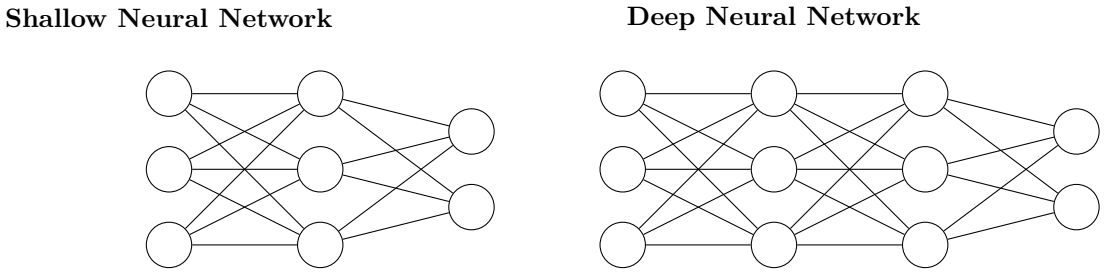



FIGURE 2.2: Comparison of shallow and deep neural networks. Deep networks contain multiple hidden layers, enabling hierarchical feature learning.

Before transformers became dominant, sequential language modeling primarily relied on Recurrent Neural Networks (RNNs). Unlike feed-forward networks, RNNs maintain an internal hidden state that is updated after processing each input token, enabling information to propagate through the sequence [33]. Formally,

$$h_t = f(x_t, h_{t-1}),$$

where $x_t$ denotes the current input and $h_{t-1}$ the hidden state from the previous time step.

Although RNNs naturally process sequential data, they struggle to capture long-range dependencies due to the vanishing and exploding gradient problems during training.

Long Short-Term Memory (LSTM) networks were proposed to address these limitations by introducing memory cells regulated through input, forget, and output gates [53]. These gating mechanisms allow relevant information to be preserved over longer sequences while discarding irrelevant context, significantly improving performance on many NLP tasks.

### 2.4.2 Transformer-based Large Language Models

The work presented in this dissertation focuses on a specific class of LLMs, namely transformer-based decoder-only architectures. In this subsection, the key concepts underlying these models are defined and discussed.

Transformer models operate on sequences of tokens rather than raw text. A token is the basic unit processed by the model and may correspond to an entire word, part of a word, punctuation, or an individual character depending on the tokenizer. Example of tokenization is presented in Figure 2.3.

22:1116847000

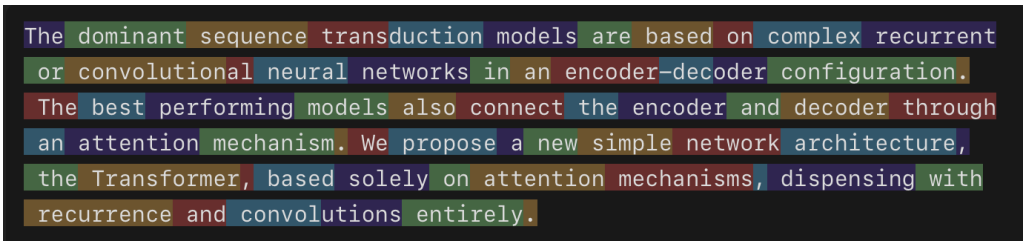



FIGURE 2.3: Example of text tokenization performed by a subword tokenizer. Each colored segment represents an individual token, which serves as the basic unit processed by a transformer model. Tokens may correspond to complete words, subwords, or punctuation symbols, enabling efficient representation of arbitrary text using a fixed-size vocabulary.

Using subword tokenization enables the model to efficiently represent rare or previously unseen words while maintaining a manageable vocabulary size.

The transformer architecture, introduced by Vaswani et al. [135], represents a significant departure from earlier sequence modeling approaches by relying entirely on attention mechanisms rather than recurrence or convolution. The original transformer consists of two main components: an encoder and a decoder. The encoder processes the input sequence into a contextualized representation, while the decoder generates the output sequence based on both the encoded input and previously generated tokens. This architecture was initially proposed for machine translation tasks, where the encoder and decoder operate on different languages.

At the core of the transformer architecture lies the self-attention mechanism. Unlike recurrent models, which process tokens sequentially, self-attention allows every token in a sequence to directly attend to every other token. For each token, the model computes attention weights indicating the relevance of surrounding tokens when constructing its contextual representation.

Given query ($Q$), key ($K$), and value ($V$) matrices derived from the input representations, scaled dot-product attention is computed as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V,$$

where $d_k$ denotes the dimensionality of the key vectors. The resulting weighted combination allows the model to capture both local and long-range contextual dependencies while remaining fully parallelizable during training.

While the encoder - decoder architecture is well-suited for sequence-to-sequence tasks, many NLP applications, particularly those involving text generation, can be effectively addressed using only the decoder component. Decoder-only transformers are trained using an autoregressive objective. During training, the model receives a sequence of tokens

$$(x_1, x_2, \ldots, x_{t-1})$$

and learns to predict the next token $x_t$. Consequently, the joint probability of a sequence is factorized as

$$P(x_1, \ldots, x_n) = \prod_{t=1}^{n} P(x_t | x_1, \ldots, x_{t-1}).$$

During inference, text is generated iteratively: after predicting one token, the generated token is appended to the input and used to predict the following token. This process continues until a stopping criterion, such as an end-of-sequence token or maximum output length, is reached. An example of simplified decoder-only architecture is presented in Figure 2.4.

23:6662743157

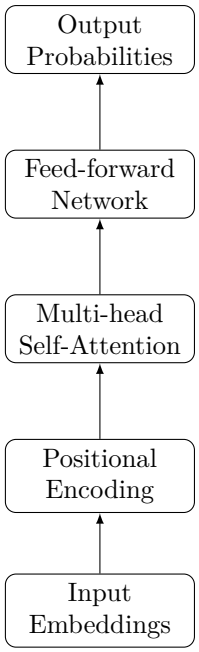



FIGURE 2.4: Simplified architecture of a decoder-only transformer block. In practice, multiple identical blocks are stacked to form a large language model.

Models trained using the autoregressive objective are referred to as generative language models. Rather than assigning labels to input text, they learn the probability distribution of natural language and are therefore capable of generating coherent text, answering questions, summarizing documents, translating languages, or producing computer code. GPT-family models is one of the most prominent examples of such decoder-only generative architectures [12].

Besides text generation, transformer architectures are widely used to compute semantic vector representations of text, commonly known as embeddings. An embedding is a dense numerical vector in which semantically similar texts are represented by nearby points according to a chosen distance metric, such as cosine similarity. Instead of representing words or documents symbolically, embeddings encode their semantic meaning into numerical features learned during model training [111]. These representations form the basis of semantic search, document clustering, recommendation systems, and retrieval-augmented generation.

Another emerging paradigm is the use of LLM-based agents. These systems extend traditional LLM capabilities by incorporating mechanisms for planning, tool usage, and iterative reasoning [142, 146]. Agents are designed to autonomously accomplish specified tasks under given constraints, often interacting with external tools or environments. This paradigm shifts LLM usage from passive text generation toward more active and goal-directed behavior.

### 2.4.3 Fine-tuning and Optimization of LLMs

Although LLMs are typically pre-trained on large-scale corpora using self-supervised objectives, their performance on specific downstream tasks can be significantly improved through fine-tuning. Fine-tuning involves adapting a pre-trained model to a particular task or domain using labeled or instruction-based datasets [107]. This process allows models to specialize their general linguistic knowledge toward more narrowly defined objectives.

A common initial stage of adaptation is Supervised Fine-Tuning (SFT), in which the model is trained on curated input-output pairs that reflect desired behavior. SFT is widely used to

24:4565435729

transform base language models into instruction-following systems and serves as a foundation for further alignment procedures [98].

Beyond SFT, more advanced alignment techniques have been proposed to better align model outputs with human preferences. Reinforcement Learning from Human Feedback (RLHF) remains a prominent approach, where a reward model - trained on human preference data - is used to guide policy optimization [98]. However, RLHF can be complex and computationally expensive. As a result, alternative methods such as Direct Preference Optimization (DPO) [106] and Odds Ratio Preference Optimization (ORPO) [54] have been introduced. These approaches simplify the alignment process by directly optimizing preference data without requiring an explicit reward model or reinforcement learning loop, thereby improving training stability and efficiency.

Another important aspect of LLM optimization is model compression and deployment efficiency. Quantization techniques reduce the numerical precision of model parameters (e.g., from 32-bit floating point to 8-bit or 4-bit representations), thereby decreasing memory footprint and improving inference speed with minimal impact on performance [2].

Additionally, model merging methods have recently gained attention as a way to combine multiple fine-tuned models into a single model without retraining [42]. Such approaches enable the integration of diverse capabilities learned from different datasets or tasks.

Finally, alternative adaptation paradigms that do not require updating the model's parameters have also become widely used. One such approach is prompt engineering, which involves carefully designing the input prompt to guide the model toward producing the desired output [82]. Prompt engineering may include providing explicit instructions, specifying the desired output format, assigning the model a particular role, or supplying relevant contextual information. By structuring prompts appropriately, users can substantially influence model behavior without modifying the underlying weights.

A related technique is in-context learning, in which the model is provided with one or more examples of the target task directly within the prompt [92]. Rather than learning through parameter optimization, the model infers the intended task from the examples and generalizes this behavior to new inputs during inference. Depending on the number of demonstrations included, this paradigm is commonly referred to as zero-shot, one-shot, or few-shot learning. In-context learning has proven effective for a wide range of downstream tasks, particularly when task-specific fine-tuning is impractical or unavailable.

Collectively, these approaches complement parameter-based adaptation methods by enabling flexible and efficient customization of LLM behavior at inference time. Together, they form a comprehensive toolkit for adapting, aligning, and deploying large language models across a wide range of applications.

### 2.4.4 Vector Representations and Databases

Vector representations produced by embedding models are often stored and managed using specialized data structures known as vector databases. These systems are designed to efficiently handle high-dimensional data and support similarity search operations, typically based on metrics such as cosine similarity, dot product, or Euclidean distance [61]. Unlike traditional relational databases, which are optimized for exact matching, filtering, and structured queries, vector databases are optimized for approximate nearest-neighbor search in continuous embedding spaces. This makes them particularly suitable for applications in which semantic similarity is more important than exact lexical overlap.

25:5197951972

The need for such systems arises from the scale and dimensionality of modern embedding-based representations. In many natural language processing tasks, documents, sentences, biomedical entities, or relations may be represented as dense vectors containing hundreds or thousands of dimensions. Performing exhaustive pairwise comparison between a query vector and all vectors in a large corpus can be computationally expensive and impractical. Vector databases address this limitation by using indexing techniques that accelerate nearest-neighbor retrieval while maintaining an acceptable trade-off between search accuracy and computational efficiency. Common approaches include inverted file indexes, product quantization, and other approximate nearest-neighbor methods [61].

Vector databases, such as FAISS [61], ChromDB [24] or Milvus [136, 45], enable scalable retrieval of semantically similar items from large corpora. In a typical workflow, textual units such as abstracts, paragraphs, sentences, or entity descriptions are first transformed into embedding vectors. These vectors are then indexed in the database together with relevant metadata, such as document identifiers, source information, entity labels, or publication details. At query time, the input query is encoded into the same vector space and compared against the indexed representations. The database returns the most similar items, which can then be used for downstream processing, ranking, filtering, or interpretation.

## 2.5 Named Entity Recognition

NER is a fundamental sub-task within the broader field of NLP, concerned with the automatic identification of entity mentions in unstructured text and their classification into predefined semantic categories such as persons, organizations, locations, or others [88, 138]. As an essential component in information extraction pipelines, NER facilitates downstream applications including question answering, Knowledge Base (KB) construction, and information retrieval.

The development of NER systems has undergone several distinct methodological phases. Early approaches were predominantly rule-based, relying on handcrafted linguistic patterns and domain-specific heuristics [87, 74, 104]. While such systems demonstrated high precision in constrained settings, they suffered from limited scalability and poor adaptability to new domains. Subsequently, statistical and traditional machine learning methods, including Hidden Markov Models (HMMs), Conditional Random Fields (CRFs), and feature-engineered classifiers, became prominent [87, 116, 15]. These approaches enabled more robust generalization by leveraging annotated corpora, although they remained heavily dependent on manual feature design.

The advent of deep learning marked a significant paradigm shift in NER research. Neural architectures such as RNNs, Convolutional Neural Networks (CNNs), and, more recently, transformer-based models (e.g., BERT and its domain-specific variants like BioBERT) have demonstrated substantial performance improvements by automatically learning contextualized representations of text [76, 72]. These models reduced the reliance on feature engineering and enabled more effective handling of complex linguistic phenomena.

In recent years, LLMs have further advanced the state of the art in NER [147, 145, 125]. Owing to their large-scale pretraining on diverse corpora and their capacity for in-context learning, LLM-based approaches exhibit strong generalization capabilities, even in low-resource or zero-shot settings. Fine-tuned LLMs, as well as instruction-based and prompt-driven methods, have shown competitive or superior performance compared to traditional supervised systems, particularly in handling ambiguous, rare, or previously unseen entity types.

Despite the methodological diversity, the overarching objective of NER remains consistent - to accurately extract and categorize all relevant entity mentions within a given text corpus. To

26:8977261579

illustrate this task, consider the following example sentence derived from the Tweeki dataset [48]:

> "AHAHAH RT @politico The Roots use entrance music to bash Michele Bachmann on Jimmy Fallon"

An effective NER system should identify the following entity mentions: *politico*, *The Roots*, *Michele Bachmann*, and *Jimmy Fallon*. While the recognition of canonical person entities such as *Michele Bachmann* and *Jimmy Fallon* is typically straightforward, more nuanced cases - such as *The Roots* (a musical group) or *politico* (a media organization) - present challenges, particularly for earlier approaches that relied on limited lexical resources or rigid feature representations. In contrast, contemporary models, especially those based on LLMs, are better equipped to disambiguate such entities due to their richer contextual understanding and broader world knowledge. This capability underscores the increasing robustness and versatility of modern NER systems.

## 2.6 Entity Linking

EL represents another fundamental sub-task within the domain of NLP. Its primary objective is to associate mentions of entities occurring in a text corpus with their corresponding entries in a structured KB, such as Wikidata or DBpedia. By grounding textual mentions in a canonical representation, EL enables the integration of unstructured text with structured knowledge, thereby supporting applications such as semantic search, knowledge graph construction, and question answering.

The EL task is closely related to NER [69]. In most practical pipelines, NER is the initial step, responsible for detecting entity mentions within the text, while Entity Disambiguation (ED) - a core component of EL - subsequently resolves these mentions by linking them to the correct entities in one or more KBs. This decomposition highlights the dual challenge of EL: identifying candidate entities and selecting the most appropriate match given the context.

Over time, a wide range of methodological approaches have been proposed to address the EL problem. Early techniques relied on relatively simple strategies, such as exact string matching or approximate matching based on edit distance and lexical similarity [10]. While computationally efficient, these approaches are highly sensitive to variations in entity names and lack robustness in ambiguous contexts. More advanced statistical and machine learning methods introduced probabilistic models and feature-based ranking mechanisms to better capture contextual signals and prior probabilities [10, 123]. These approaches improved performance but often required extensive feature engineering and domain-specific tuning.

Recent advances have been driven by deep learning techniques, which leverage neural representations to model both entity mentions and KB entries in a shared semantic space [122, 141, 71, 6]. Transformer-based architectures and pre-trained language models have further enhanced EL systems by enabling richer contextual understanding and improved generalization. Such models are capable of capturing subtle semantic relationships and resolving ambiguities even in challenging or low-resource scenarios.

To illustrate the EL task, consider the same example sentence used in the NER discussion:

> "AHAHAH RT @politico The Roots use entrance music to bash Michele Bachmann on Jimmy Fallon"

From this sentence, the extracted entity mentions are: *politico*, *The Roots*, *Michele Bachmann*, and *Jimmy Fallon*. An EL system operating over Wikidata is expected to map these mentions

27:1133617857

to their corresponding unique identifiers, namely: Q3109740 for *politico*, Q1052139 for *The Roots*, Q160582 for *Michele Bachmann*, and Q335680 for *Jimmy Fallon*.

However, this task is non-trivial due to the inherent ambiguity of natural language. For instance, the mention *politico* could be incorrectly linked to Q991826 (Politico Europe) or even unrelated entities such as Q48781731 (album by Mexican Institute of Sound), depending on the system's ability to interpret contextual cues. While such distinctions are typically straightforward for human readers, computational systems must explicitly model contextual relevance, entity popularity, and semantic coherence to resolve these ambiguities effectively. This challenge underscores the importance of robust disambiguation mechanisms in modern EL systems.

## 2.7 Relation Extraction

RE is another fundamental sub-task within the field of NLP. Its primary objective is to identify and classify semantic relationships between entities mentioned in text. By determining how entities are related, RE plays a crucial role in enabling structured understanding of unstructured data, and serves as a key component in applications such as knowledge graph construction, information retrieval, and question answering systems [30].

Similar to NER and EL, the evolution of RE methodologies has followed a progression from rule-based systems to data-driven approaches. Early work in RE relied on manually crafted patterns and linguistic rules, often leveraging syntactic parsing and domain-specific heuristics [134, 133]. While effective in narrowly defined settings, such approaches suffered from limited scalability and required significant expert effort to maintain and extend [1, 144, 91, 57].

Subsequently, machine learning-based methods introduced supervised models that utilized engineered features derived from lexical, syntactic, and semantic properties of text [57]. These approaches improved generalization capabilities and reduced reliance on explicit rules, but still depended heavily on feature design and annotated training data. In recent years, deep learning techniques have become the dominant paradigm in RE research [57, 121]. Neural architectures, particularly those based on transformer models, enable the automatic learning of contextual representations and have significantly advanced the state of the art in relation classification and extraction tasks.

Despite these advancements, RE remains inherently more challenging than NER and EL. One of the primary sources of difficulty lies in the potentially long-range and complex nature of relationships between entities. Unlike NER, where entity recognition is typically localized, or EL, where disambiguation is largely context-dependent within a limited scope, relations in RE may span across sentences, paragraphs, or even entire documents. In some cases, relevant information required to infer a relationship may be distributed across multiple documents, necessitating cross-document reasoning.

This extended scope introduces additional challenges, including the need for effective modeling of discourse structure, coreference resolution, and long-range dependencies. Consequently, modern RE systems must go beyond sentence-level analysis and incorporate broader contextual understanding to accurately capture relationships [121]. These complexities highlight the importance of continued research in developing robust and scalable methods for RE in increasingly diverse and large-scale textual corpora.

28:9831112930

# Chapter 3

# Literature review

## 3.1 Introduction

This chapter presents a comprehensive review of the literature relevant to the research undertaken in this dissertation. It begins with Section 3.2, which outlines current knowledge on HDL, with particular emphasis on its role in inflammatory processes. Section 3.3 then surveys state-of-the-art approaches to NER, highlighting recent methodological advances. Subsequently, Section 3.4 reviews the literature on end-to-end entity linking systems. Next, Section 3.5 presents current approaches to RE. Finally, Section 3.6 discusses the use of Petri nets for modeling biomedical phenomena and reviews approaches for the automatic construction of Petri net-based models.

## 3.2 High-Density Lipoproteins and Inflammation

### 3.2.1 Introduction

HDLs have historically been regarded as cardioprotective particles, primarily based on epidemiological evidence demonstrating an inverse relationship between HDL cholesterol (HDL-C) levels and cardiovascular disease (CVD) risk. However, emerging evidence has challenged this paradigm, showing that increasing HDL-C levels does not necessarily translate into improved clinical outcomes [62, 63, 95]. Consequently, contemporary research has shifted focus from HDL quantity to HDL quality and functionality, particularly in the context of inflammation and immune modulation.

### 3.2.2 HDL Structure, Function, and Heterogeneity

HDL particles are highly heterogeneous in size, composition, and function, containing a diverse array of lipids, proteins, and bioactive molecules such as apolipoproteins, enzymes, and microRNAs [95, 62, 34]. The major structural protein, apolipoprotein A-I (apoA-I), plays a central role in HDL biogenesis and function, particularly in Reverse Cholesterol Transport (RCT), whereby cholesterol is removed from peripheral tissues and delivered to the liver for excretion [62, 34, 95]. Beyond RCT, HDL exerts multiple biological effects, including anti-inflammatory, antioxidant, anti-thrombotic, and endothelial protective functions [95, 63, 34, 23, 62]. These properties are highly dependent on HDL composition and structural integrity, which can be altered under pathological conditions.

### 3.2.3 Paradigm Shift: From HDL Quantity to Functionality

The lack of cardiovascular benefit observed with HDL-C–raising therapies, including cholesteryl ester transfer protein (CETP) inhibitors and niacin, has shifted the focus from increasing HDL-C



29:3325048422

concentrations to improving HDL functionality [62, 63, 34]. Mendelian randomization studies further support this concept by demonstrating that genetic variants associated with elevated HDL-C do not give protection against myocardial infarction [62]. HDL functionality encompasses several key processes, including cholesterol efflux capacity, antioxidant activity, and modulation of inflammatory pathways. Among these, cholesterol efflux capacity has emerged as a stronger predictor of cardiovascular risk than HDL-C levels alone [34]. Importantly, HDL function is influenced by its proteomic and lipidomic composition, which can be significantly altered during inflammatory states.

### 3.2.4 HDL as a Modulator of Inflammation

HDL plays a crucial role in regulating immune responses and inflammation. It modulates immune cell activation, cytokine production, and oxidative stress, thereby contributing to the maintenance of vascular homeostasis [95]. HDL can inhibit the expression of adhesion molecules on endothelial cells, reduce monocyte recruitment, and neutralize pro-inflammatory lipids, all of which are critical steps in the development of atherosclerosis. Additionally, HDL-associated enzymes such as paraoxonase-1 (PON1) contribute to its antioxidant and anti-inflammatory properties by preventing the oxidation of LDL, a key driver of vascular inflammation [23]. Native HDL can also remove oxidized lipids and reactive oxygen species, thereby attenuating inflammatory signaling pathways.

### 3.2.5 Dysfunctional HDL in Inflammatory States

Under conditions of acute or chronic inflammation, HDL undergoes structural and functional modifications that impair its protective properties. These modifications include oxidation, glycation, and carbamylation of apoA-I and other HDL components [95]. Such post-translational changes lead to the formation of so-called “dysfunctional HDL”, which may lose its anti-inflammatory capacity and even acquire pro-inflammatory characteristics. Inflammatory conditions such as diabetes, chronic kidney disease, autoimmune disorders, and infections are associated with alterations in HDL composition and function [95]. For example, during bacterial infections, serum amyloid A (SAA) is increased, while apoA-I is decreased in HDL particles, impairing cholesterol efflux and promoting inflammatory responses. Furthermore, HDL quantity itself is often reduced in inflammatory states, including acute infections and chronic diseases, reflecting both increased catabolism and impaired synthesis [23]. However, functional impairment appears to be more critical than reduced levels alone.

### 3.2.6 Bidirectional Relationship Between HDL and Inflammation

The relationship between HDL and inflammation is bidirectional. Inflammation can alter HDL composition and impair its function, while dysfunctional HDL can worsen inflammatory processes. HDL influences immune cell behavior, including macrophage polarization and T-cell activation, thereby modulating both innate and adaptive immune responses [95]. Recent concepts also propose that HDL acts as a bidirectional lipid vector, enabling both lipid efflux and influx depending on tissue requirements [85]. This dynamic role may be particularly relevant during tissue injury and repair, where HDL participates in resolving inflammation and restoring cellular homeostasis.

### 3.2.7 Clinical Implications

The recognition that HDL function is more important than HDL-C levels has significant clinical implications. Traditional lipid profiles may not adequately capture cardiovascular risk associated

30:5544091400

with HDL, and functional assays such as cholesterol efflux capacity are being explored as more informative biomarkers [34]. Therapeutic strategies are increasingly focused on improving HDL functionality rather than simply raising HDL-C levels. These approaches include the use of reconstituted HDL (rHDL), apoA-I mimetics, and modulation of HDL-associated bioactive molecules such as sphingosine-1-phosphate (S1P) [95, 34, 63, 62]. However, clinical translation remains challenging due to the complexity and heterogeneity of HDL particles.

### 3.2.8 Synthesis

In summary, HDL plays a multifaceted role in inflammation, extending far beyond its traditional function in lipid transport. The emerging understanding of HDL as a dynamic and functionally diverse particle has shifted the focus from quantity to quality and functionality. Inflammatory conditions alter HDL structure and impair its protective functions, contributing to disease progression.

## 3.3 Named Entity Recognition

NER has undergone a fundamental architectural shift. Traditionally framed as a sequence labeling task, the field has moved toward open-vocabulary extraction and generative distillation, driven by the emergence of LLMs.

### 3.3.1 Traditional Discriminative Approaches

For much of the last decade, the state-of-the-art in NER was defined by encoder-only architectures, most notably `BERT` [31] and its derivatives (e.g., `RoBERTa` [83], `DeBERTa` [50]). These models typically predict labels based on the *BIO* (Beginning, Inside, Outside) tagging scheme.

While highly efficient, these ”closed-world” models suffer from primary limitation - Label Sparsity. They are restricted to a predefined set of entity types (e.g., PER, ORG, LOC) established during training [5, 22].

### 3.3.2 Zero-Shot and Open-Vocabulary NER

To address the rigidity of traditional taggers, recent advancements have introduced models capable of generalizing to arbitrary entity types. `GLiNER` [145] represents a significant breakthrough in this domain. By framing NER as a span-matching task rather than sequence labeling, `GLiNER` achieves strong generalization across diverse benchmarks with only 300 million parameters.

The `GLiNER` architecture integrates three core components: (1) a pretrained textual encoder for context, (2) a span representation module that aggregates token-level embeddings, and (3) an entity representation module that encodes the labels themselves as natural language queries. By computing a matching score between the latent representations of text spans and entity types, `GLiNER` facilitates ”zero-shot” extraction, where a user can define new entity types at inference time without retraining.

### 3.3.3 LLM-Based Generative Extraction and Distillation

The rise of LLMs like `GPT` and `Claude` has introduced the “Generative NER” paradigm. In this context, NER is reformulated as a sequence-to-sequence task where the model outputs entities

31:4853363626

in structured formats such as JSON or inline XML tags. Research indicates that while general-purpose LLMs excel at complex reasoning and ED [137], their high computational cost and latency often preclude them from large-scale production use.

Consequently, frameworks like `UniversalNER` (`UniNER`) [147] focus on task-specific distillation. Built upon the `LLaMA` LLM architecture [130], `UniNER` was fine-tuned using high-quality synthetic data generated by `ChatGPT` on the Pile corpus [40]. `UniNER` utilizes a conversation-style tuning approach, prompting the model with natural language questions for each potential entity type.

A technical contribution of `UniNER` is its negative sampling strategy. During training, the model is presented with entity types not present in the text and is taught to return empty lists. This approach significantly mitigates hallucination tendencies and improves the model's instruction-following capabilities. By distilling the broad knowledge of a large teacher model into a more compact student model, `UniNER` achieves superior performance on specialized benchmarks while maintaining open-domain flexibility.

### 3.3.4 Synthesis

The current SOTA reflects a convergence where the speed of encoder-based models (like `GLiNER`) and the linguistic intelligence of LLMs (like `UniNER`) provide a robust toolkit for modern Information Extraction. These models shift the bottleneck from data annotation to prompt engineering and ontology design, allowing for the rapid adaptation of NER systems to specialized domains.

## 3.4 Entity Linking Landscape

The contemporary landscape of EL is dominated by multi-stage pipelines that balance computational efficiency with semantic depth. Generally, state-of-the-art approaches follow a three-step paradigm: (1) generating dense or symbolic entity representations, (2) performing candidate retrieval to identify a subset of $k$ potential matches from a KB, and (3) executing a final disambiguation or re-ranking step to select the most probable entity.

### 3.4.1 Bi-Encoder and Cross-Encoder Architectures

A seminal contribution to scalable EL is the `BLINK` framework [141]. Designed specifically for ED, `BLINK` utilizes a staged pipeline to manage the trade-off between speed and accuracy. The initial stage employs a bi-encoder architecture, mapping both the mention context and KB entities into a shared dense vector space. This allows for rapid retrieval via $k$-nearest neighbor ($k$-NN) search.

To refine these results, `BLINK` introduces a cross-encoder in the final stage. Unlike the bi-encoder, which processes inputs independently, the cross-encoder jointly encodes the mention and candidate entities. This allows the model to capture fine-grained interactions between the context and the entity description, yielding a probability distribution over the $k$ candidates. While highly effective, `BLINK` is a disambiguation-only system, assuming that entity boundaries have been pre-identified.

Similarly, the `EPGEL` framework [71] enhances the retrieval phase by leveraging the generative capabilities of `BART` models. `EPGEL` constructs comprehensive entity profiles, combining titles and descriptions, which are indexed within an Elasticsearch engine. By augmenting the retrieval set with aliases and entity types, `EPGEL` achieves superior performance over `BLINK` on several benchmarks. However, the framework's reliance on extensive pre-training suggests potential challenges in domain adaptation or when applied to lower-resource KBs.

32:6824103157

### 3.4.2 End-to-End Frameworks

In contrast to disambiguation-only systems, recent research has shifted toward holistic, end-to-end architectures. `ReFinED` [6] integrates entity recognition and linking into a singular transformer-based workflow. The process begins with token-level classification to detect entity mentions, which are subsequently encoded into fixed-length embeddings.

`ReFinED` is distinguished by its scoring mechanism, which decomposes the linking decision into two components:

1. **Typing Score:** Evaluates the semantic consistency between the mention context and the entity's taxonomic category (e.g., Person vs. Organization).
2. **Description Score:** Measures the alignment between the context and the KB entity's textual description.

A significant advantage of `ReFinED` is its native support for `NIL` entities - identifying mentions that do not exist within the reference KB. This makes the framework particularly robust for real-world applications where KB coverage may be incomplete.

### 3.4.3 Synthesis and Limitations

While `BLINK`, `EPGEL`, and `ReFinED` represent the current state-of-the-art, they remain heavily coupled with large-scale KBs such as Wikipedia and Wikidata. Their performance is often a function of the rich textual descriptions and structured metadata available in these sources. Consequently, extending these models to specialized or proprietary KBs typically necessitates significant fine-tuning and architectural adaptation to account for variations in data density and schema structure.

## 3.5 Relation Extraction in Biomedical Text Mining

RE is a central task in Biomedical Natural Language Processing (BioNLP), aiming to identify semantic relationships between entities such as genes, proteins, diseases, and drugs from unstructured text. This task is critical for constructing knowledge graphs, supporting pathway modeling, and enabling downstream applications such as drug discovery and clinical decision support. Recent surveys highlight that biomedical RE has evolved alongside the availability of large-scale textual resources and annotated datasets, including scientific literature, electronic health records, and curated corpora [57].

### 3.5.1 Evolution of Relation Extraction Methods

Biomedical RE methods have undergone several paradigm shifts, transitioning from rule-based approaches to machine learning and, more recently, to deep learning and transformer-based architectures.

**Rule-based methods** Early RE systems relied on manually crafted rules and linguistic patterns to identify relations. These approaches typically leveraged syntactic structures such as dependency parses, part-of-speech tags, and predefined trigger words. Systems like `REACH` [134, 133] illustrate this paradigm, focusing on extracting biochemical interactions and causal relationships through curated rules and domain knowledge. While rule-based systems offer interpretability and precision, they are limited by their inability to generalize across domains and their dependence on expert-defined patterns [1, 144, 91, 57].

33:1973770529

**Traditional machine learning methods** With the availability of annotated corpora, RE was increasingly formulated as a supervised classification problem. Methods based on support vector machines and kernel techniques utilized engineered features derived from lexical, syntactic, and semantic representations. Although these approaches improved scalability and adaptability compared to rule-based systems, their performance remained constrained by the quality of feature engineering and the availability of labeled data [57].

**Deep learning methods** The introduction of deep neural networks marked a significant advancement in RE. RNNs and LSTM models, enabled automatic feature learning from raw text. These models demonstrated strong performance by capturing contextual semantics and long-range dependencies, reducing the reliance on manual feature design. Hybrid architectures combining multiple neural components further improved performance across various biomedical RE benchmarks [57].

**Transformer-based methods** Transformer architectures, particularly `BERT` and its biomedical variants such as `BioBERT` [73], have become the dominant paradigm in RE. These models leverage attention mechanism to capture bidirectional context and long-range dependencies effectively. Fine-tuned transformer models have achieved state-of-the-art results across multiple biomedical datasets, significantly outperforming earlier approaches. However, they are constrained by input length limitations and require substantial computational resources for training and adaptation [57, 121].

### 3.5.2 Document-level Relation Extraction

While early work focused primarily on sentence-level RE, recent research emphasizes document-level extraction, where relations may span multiple sentences. This setting is particularly important in biomedical texts, where entities often appear in different contexts and may be referenced through aliases or abbreviations. Studies indicate that a substantial proportion of biomedical relations require cross-sentence reasoning, highlighting the limitations of sentence-level approaches [121].

Document-level RE introduces additional challenges, including the need for cross-sentence inference, handling co-reference and entity normalization, and integrating external knowledge. Graph-based methods address these challenges by constructing document-level representations and performing relational reasoning over entity and sentence graphs. Transformer-based models extend their capabilities through attention mechanism that capture long-distance dependencies. However, both approaches face limitations, such as reliance on graph quality or input length constraints [121].

### 3.5.3 Large Language Models for Relation Extraction

Recent advances in LLMs have opened new possibilities for RE. Unlike earlier approaches, LLMs can interpret complex linguistic structures and capture implicit relationships without explicit feature engineering. They enable flexible, context-aware extraction and can be adapted through prompting, fine-tuning, and retrieval-augmented generation.

Emerging frameworks integrate LLMs with task-specific adaptations, such as synthetic data generation, document-relation cross-mapping, and retrieval of external biomedical knowledge. These approaches aim to address key challenges in document-level RE, including data scarcity, cross-sentence reasoning, and domain-specific knowledge integration [121].

34:1492611877

### 3.5.4 Synthesis and Limitations

The evolution of biomedical RE reflects a broader trend in NLP, moving from rule-based systems toward data-driven and pre-trained models. While systems such as `REACH` remain valuable for structured and interpretable extraction, modern approaches increasingly rely on deep learning and LLMs to achieve higher coverage and flexibility.

However, direct application of LLMs to biomedical RE remains challenging. Limitations include insufficient domain knowledge, difficulty in handling complex multi-class relation settings, and sensitivity to incomplete or noisy input data. As a result, recent research focuses on hybrid approaches that combine LLMs with structured knowledge, fine-tuning strategies, and domain-specific augmentation techniques [121].

## 3.6 Petri net-based models automatic creation

This section first reviews the application of PN-based to biomedical process modeling, highlighting their suitability for representing biological knowledge. It then surveys computational approaches for the automatic construction of Petri net-based models, focusing on methods that translate structured biological knowledge, interaction databases, and information extracted from scientific literature into executable Petri net representations.

### 3.6.1 Petri Nets in Biomedical Process Modeling

Early work demonstrated the usefulness of Petri nets for representing metabolic pathways and biochemical reaction systems [110, 52, 99], while subsequent surveys emphasized their applicability to molecular networks, signaling pathways, metabolic systems, and genetic regulatory networks [20, 51]. More recent studies further confirm the continued relevance of Petri-net-based modeling in biomedical contexts, including signal transduction networks [67], immune-response modeling [118], and multilevel biological systems modeled with colored Petri nets [78]. These applications show that Petri nets are well suited for encoding the structure and dynamics of biomedical processes, supporting qualitative, semi-quantitative, stochastic, and hybrid analyses.

### 3.6.2 Knowledge transformation

A group of methods focuses on the transformation of already curated biological pathway data into Petri net formalisms. For instance, VANESA [9] provides an integrated environment for the visualization, construction, and analysis of PN models, combining simulation capabilities with access to pathway information from multiple biological databases. This enables users to construct Petri nets grounded in experimentally validated and curated knowledge. Similarly, GINtoSPN [77] automates the conversion of molecular interaction networks from the Global Integrative Network (GIN), a comprehensive knowledge graph integrating molecular interactions from multiple biological databases and literature-derived sources [19], into stochastic Petri nets, allowing for the generation of executable models from large-scale interaction datasets.

More broadly, this category includes approaches that rely on established pathway repositories such as KEGG [65], Reactome [108], or BioModels [124] as primary knowledge sources. In such workflows, biological entities and their interactions are mapped onto PN elements, such as places, transitions, and arcs, often following predefined transformation rules [9, 77]. This mapping process can be partially automated and may incorporate domain-specific heuristics to preserve biochemical semantics, such as distinguishing between different types of regulatory or catalytic

35:2787744601

interactions [9, 77]. In some cases, additional layers of annotation, including kinetic parameters or compartmentalization, are incorporated to enhance the expressiveness of the resulting PN models [13].

The main advantage of these approaches lies in their reliability and interpretability. Since they are grounded in curated databases, the resulting models benefit from a high level of biological accuracy and consistency [90, 65]. Furthermore, the use of standardized data formats and ontologies facilitates reproducibility and interoperability with other modeling frameworks [13].

However, these methods are inherently constrained by the scope and update frequency of the underlying databases. Curated resources often lag behind the rapidly expanding body of biomedical literature, leading to incomplete or outdated representations of biological processes [90, 65]. Additionally, they may exhibit biases toward well-studied pathways and organisms, limiting their applicability in emerging or less-characterized domains [90]. As a result, while database-driven PN construction provides a solid foundation for model development, it does not fully address the need for incorporating up-to-date and comprehensive knowledge directly from unstructured scientific sources.

### 3.6.3 Knowledge synthesis

Beyond direct conversion tools, other frameworks aim to support the integration of diverse knowledge sources into computational modeling workflows. The MAGCC framework [25] represents one such effort, combining curated biological models with information derived from textual sources. Within MAGCC, textual knowledge is incorporated into a Structured Scientific Knowledge Representation, which can subsequently be utilized by automated systems such as the Computational Modeling Assistant to generate executable models. While this approach demonstrates the feasibility of integrating textual information into modeling pipelines, its primary emphasis lies in knowledge integration and model assembly rather than in the systematic construction of PN-based modelss from raw literature.

A more general and influential line of work is represented by the Integrated Network and Dynamical Reasoning Assembler (INDRA) [46, 7]. INDRA is designed as a unifying framework for assembling computational models from heterogeneous sources, including pathway databases, experimental data, and outputs of machine reading systems. Central to INDRA is the concept of standardized mechanistic statements, which capture biological interactions such as activation, inhibition, binding, and modification in a structured and semantically grounded form. These statements are normalized across sources and can be used to generate a variety of downstream models, including rule-based models, ordinary differential equation systems, and network representations.

An important contribution of INDRA lies in its ability to reconcile and aggregate information from multiple extraction engines and databases, thereby improving coverage and robustness. The framework also incorporates mechanisms for evidence tracking, provenance management, and belief scoring, which are critical for assessing the reliability of automatically assembled models. Although INDRA is not specifically designed for Petri net construction, its intermediate representation of mechanistic knowledge provides a valuable foundation that can, in principle, be mapped to PN formalisms. This highlights a broader trend in systems biology toward the use of intermediate, semantically rich representations as a bridge between data sources and executable models.

### 3.6.4 Synthesis and Limitations

Despite these advances, existing approaches to PN model construction remain largely reliant on structured inputs or intermediate representations derived from curated or semi-structured data.

36:5384098225

Direct pipelines that translate unstructured scientific literature into PN-based models are still relatively underdeveloped. Most current systems either (i) convert between structured representations (e.g., pathway databases to Petri nets) or (ii) extract biological knowledge into general-purpose interaction formats that require additional processing before they can be used for PN modeling.

37:2732956905

# Chapter 4

# HDL in inflammation Petri net-based model

## 4.1 Introduction

HDL has traditionally been recognized for its essential role in maintaining lipid homeostasis, primarily through facilitating reverse cholesterol transport. Beyond this foundational function, it plays an active part in cardiovascular defense by exerting anti-inflammatory, anti-diabetic, and anti-thrombotic effects. However, contemporary research challenges the classic, quantity-centric view of HDL, demonstrating that simply increasing its circulating levels does not guarantee clinical benefits [62, 63, 34]. Instead, attention has shifted to HDL functionality, which is highly sensitive to the molecule's structural composition and the surrounding physiological context [95].

In particular, inflammatory conditions profoundly disrupt HDL metabolism. Chronic inflammation triggers a functional shift, transforming HDL from a protective, anti-inflammatory agent into a dysfunctional, pro-inflammatory molecule. This creates a pathological feedback loop that worsens inflammatory signaling and alters lipoprotein remodeling [58, 26]. Because these interactions span multiple organ systems - including the liver, intestines, peripheral tissues, and immune cells - traditional, reductionist approaches that isolate single molecular pathways fail to capture the complete picture. A holistic, systems-level approach is required to understand the dynamic, multiorgan network that governs HDL behavior under stress.

To address this complexity, this chapter details the manual construction, expansion, and computational analysis of a systems biology model designed to represent HDL metabolism under homeostatic, pathological and treatment conditions [39]. Using a PNs approach, the baseline metabolic framework is progressively extended to incorporate inflammatory perturbations and selected therapeutic interventions.

Through systematic in silico knock-out simulations, this research evaluates how disabling individual metabolic processes influences the global system. The analysis focuses on identifying the critical regulatory control points. Ultimately, this chapter maps out the transition from pathology-driven to therapy-dependent regulation of HDL, establishing a conceptual framework to guide the development of targeted, systems-oriented therapeutic strategies.

## 4.2 Baseline model

Before describing the Petri net representation in detail, Figure 4.1 summarizes the biological processes captured by the model at a conceptual level. The diagram illustrates the physiological



38:9202029440

HDL biogenesis and remodeling cycle together with the pathological influence of inflammation and the restorative effects of therapeutic intervention. The baseline model represents the homeostatic HDL lifecycle, whereas the inflammation and treatment modules are introduced in subsequent sections as extensions of this core network. This conceptual overview provides the biological context for the detailed Petri net formulation presented below.

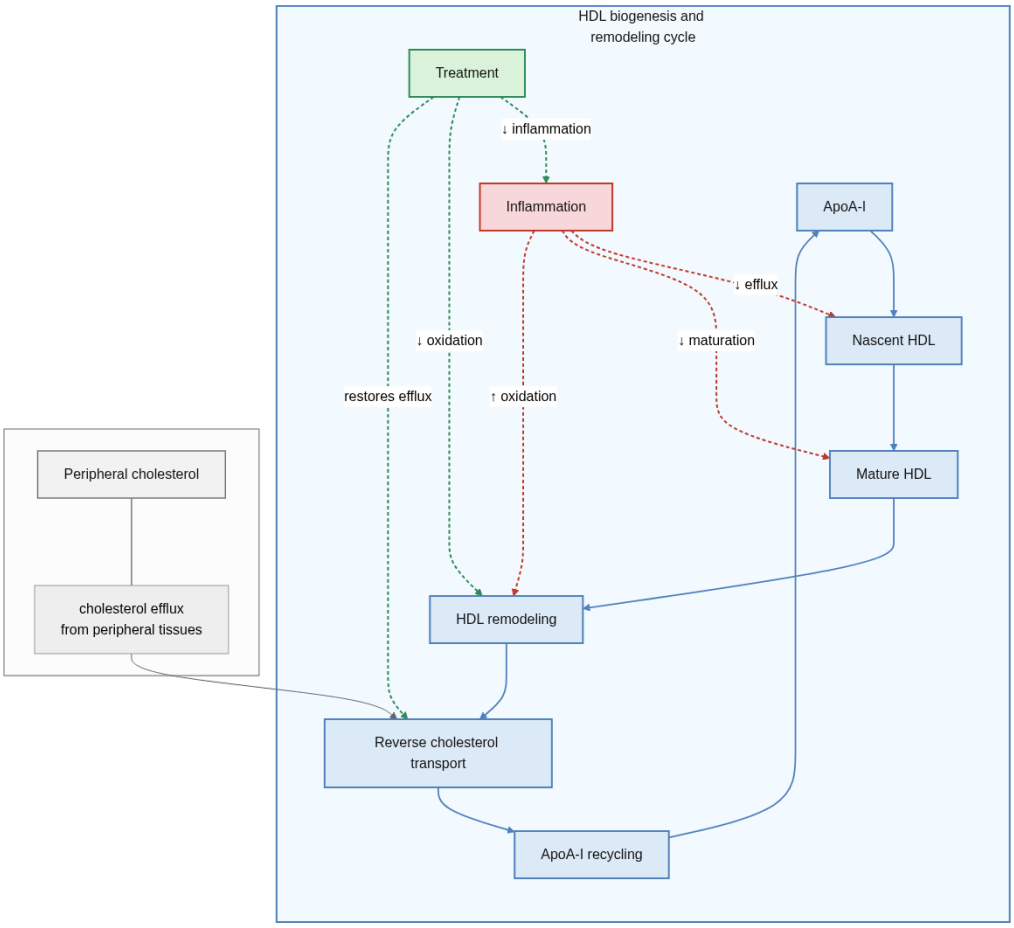



FIGURE 4.1: Conceptual overview of the HDL metabolism model. The baseline model captures HDL biogenesis, maturation, remodeling, reverse cholesterol transport, and Apolipoprotein A-I recycling under homeostatic conditions. Additional modules represent the disruptive effects of inflammation on cholesterol efflux, HDL maturation, and remodeling, as well as therapeutic interventions that partially restore physiological HDL function.

The baseline PN model simulates the homeostatic lifecycle of HDL across a multiorgan pathway. This configuration captures the physiological balance of lipid transport, apolipoprotein production, enzymatic alterations, and clearance under non-pathological conditions.

**System Initialization and Biogenesis** The lifecycle within the model begins with the generation of the primary structural protein of HDL. In the following descriptions, places are denoted by $p_i$ and transitions by $t_i$, where $i$ is the unique index assigned to each place or transition in the Petri net.

- **Apolipoprotein A-I Production:** The synthesis and cellular release of ApoA-I originate from hepatic ($p_0$) and intestinal ($p_1$) sources, which are driven by specific secretion transitions ($t_2$, $t_3$).

39:7327752738

- **Nascent Particle Formation:** Upon entering the circulatory system, ApoA-I exists in an unlipidated or lipid-poor state ($p_2$). This state acts as the initial substrate for peripheral lipid acquisition. Cellular lipid removal is mediated by ATP-binding cassette transporter A1 (ABCA1) ($p_3$). This interaction triggers the initial lipidation phase via nascent particle formation transitions ($t_4$, $t_{38}$), yielding structurally immature, discoidal HDL particles ($p_4$).

**Maturation and Structural Remodeling** Once discoidal intermediates are established, the model coordinates their expansion into mature, spherical forms through continuous enzymatic modifications and lipid transfer events.

- **Esterification and Expansion:** The transition from discoidal ($p_4$) to dense, spherical HDL variants ($p_7$, $p_8$) is regulated by lecithin-cholesterol acyltransferase (LCAT) ($p_9$). This process is governed by esterification transitions ($t_5$, $t_8$). Concurrently, circulating HDL continues to accumulate lipids from peripheral tissues via both ABCA1 ($p_3$) and ATP-binding cassette sub-family G member 1 (ABCG1) ($p_5$) pathways.
- **Lipid Exchange Dynamics:** The heterogeneity of the circulating HDL pool is maintained by active molecular exchanging mechanisms. CETP ($p_{10}$) handles the shuffling of triglycerides and cholesteryl esters between HDL and apolipoprotein B-containing particles via dedicated transfer transitions ($t_{10}$, $t_{11}$).
- **Phospholipid and Lipolytic Processing:** Phospholipid redistribution is managed by phospholipid transfer protein (PLTP) ($p_{15}$), while particle size and core lipid compositions are further refined by hepatic lipase ($p_{16}$) and endothelial lipase ($p_{17}$) through hydrolytic remodeling transitions ($t_{49}$, $t_{50}$).

**Clearance and Excretion Pathways** The final phase of the baseline network tracks the elimination and recycling of lipoprotein components.

- **Selective Lipid Delivery:** Cholesteryl esters are selectively delivered to hepatic tissue via scavenger receptor class B type 1 (SR-B1) ($p_6$), a process mapped through cholesterol export transitions ($t_6$, $t_{23}$).
- **Terminal Degradation:** The structural lifecycle concludes with terminal catabolism. This includes the breakdown of the mature lipoprotein complex and the subsequent renal excretion of liberated, lipid-poor ApoA-I ($p_{14}$), which is driven by metabolic removal transitions ($t_{20}$-$t_{22}$).

For the creation and analysis of the PN models, Holmes software was used [105]. The structural topology of the baseline PN, illustrated in Figure 4.2, comprises 25 places and 40 transitions. For comprehensive enumeration, the complete sets of places and transitions across all incremental models are detailed in Table 4.3 and Table 4.4, respectively. This unified presentation accounts for the sequential extension of the baseline architecture into the inflammation model, which is subsequently extended by the treatment mechanisms. Structural analysis reveals that the net is fully covered by 345 $t$-invariants. Furthermore, the network contains 10 MCT sets, though their analytical utility is limited due to their size distribution. Specifically, the majority of these sets consist of merely two transitions, while two sets comprise three and four transitions, respectively. The remaining 17 transitions do not belong to any non-trivial MCT set, indicating a high degree of local independence in those functional pathways.

40:9433041814

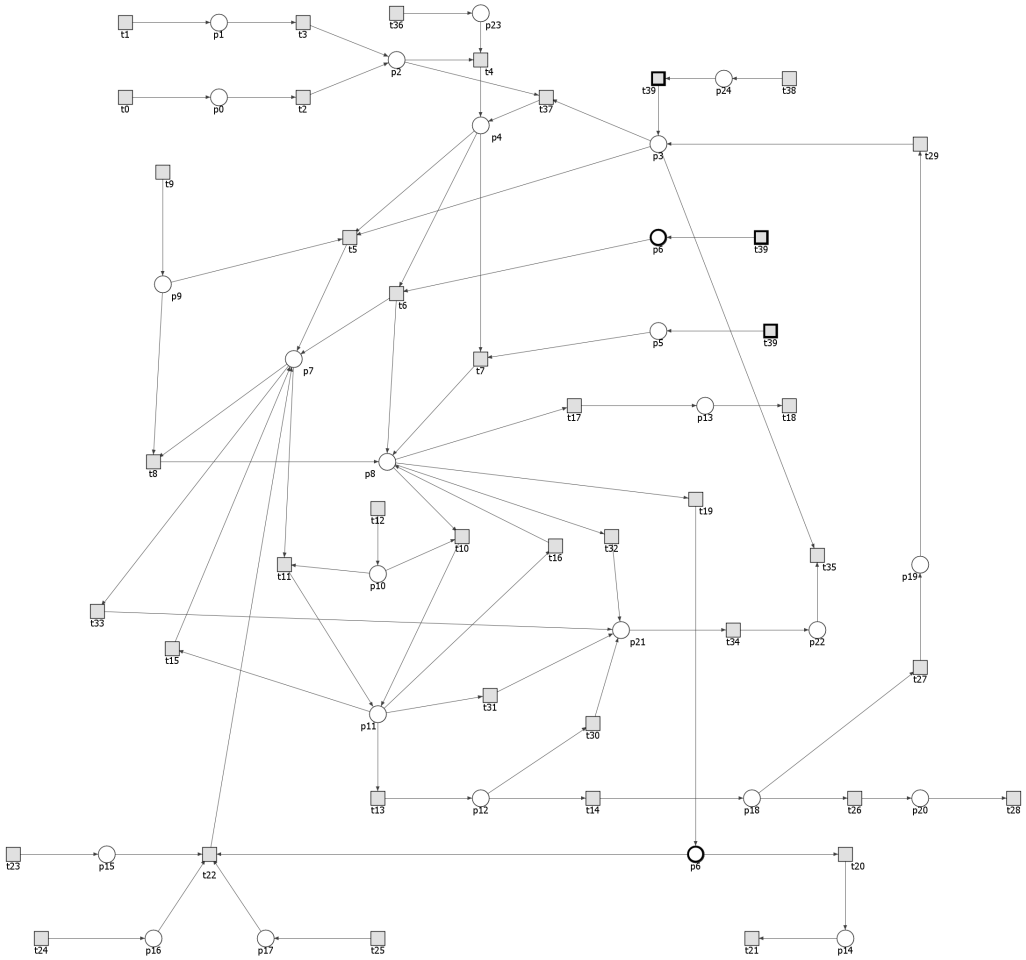



FIGURE 4.2: Petri net-based model of HDL metabolism - the baseline model. To improve the readability of the Petri net, logical places and logical transitions (shown in bold) are used as graphical references to existing model elements. They duplicate the graphical representation of the corresponding place or transition, allowing long or intersecting arcs to be omitted. These logical elements are semantically identical to their originals and do not introduce additional places, transitions, or firing behavior.

## 4.3 Model extension - inflammation

To evaluate how network-wide stress destabilizes lipid homeostasis, the network was expanded by integrating inflammatory perturbations. This state was achieved by introducing dedicated pathological network components and adjusting arc weights to reflect the molecular down-regulation, acceleration, or impairment of standard baseline processes.

**Molecular Alterations and Transporter Suppression** Inflammation transforms the network from an atheroprotective system into a dysfunctional loop. Pro-inflammatory signaling cascades, driven by cytokines such as Tumor Necrosis Factor alpha (TNF-$\alpha$) and Interferon gamma (IFN-$\gamma$), alter the structural attributes and expression profiles of crucial metabolic nodes.

- **Transporter Down-regulation:** The functional capacity of key cellular efflux transporters (ABCA1, ABCG1) and selective clearance receptors (SR-B1) is reduced. This shift impairs peripheral cholesterol clearance.
- **Enzymatic Decoupling:** LCAT-driven esterification is suppressed, stalling the conversion

41:2053996593

of discoidal particles into mature spheres.

**Pathological Transitions and Feedback Loops** The inflammatory extension incorporates specific regulatory bottlenecks that dictate network behavior under chronic stress:

- **Adiponectin Deprivation:** Triggers a drop in protective signaling via a decreased production ($t_{44}$).
- **Esterification Failure:** Captured through an LCAT-impairment transition ($t_{47}$).
- **Oxidative Modification:** Circulating HDL undergoes chemical alteration, simulated by an oxidized HDL activity transition ($t_{56}$), which yields structurally compromised particles with diminished transport capabilities.
- **Cytokine Domination:** Cytokine-mediated expression transitions ($t_{52}$, $t_{54}$) and an inflammatory suppression transition ($t_{39}$) shift the system's focus away from lipid clearance. This creates a self-sustaining feedback loop where dysfunctional HDL amplifies the underlying inflammatory state.

A complete summary of these inflammation-induced network adaptations and their literature validations is compiled in Table 4.1.

Table 4.1: Modifications introduced to the HDL metabolism model to represent the inflammatory state.

| **Place/transition name; modification** | **Literature support** |
|---|---|
| inflammatory function suppression; increase in weights of arcs:<br>• ABCA1 ($p_3$) $\rightarrow$ discoidal HDL formation ($t_{38}$) (new weight: 3)<br>• ABCA1 ($p_3$) $\rightarrow$ enzymatic reactions ($t_5$) (new weight: 3) | Adipose tissue inflammation impairs the expression and activity of cholesterol transporters such as ABCA1, reducing cholesterol efflux from adipocytes. Since ABCA1 is a key mediator of both cholesterol transfer and the generation of nascent HDL particles, its dysfunction promotes the production of immature HDL with diminished capacity for reverse cholesterol transport [58]. |
| weight 3 on cholesterol export ($t_{23}$) to ABCG1 ($p_5$) arc | Loss of adipocyte ABCA1 has been associated with a compensatory increase in ABCG1 expression, indicating that cells enhance alternative cholesterol export pathways to preserve intracellular cholesterol homeostasi [26]. |
| Added new place - TG-enriched HDL-C clearance ($p_{26}$) | Elevated CETP activity promotes the exchange of triglycerides for cholesteryl esters between lipoproteins, generating HDL particles enriched in triglycerides. These modified HDL particles exhibit reduced stability and are cleared more rapidly from the circulation, contributing to lower circulating HDL-C concentrations [58]. |

42:7791691505

| **Place/transition name; modification** | **Literature support** |
|---|---|
| adiponectin decreased production; increase in weights of arcs:<br>• ABCA1 ($p_3$) $\rightarrow$ discoidal HDL formation ($t_{38}$) (new weight: 5)<br>• ABCA1 ($p_3$) $\rightarrow$ enzymatic reactions ($t_5$) (new weight: 5)<br>• ABCA1 ($p_3$) $\rightarrow$ absorbed FA in peripheral tissue ($t_{36}$) (new weight: 3)<br>• lipid free ApoA-I ($p_2$) $\rightarrow$ discoidal HDL formation ($t_4$) (new weight: 3)<br>• lipid free ApoA-I ($p_2$) $\rightarrow$ discoidal HDL formation ($t_{38}$) (new weight: 3) | Reduced adiponectin levels have been associated with impaired HDL metabolism and disrupted cholesterol homeostasis. Experimental studies in animal models have shown that adiponectin deficiency lowers the expression of both ApoA-I and ABCA1, leading to reduced ApoA-I availability and diminished HDL-C formation, although comparable evidence has not been consistently demonstrated in humans [58]. |
| reduced esterification; increase in weights of arcs:<br>• LCAT ($p_9$) $\rightarrow$ enzymatic reactions ($t_5$) (new weight: 3)<br>• LCAT ($p_9$) $\rightarrow$ enzymatic reactions ($t_8$) (new weight: 3) | The association of SAA with HDL particles during chronic inflammation disrupts normal HDL remodeling by reducing the efficiency of LCAT-mediated cholesterol esterification. As a consequence, the conversion of nascent HDL into mature HDL particles is impaired, leading to defective HDL maturation and reduced reverse cholesterol transport [58]. |
| Added new place - HDL hydrolysis increased by sPLA2 ($p_{29}$) | Inflammatory conditions promote the release of sPLA2, which enhances the hydrolysis of phospholipids within HDL particles. This accelerates HDL particle degradation and contributes to a reduction in circulating HDL-C levels [58]. |

43:2275074488

| **Place/transition name; modification** | **Literature support** |
|---|---|
| decreased expression; increase in weights of arcs:<br>• ABCA1 ($p_3$) → discoidal HDL formation ($t_{38}$) (new weight: 7)<br>• ABCA1 ($p_3$) → enzymatic reactions ($t_5$) (new weight: 7)<br>• ABCA1 ($p_3$) → absorbed FA in peripheral tissue ($t_{36}$) (new weight: 5)<br>• SR-B1 ($p_6$) → cholesterol export ($t_6$) (new weight: 3)<br>• ABCG1 ($p_5$) → cholesterol export ($t_7$) (new weight: 3) | Low-grade systemic inflammation impairs reverse cholesterol transport by suppressing the expression of key cholesterol transporters. Inflammatory signaling reduces retinoic acid receptor-$\alpha$ activity, leading to decreased expression of ABCA1, ABCG1, and SR-B1 in macrophages, thereby limiting cholesterol efflux and compromising HDL function [58]. |
| decreased expression by INF gamma; increase in weights of arc:<br>• SR-B1 ($p_6$) → cholesterol export ($t_6$) (new weight: 5) | IFN-$\gamma$ has been shown to impair macrophage cholesterol handling by reducing the expression of SR-B1 while promoting macrophage activation. The resulting decrease in SR-B1 expression limits cholesterol efflux and contributes to impaired reverse cholesterol transport under inflammatory conditions [58]. |
| activity decrease; increase in weights of arcs:<br>• ABCA1 ($p_3$) → discoidal HDL formation ($t_{38}$) (new weight: 9)<br>• ABCA1 ($p_3$) → enzymatic reactions ($t_5$) (new weight: 9)<br>• ABCA1 ($p_3$) → absorbed FA in peripheral tissue ($t_{36}$) (new weight: 7)<br>• ABCG1 ($p_5$) → cholesterol export ($t_7$) (new weight: 5) | Oxidized HDL particles amplify inflammatory signaling through activation of the NOD-like receptor protein 3 (NLRP3) inflammasome, promoting the release of pro-inflammatory mediators such as IL-1$\beta$ and IL-18. This inflammatory response is accompanied by reduced activity of the cholesterol transporters ABCA1 and ABCG1, further impairing cholesterol efflux and HDL functionality [58]. |

The inflammation model, presented in Figure 4.3, consists of 36 places and 59 transitions. Similarly to the baseline model, lists of places and transitions are available in Table 4.3 and Table 4.4 respectively. The net is fully covered by 1735 $t$-invariants. Also similarly to the baseline model, MCT sets offer minimal data, as 10 out of 14 consist only of two transitions.

The structural architecture of the inflammation model, illustrated in Figure X, expands the

44:5261304950

baseline topology to 36 places and 59 transitions. In alignment with the documentation strategy of the baseline system, the exhaustive inventories of these components are cataloged within Table 4.3 and Table 4.4, respectively. Structural invariant analysis demonstrates that the underlying Petri net is fully covered by 1735 $t$-invariants, reflecting a significant increase in behavioral complexity. However, much like the baseline net, the analytical insights derived from the MCT sets remain marginal. Out of the 14 identified MCT sets, 10 are trivial, each containing only two transitions, thereby offering limited explanatory power regarding large-scale functional sub-systems.

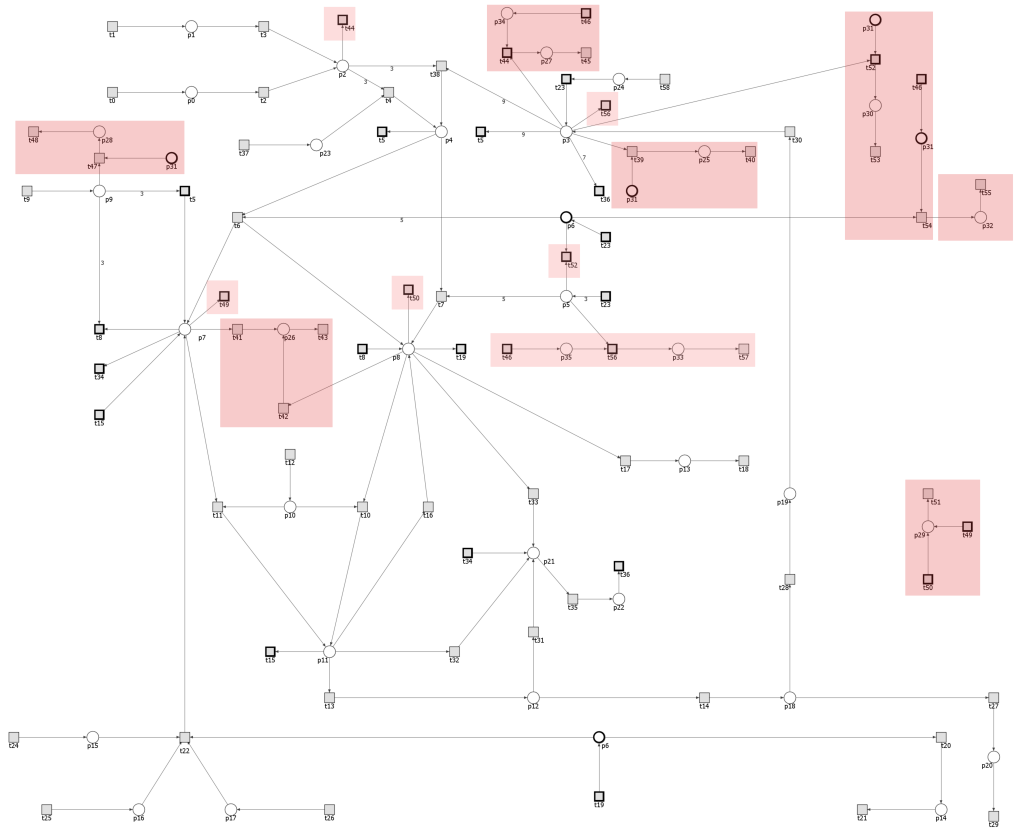


FIGURE 4.3: Petri net-based model of HDL metabolism - inflammation extension. Red highlighted regions denote inflammatory processes. The color coding indicates only the process category; it does not represent the direction or magnitude of changes in the corresponding arc weights (i.e., whether weights increase).

## 4.4 Model extension - treatment

The final model extension introduces targeted therapeutic mechanisms into the inflamed network. This model evaluates how therapeutic strategies counteract inflammatory disruptions and restore lipid transport pathways.

**Therapeutic Nodes and Mechanisms** The treatment extension introduces exogenous components and regulatory interventions designed to mitigate oxidative stress and bypass damaged endogenous pathways:

- **rHDL Infusion:** Modeled as an independent, exogenous lipid acceptor that accelerates peripheral lipid removal. This mechanism operates via rHDL-specific cholesterol export transitions ($t_{60}$-$t_{62}$), leveraging ABCA1, ABCG1, and SR-B1 pathways to offset the functional deficits of endogenous HDL.

45:2749136461

- **Peptide 5A Therapy:** Represented as an anti-inflammatory peptide node ($p_{39}$) that directly lessens the rate of inflammation. This component targets the core signaling pathways driving transporter down-regulation.
- **Antioxidant Supplementation:** Simulated through an oxidation reduction transition ($t_{67}$), which replicates the biochemical impact of vitamins C and E. This node reduces the generation rate of oxidized HDL, preserving particle integrity.
- **Inflammation Mitigation:** Managed by an explicit inflammation inhibition transition ($t_{65}$).

These therapeutic changes and their corresponding mechanisms of action are detailed in Table 4.2. From a structural perspective, this extension repositions the network's primary control points, shifting dominance away from pathology-driven transitions toward pathways that reestablish lipid clearance and restore metabolic balance.

TABLE 4.2: Modifications introduced to the model to represent therapeutic interventions.

| **Place/transition name; modification** | **Literature support** |
|---|---|
| increase in weights of arcs:<br>• cholesterol transfer ($t_{30}$) → ABCA1 ($p_3$) (new weight: 3)<br>• cholesterol export ($t_{23}$) → ABCA1 ($p_3$) (new weight: 3) | The HDL-modulating effect of theobromine has been attributed to enhanced ABCA1 activity. This mechanism is proposed to involve phosphodiesterase inhibition, which alters intracellular cyclic adenosine monophosphate (cAMP) signaling and promotes ABCA1-mediated cholesterol efflux, thereby supporting HDL biogenesis [95]. |
| vitamins C and E effect; increase in weights of arcs:<br>• ABCA1 ($p_3$) → activity decrease ($t_{56}$) (new weight: 3)<br>• ABCG1 ($p_5$) → activity decrease ($t_{56}$) (new weight: 3) | Vitamin C and E supplementation has been shown to enhance the antioxidant properties of HDL by promoting beneficial changes in its protein composition, including increased levels of PON1. Since PON1 is a major determinant of HDL antioxidant capacity, its increased abundance improves protection against HDL oxidation, whereas PON1 deficiency is associated with diminished antioxidant activity [95]. |
| niacin - increase in weight of arc spherical HDL2 ($p_8$) → HDL catabolism ($t_{17}$) to 3 | The interaction between niacin and HDL catabolism is not explicitly discussed in the main text of the cited publication. The modification of this interaction in the Petri net model is therefore based on the conceptual pathway illustrated in Figure 4 rather than on direct evidence reported in the article [95]. |

46:4475039848

| **Place/transition name; modification** | **Literature support** |
|---|---|
| increase in weights of arcs:<br>• fatty acid release ($t_{32}$) → fatty acid ($p_{21}$) (new weight: 3)<br>• ApoA-I synthesis ($t_2$, $t_3$) → lipid free ApoA-I ($p_2$) (new weight: 3)<br>• cholesterol export ($t_{23}$) → ABCA1 ($p_3$) (increased weight by 2; new weight: 5)<br>• cholesterol transfer ($t_{30}$) → ABCA1 ($p_3$) (increased weight by 2; new weight: 5) | Activation of Peroxisome proliferator-activated receptor (PPAR) $\alpha$ enhances lipid metabolism by increasing the expression of lipoprotein lipase (LPL), thereby promoting triglyceride hydrolysis and fatty acid release. In addition, fibrates stimulate the synthesis of ApoA-I and upregulate ABCA1 expression, supporting HDL biogenesis and cholesterol efflux [95]. |
| CETP inhibition; increase in weights of arcs:<br>• CETP ($p_{10}$) → cholesteryl ester transfer ($t_{10}$) (new weight: 3)<br>• CETP ($p_{10}$) → cholesteryl ester transfer ($t_{11}$) (new weight: 3) | CETP activity is inversely associated with circulating HDL-C concentrations because it mediates the exchange of cholesteryl esters and triglycerides between lipoproteins. Consequently, pharmacological inhibition of CETP has been extensively investigated as a strategy to reduce cholesteryl ester transfer and increase plasma HDL-C levels [95]. |
| increased in weight of arc ApoA-I synthesis ($t_2$, $t_3$) to lipid free ApoA-I ($p_2$) by 2; new weight: 5 | Apabetalone (RVX-208), a bromodomain and extraterminal (BET) protein inhibitor, has been shown to enhance ApoA-I synthesis through epigenetic regulation. By increasing ApoA-I production, apabetalone promotes HDL biogenesis and favorably influences HDL-related metabolic pathways [95]. |
| increase in weights of arcs:<br>• cholesterol export ($t_{23}$) to ABCA1 ($p_3$) (increased weight by 2; new weight: 7)<br>• cholesterol transfer ($t_{30}$) to ABCA1 ($p_3$) (increased weight by 2; new weight: 7)<br>• cholesterol export ($t_{23}$) to ABCG1 ($p_5$) (increased weight by 2; new weight: 5) | The ApoA-I mimetic peptide 5A has been shown to promote cholesterol efflux by activating the ABCA1- and ABCG1-mediated transport pathways. Enhanced cholesterol export through these transporters improves reverse cholesterol transport and has been associated with reduced atherosclerotic burden in preclinical studies [95]. |

47:6790132411

| **Place/transition name; modification** | **Literature support** |
|---|---|
| added new place - rHDL ($p_{36}$) | rHDL particles have been developed to enhance ABCA1-mediated cholesterol efflux from cholesterol-loaded macrophages. In addition to promoting reverse cholesterol transport, rHDL has been investigated as a therapeutic carrier for targeted delivery to macrophages and atherosclerotic plaques [56]. |
| added new arcs cholesterol export from SR-B1 ($p_6$) and ABCG1 ($p_5$) to rHDL ($p_{36}$) | Based on expert answer. |
| peptide 5A ($p_{39}$); changes:<br>• increased weight of every arc representing inflammation by 2<br>• increased weight of every arc representing oxidation by 2 (new weight: 5) | The ApoA-I mimetic peptide 5A has been shown to attenuate both inflammatory responses and oxidative stress in experimental models, including human endothelial cells and rabbit carotid arteries. These findings indicate that, beyond enhancing cholesterol efflux, 5A exerts anti-inflammatory and antioxidant effects that contribute to its atheroprotective properties [95]. |

The final Petri net-based model is illustrated in Figure 4.4. This model extends the inflammation PN model with therapeutic interventions. Consistent with the preceding models, the complete inventory of places is presented in Table 4.3, while the corresponding set of transitions is available in Table 4.4. Invariant analysis reveals that the model is covered by 17,516 $t$-invariants. In alignment with the structural properties observed in the baseline and inflammation nets, the analytical utility of the 19 identified MCT sets remains highly constrained. Specifically, 17 of these sets contain only two transitions each, while the remaining two sets comprise merely three and four transitions, respectively. The remaining 27 transitions do not partition into any non-trivial MCT set, reinforcing the presence of localized, independent functional pathways within the global metabolic network.

## 4.5 In silico knock-out analysis

To systematically evaluate how individual structural components regulate global system dynamics, an in silico knock-out analysis was executed. By selectively disabling specific transitions and comparing the resulting token distributions and flow properties against baseline reference simulations, the importance of each network element was quantified. This sensitivity assessment was performed independently across both the inflammatory extension and the treatment model to explore how external interventions alter internal regulatory dependencies.

In the unmitigated inflammatory state, network behavior is heavily governed by transitions reflecting pathological pathways, particularly those representing impaired HDL maturation and chronic inflammation. Specifically, disabling transitions corresponding to reduced esterification ($t_{47}$), adiponectin decreased production ($t_{44}$), and activity decrease associated with oxidized HDL ($t_{56}$) caused massive, system-wide reallocations of token flow. The profound network vulnerabilities exposed by the knock-outs of $t_{47}$, $t_{44}$, and $t_{56}$ underscore their critical role in sustaining the self-perpetuating pathological state of the system.

48:9629429017

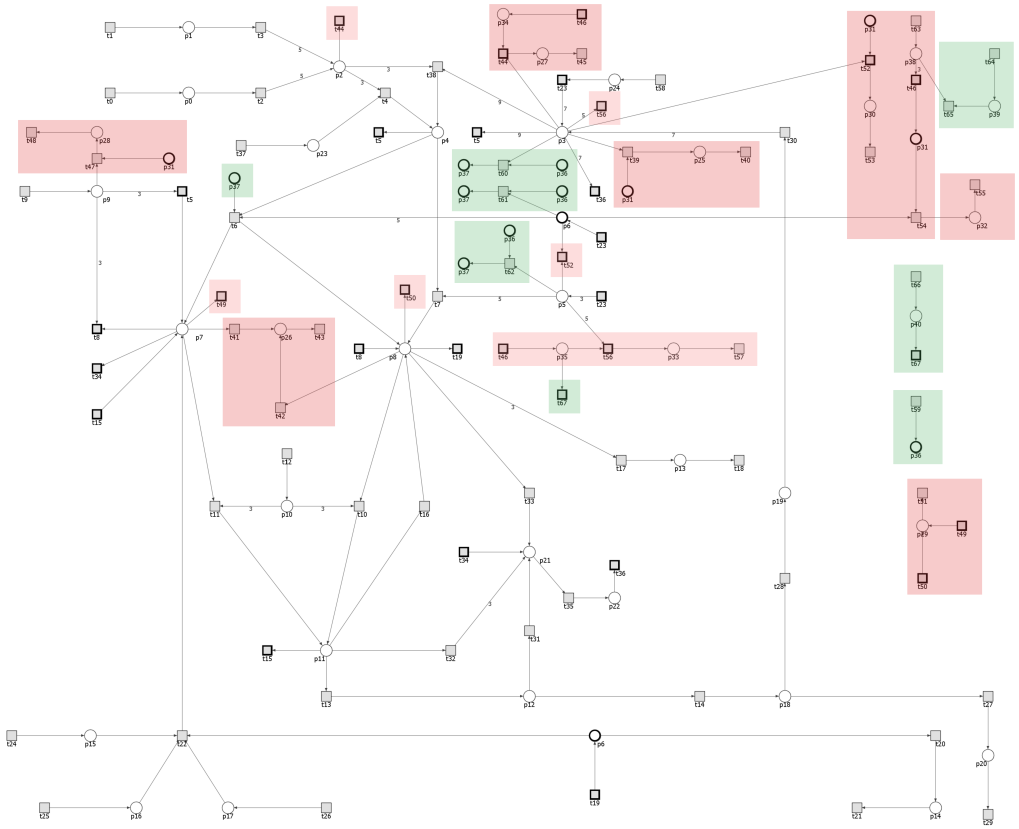


FIGURE 4.4: Petri net-based model of HDL metabolism - treatment extension. Red highlighted regions denote inflammatory processes, whereas green highlighted regions denote treatment-related processes. The color coding indicates only the process category; it does not represent the direction or magnitude of changes in the corresponding arc weights (i.e., whether weights increase or decrease).

The structural sensitivity profile of the PN shifted substantially upon the introduction of therapeutic mechanisms. Within the treatment-augmented model, the nodes possessing the highest regulatory leverage switched to those directly representing the intervention strategies. Notable among these were inflammation inhibition ($t_{65}$), oxidation reduction ($t_{67}$), and reconstituted HDL-mediated cholesterol export via ABCA1, SR-B1, and ABCG1 ($t_{60}$–$t_{62}$). Eliminating these therapeutic transitions induced a major collapse of the stabilized network structure, marked by a sharp drop in cholesterol efflux capability and the immediate return of inflammation-driven token accumulation.

A comparative analysis of both models reveals that the primary control hubs are highly context-dependent. The most influential transitions identified for each model variant are summarized in Table 4.5.

To evaluate these global shifts visually, the comprehensive impact of each individual transition knock-out was mapped onto a unified heatmap (Figure 4.5). This visualization presents a clear overview of system-wide reactions by plotting variations in token flow and distribution against the respective reference models.

The simulation results indicate that a substantial majority of the single-transition knock-outs provoked only minor, localized disruptions. This behavior indicates that large sectors of the metabolic network possess inherent topological robustness. Conversely, a select group of critical transitions acted as network-wide vulnerabilities. Deleting them caused widespread redistribution of token flow that affected multiple downstream paths, including HDL remodeling, cholesterol

49:6942260987

Table 4.3: List of places

| Place | Biological meaning |
|---|---|
| $p_0$ | liver |
| $p_1$ | intestine |
| $p_2$ | lipid free ApoA-I |
| $p_3$ | ABCA1 |
| $p_4$ | discoidal HDL |
| $p_5$ | ABCG1 |
| $p_6$ | SR-B1 |
| $p_7$ | spherical HDL3 |
| $p_8$ | spherical HDL2 |
| $p_9$ | LCAT |
| $p_{10}$ | CETP |
| $p_{11}$ | VLDL |
| $p_{12}$ | IDL |
| $p_{13}$ | High-Density Lipoprotein Receptor (HDLR) |
| $p_{14}$ | kidney |
| $p_{15}$ | PLTP |
| $p_{16}$ | Hepatic Lipase (HL) |
| $p_{17}$ | Endothelial Lipase |
| $p_{18}$ | LDL |
| $p_{19}$ | Low-Density Lipoprotein Receptor (LDLR) |
| $p_{20}$ | Lipoprotein(a) |
| $p_{21}$ | fatty acid (FA) |
| $p_{22}$ | Cluster of Differentiation 36 (CD36) |
| $p_{23}$ | phospholipid |
| $p_{24}$ | cholesterol |
| $p_{25}$ | impaired cholesterol |
| $p_{26}$ | TG-enriched HDL-C clearance |
| $p_{27}$ | decreased expression |
| $p_{28}$ | reduced formation of HDL-C |
| $p_{29}$ | HDL hydrolysis increased by sPLA2 |
| $p_{30}$ | decreased expression |
| $p_{31}$ | inflammation |
| $p_{32}$ | decreased expression |
| $p_{33}$ | impairing HDL-C |
| $p_{34}$ | TNF-$\alpha$ presence |
| $p_{35}$ | oxidized HDL |
| $p_{36}$ | rHDL |
| $p_{37}$ | exported cholesterol |
| $p_{38}$ | inflammation condition |
| $p_{39}$ | 5A |
| $p_{40}$ | vitamin C and E |

transport, and inflammatory signaling. These critical nodes are marked on the heatmap by intense positive or negative deviations from baseline values.

In particular, transitions associated with cholesterol esterification (LCAT-related processes), ABCA1- and ABCG1-mediated cholesterol efflux, and rHDL-dependent export exhibited the most pronounced network-wide consequences. These findings confirm the central placement of these pathways within the theoretical framework of HDL dynamics. The heatmap effectively contextualizes the primary regulatory junctions, illustrating that control over the network's state is not uniformly distributed but is instead concentrated within a few highly sensitive bottlenecks, which are summarized in Table 4.6. Together, these observations indicate that HDL metabolism is best

50:4539337408

understood as a hierarchically organized network in which a limited number of processes show disproportionate control over overall system behavior.

## 4.6 Chapter summary

A major insight derived from the created models is that control over the HDL metabolic network is highly centralized, relying on a small number of core regulatory transitions. As summarized in Table 4.5, the vast majority of network components have minimal impact on global dynamics, while a few specific processes function as critical regulatory bottlenecks.

Under unmitigated inflammatory conditions, these bottlenecks are defined by compromised esterification ($t_{47}$), oxidized HDL-mediated impairment ($t_{56}$), and diminished adiponectin regulation ($t_{44}$). These specific processes anchor a pathological control regime that destabilizes homeostatic HDL functions and drives the system toward a pro-inflammatory state.

Conversely, introducing therapeutic interventions restructures this control topography, shifting dependency toward processes that mitigate inflammation and restore redox equilibrium. Transitions representing inflammation inhibition ($t_{65}$) and oxidation reduction ($t_{67}$) emerge as the dominant regulatory hubs. This transition indicates that successful clinical therapies do not merely act as isolated local patches; rather, they alter the global structural behavior of the metabolic network.

Interestingly, the fact that inflammation and redox remain highly influential across both the pathological and therapeutic states indicates the existence of an invariant core regulatory backbone. This persistent backbone maintains its network-wide relevance despite wide-ranging adaptations among peripheral pathways. This suggests that therapeutic interventions targeting these core invariant nodes may yield sweeping, system-wide benefits that far exceed their direct, local biochemical interactions. Collectively, these simulations imply that HDL metabolism is best conceptualized not as a series of independent parallel pathways, but as a hierarchical network governed by a limited set of master regulatory processes.

This research offers a systems-level analysis of HDL dynamics across inflammatory and therapeutic contexts, demonstrating that HDL functionality is an emergent property of interconnected networks rather than the product of isolated molecular reactions. By mapping established biological pathways onto a Petri net framework and utilizing in silico knock-out methodologies, context-dependent regulatory behaviors that are difficult to isolate using conventional, reductionist experimental designs have been highlighted.

During chronic inflammation, the system is dominated by processes tied to defective HDL maturation and inflammatory damage, such as reduced cholesterol esterification, decreased adiponectin-dependent regulation, and loss of oxidized HDL-related activity. These computational outcomes match a large body of clinical and experimental literature showing that persistent inflammatory states trigger structural remodeling and functional damage in HDL particles, shifting them from protective elements to pro-inflammatory agents [47, 117, 119, 95].

When therapeutic interventions are introduced, the network's internal control structure undergoes an extensive reconfiguration. Control shifts toward processes associated with clinical modulation, such as the suppression of active inflammation, the lowering of oxidative stress, and rHDL-mediated cholesterol clearance via ABCA1, SR-B1, and ABCG1 pathways. This shift reinforces the clinical view that effective HDL-targeted therapies should not aim to repair single isolated reactions; instead, they must recalibrate the overall balance between metabolic pathways and inflammatory cascades. This network-level adaptation is consistent with modern clinical paradigms

51:2507520734

emphasizing that patient outcomes depend on improving overall HDL functionality rather than merely raising plasma HDL concentration [85, 34, 23].

The heatmap results further emphasize this concentration of control. While most knock-outs produced localized or negligible changes, a small subset of transitions triggered profound, network-wide reallocations of token flow across lipid transport, remodeling, and signaling pathways. Such disproportionate regulatory influence by select nodes is a well-documented feature of complex biological architectures and matches prior Petri net investigations of regulatory and metabolic processes [68]. These results validate knock-out sensitivity analyses as an effective strategy for finding regulatory bottlenecks in large-scale biological systems.

Viewing HDL metabolism through systems biology perspective clarifies that disease and therapeutic states represent distinct organizational regimes of a single underlying network. This perspective helps explain why interventions that appear promising in one network state may not produce the same effects in another, highlighting the need for context-aware therapeutic design.

Several limitations of this work should be noted. The Petri net model developed here is qualitative, relying on literature-derived insights regarding the direction and relative influence of interactions rather than precise kinetic rate constants. While this qualitative approach is optimal for uncovering global control topologies and network-wide dependencies, it cannot project absolute molecular concentrations or exact temporal profiles. Additionally, the therapeutic pathways analyzed here represent a targeted selection of clinical strategies rather than an exhaustive survey of all possible medical interventions.

52:6389727570

Table 4.4: List of transitions

| Transition | Biological meaning | Transition | Biological meaning |
|---|---|---|---|
| $t_0$ | ApoA1 gene expression in the intestine | $t_{34}$ | fatty acid release from HDL3 |
| $t_1$ | ApoA1 gene expression in the liver | $t_{35}$ | fatty acids absorption |
| $t_2$ | ApoA-I synthesis in the liver | $t_{36}$ | absorbed FA in peripheral tissue |
| $t_3$ | ApoA-I synthesis in the intestine | $t_{37}$ | phospholipids source |
| $t_4$ | discoidal HDL formation | $t_{38}$ | discoidal HDL formation |
| $t_5$ | enzymatic reactions | $t_{39}$ | inflammatory function suppression |
| $t_6$ | cholesterol export to HDL2 and HDL3 | $t_{40}$ | decreased efflux |
| $t_7$ | cholesterol export to HDL2 | $t_{41}$ | increased exchange of TGs for cholesteryl esters from HDL3 |
| $t_8$ | enzymatic reactions | $t_{42}$ | increased exchange of TGs for cholesteryl esters from HDL2 |
| $t_9$ | LCAT source | $t_{43}$ | decreased concentration of HDL |
| $t_{10}$ | cholesteryl ester transfer from HDL2 | $t_{44}$ | adiponectin decreased production |
| $t_{11}$ | cholesteryl ester transfer from HDL3 | $t_{45}$ | TNF-alpha removal |
| $t_{12}$ | CETP source | $t_{46}$ | inflammation source |
| $t_{13}$ | formation of VLDL remnants | $t_{47}$ | reduced esterification |
| $t_{14}$ | loss of fatty acids | $t_{48}$ | reduced esterification removal |
| $t_{15}$ | cholesteryl ester transfer from VLDL to HDL3 | $t_{49}$ | HDL3 hydrolysis |
| $t_{16}$ | cholesteryl ester transfer from VLDL to HDL2 | $t_{50}$ | HDL2 hydrolysis |
| $t_{17}$ | HDL catabolism | $t_{51}$ | hydrolyzed HDL |
| $t_{18}$ | catabolized HDL | $t_{52}$ | retinoic acid receptor-alpha alterations |
| $t_{19}$ | cholesterol binding | $t_{53}$ | retinoic acid receptor-alpha alterations removal |
| $t_{20}$ | cholesterol excretion | $t_{54}$ | decreased expression by INF gamma |
| $t_{21}$ | excreted cholesterol | $t_{55}$ | INF-gamma removal |
| $t_{22}$ | HDL2 remodeling | $t_{56}$ | activity decrease |
| $t_{23}$ | cholesterol export | $t_{57}$ | oxHDL removal |
| $t_{24}$ | PLTP source | $t_{58}$ | cholesterol source |
| $t_{25}$ | HL source | $t_{59}$ | HDL infusion |
| $t_{26}$ | Endothelial Lipase source | $t_{60}$ | cholesterol export promoted by rHDL in ABCA1 |
| $t_{27}$ | Lp(a) forming | $t_{61}$ | cholesterol export promoted by rHDL in SR-B1 |
| $t_{28}$ | LDL binding | $t_{62}$ | cholesterol export promoted by rHDL in ABCG1 |
| $t_{29}$ | Lp(a) metabolism | $t_{63}$ | inflammation state |
| $t_{30}$ | cholesterol transfer | $t_{64}$ | peptide 5A source |
| $t_{31}$ | fatty acid release from IDL | $t_{65}$ | inflammation inhibition |
| $t_{32}$ | fatty acid release from VLDL | $t_{66}$ | treatment source |
| $t_{33}$ | fatty acid release from HDL2 | $t_{67}$ | oxidation reduction |

53:9837356603

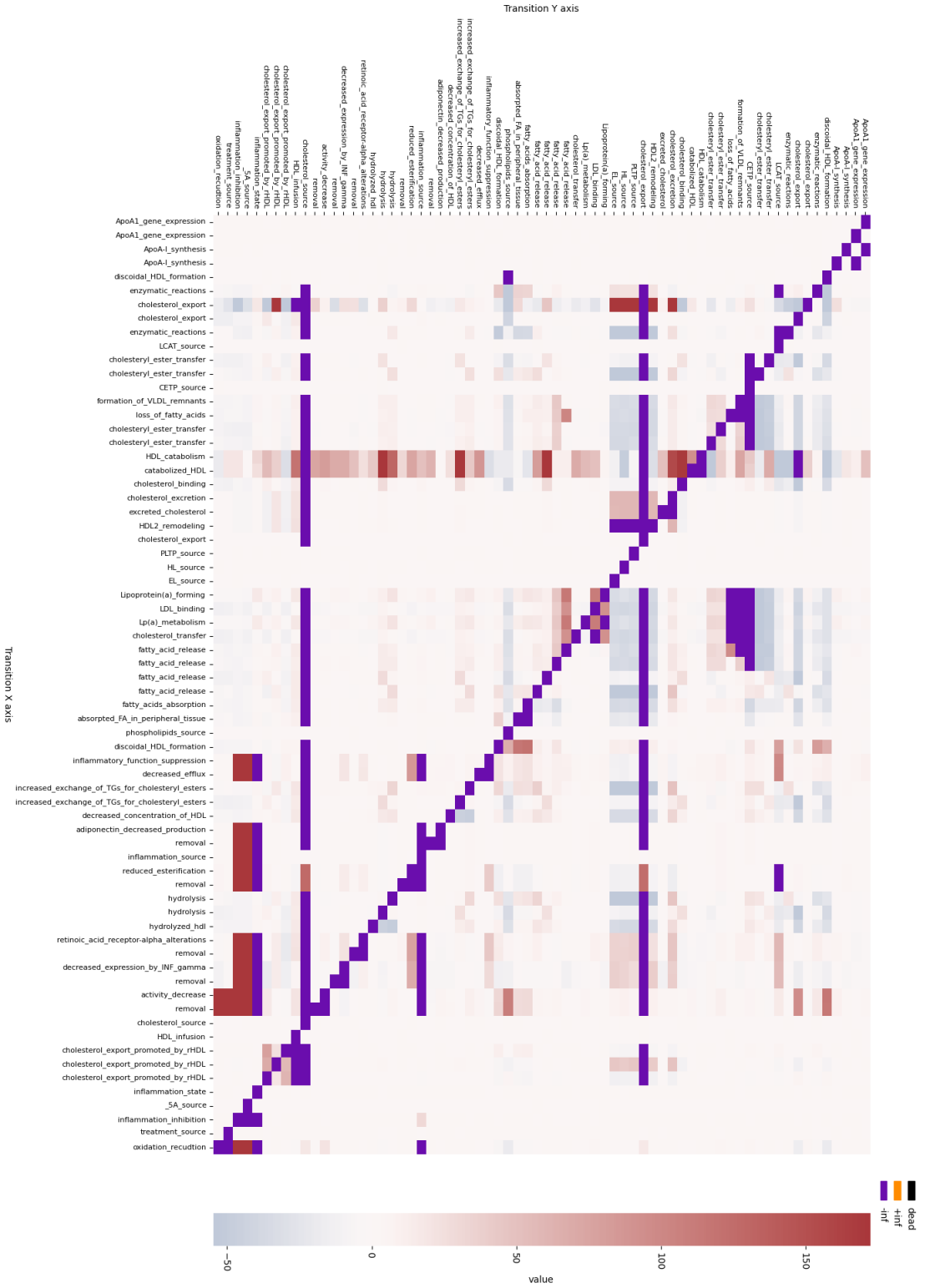



FIGURE 4.5: Heatmap of in silico knock-out analysis results. Positive values (red shades) indicate increased token flow through a given transition relative to the reference model, whereas negative values (blue shades) indicate reduced activity. Values marked as *+inf* correspond to transitions that become active only in the knock-out scenario, while *-inf* indicates complete loss of activity when a given transition is disabled. Transitions marked as *dead* would correspond to transitions inactive in both reference and knock-out conditions; such cases do not occur in the present model.

54:3432673773

TABLE 4.5: Key transitions identified by in silico knock-out analysis in inflammatory and therapeutic HDL Petri net-based models.

| **Model context** | **Transition ID** | **Biological process represented** | **Observed system-level impact (summary)** |
|---|---|---|---|
| **Inflammation model (no therapy)** | $t_{44}$ | Adiponectin decreased production | Strong redistribution of LDL/HDL remodeling fluxes; increased inflammatory load and impaired HDL-C formation |
| | $t_{56}$ | Activity decrease (oxidized HDL-related) | Marked shift toward oxidized HDL and compensatory activation of alternative efflux pathways |
| | $t_{47}$ | Reduced esterification (LCAT-related) | Impaired maturation of HDL particles and increased inflammatory signaling |
| **Treatment model** | $t_{65}$ | Inflammation inhibition | Global suppression of inflammatory token accumulation and partial normalization of HDL functionality |
| | $t_{67}$ | Oxidation reduction | Strong reduction of oxidized HDL burden and restoration of antioxidant-related processes |
| | $t_{60}$ | Cholesterol export promoted by rHDL (ABCA1-dependent) | Enhanced cholesterol efflux toward rHDL and reduced intracellular cholesterol accumulation |
| | $t_{61}$ | Cholesterol export promoted by rHDL (SR-B1-dependent) | Increased hepatic and renal cholesterol handling via SR-B1-mediated pathways |
| | $t_{62}$ | Cholesterol export promoted by rHDL (ABCG1-dependent) | Reinforcement of peripheral cholesterol efflux and HDL remodeling |

55:1167359418

Table 4.6: Regulatory bottlenecks of HDL metabolism

| **Bottleneck transition** | **Core biological function** | **System-level role** | **Clinical / translational implication** |
|---|---|---|---|
| $t_{47}$ (Reduced esterification, LCAT-related) | Conversion of free cholesterol into cholesteryl esters; HDL maturation | Metabolic choke point controlling transition from immature to functional HDL | Impaired LCAT activity may explain dysfunctional HDL and ineffective reverse cholesterol transport despite normal HDL levels. |
| $t_{56}$ (Oxidized HDL / activity decrease) | Oxidative modification of HDL and inhibition of transporter activity (ABCA1/ABCG1) | Inflammation - oxidation hub amplifying network instability and compensatory responses | Supports the role of oxidative stress as a driver of HDL dysfunction and a potential target for redox-based therapies. |
| $t_{44}$ (Adiponectin decrease) | Regulation of lipid metabolism and anti-inflammatory signaling | Integrator of metabolic and inflammatory signaling | Links adipose tissue dysfunction with HDL impairment; highlights the relevance of metabolic syndrome in HDL biology. |
| $t_{65}$ (Inflammation inhibition) | Suppression of pro-inflammatory signaling pathways | Global master regulator restoring network stability | Suggests anti-inflammatory strategies as key drivers of HDL functional recovery. |
| $t_{67}$ (Oxidation reduction) | Reduction of oxidative stress and restoration of redox balance | Core regulatory backbone shared across states | Identifies redox modulation as a central therapeutic axis in HDL-related disorders. |
| $t_{60}$–$t_{62}$ (rHDL-mediated cholesterol export) | Enhanced cholesterol efflux via ABCA1, SR-B1, and ABCG1 pathways | Effector layer restoring lipid flux | Explains why direct HDL infusion strategies may work only when upstream regulation is restored. |

56:7340428343

# Chapter 5

# Named Entity Recognition

## 5.1 Introduction

NER was initially conceived in this work as a foundational step for extracting structured semantic elements - such as substrates, modifiers, and products - that could subsequently be mapped onto places in a Petri net representation. From this perspective, NER is not only a text classification task, but also an interface between unstructured natural language and a formal process representation. Reliable identification of entity boundaries and types is therefore a prerequisite for preserving the semantic content of the source text during the construction of downstream process models.

Early experiments revealed that large language models can perform NER with high accuracy and can recognize considerably more diverse entity categories than traditional supervised sequence-labeling models. However, existing LLM-based approaches frequently impose constraints that limit their applicability in realistic information-extraction pipelines. In particular, prompt-based methods may require a separate inference pass for each entity type, return entity strings without unambiguous character or token offsets, or produce outputs whose format varies between generations. Their relatively high computational and memory requirements can additionally impact large-scale processing.

These limitations motivated a closer investigation of NER architectures that combine the generalization capabilities of instruction-tuned language models with predictable, structured output and efficient inference. Of particular interest were models capable of recognizing domain-specific entity categories without requiring a separately trained classifier for every target ontology. Such flexibility is important in process-oriented applications, where the relevant semantic categories may differ substantially between domains and may not correspond to conventional NER classes such as person, organization, or location.

This chapter presents two complementary contributions arising from this investigation. First, it introduces a set of practical enhancements to the `UniversalNER` model [147], focusing on improved usability and computational efficiency through explicit tokenization and weight quantization. Second, it proposes a new model family, `ner-on`, designed to perform multi-type entity recognition in a single forward pass while producing explicit span-level annotations. Although subsequent advances in LLM capabilities ultimately enabled more direct end-to-end generation of Petri-net-based models without a separate NER stage, the models developed in this chapter provide insights into scalable information extraction, constrained generation, and the conversion of text into structured representations.



57:2496453444

## 5.2 UniversalNER enhancements

### 5.2.1 UniversalNER model family

`UniversalNER` is a family of instruction-tuned, generative models developed for open named entity recognition [147]. In contrast to conventional NER systems, which are commonly trained to predict a fixed set of labels, `UniversalNER` formulates entity recognition as a natural-language instruction-following task. The model receives an input passage together with a requested entity type and generates the textual mentions belonging to that type.

For example, given a sentence and an instruction requesting entities of type "chemical compound", the model is expected to return all expressions in the sentence that satisfy the supplied semantic description. Consequently, the set of recognizable entity types is not restricted to a small ontology encoded in a task-specific classification layer. Entity categories can instead be specified through natural-language labels, allowing the same model to be applied across different domains and annotation schemes.

This formulation is referred to as open NER because the model is intended to generalize beyond a predefined inventory of entity classes. It is particularly relevant in specialized domains, where categories may be fine-grained, infrequent, or absent from standard NER corpora. Examples include scientific concepts, legal terms, biomedical entities, musical categories, and domain-specific process components.

`UniversalNER` was created using a targeted distillation procedure in which a larger language model was used to generate task-specific instruction data from unlabelled text. A smaller student language model was subsequently instruction-tuned on the resulting data. This approach transfers the instruction-following and semantic generalization capabilities of a larger model into a substantially smaller model optimized for entity recognition. The original `UniversalNER` benchmark contains 43 datasets drawn from nine domains, including biomedicine, programming, social media, law, and finance [147]. The breadth of this evaluation reflects the central objective of the model family: robust recognition of heterogeneous entity types across domains rather than specialization in a single benchmark.

The experiments presented in this chapter use `UniNER-7B-all`, a seven-billion-parameter member of the UniversalNER family. The suffix `all` denotes the model variant trained using the complete `UniversalNER` instruction-training collection rather than a restricted domain-specific subset. This variant was selected because it provides a useful balance between broad entity-type coverage, generalization ability, and a model size that remains feasible for deployment on locally available accelerators.

Despite these advantages, the original `UniversalNER` interface has two properties that complicate its integration into larger processing pipelines. First, the reference implementation relies on a conversational prompt template provided through FastChat. This introduces an additional software dependency and obscures the exact token sequence submitted to the model. Second, the full-precision model requires substantial GPU memory, limiting its use on commodity hardware and reducing the number of documents that can be processed concurrently.

Several modifications were therefore introduced to the original `UniversalNER` configuration, specifically to the `UniNER-7B-all` model, to improve usability and efficiency. First, an explicit tokenizer and prompt-construction procedure were integrated, enabling direct model invocation without reliance on the FastChat conversation template. This modification makes the inference procedure easier to reproduce, reduces external dependencies, and provides explicit control over tokenization, prompt formatting, truncation, and generation parameters.

58:8193782230

Second, weight quantization was applied to reduce the model's memory footprint. Quantization decreases the numerical precision used to store model parameters, thereby reducing memory consumption and, depending on the hardware and implementation, potentially improving inference throughput. The principal trade-off is that the approximation introduced by lower-precision weights may reduce predictive accuracy. It was therefore necessary to evaluate whether the memory savings could be achieved without entity recognition quality degradation.

### 5.2.2 Quantization procedure

Quantization was performed using `LLM Compressor` [2], reducing the precision of model weights to 4 bits while preserving activation precision at 16 bits. The process was calibrated using 512 randomly selected samples from the `UniversalNER` dataset. The calibration samples were used to estimate suitable quantization parameters and to adapt the compressed representation to the activation and weight distributions encountered during NER inference.

This quantization configuration is commonly denoted as *W4A16*, where *W4* indicates four-bit model weights and *A16* indicates sixteen-bit activations. Keeping activations at higher precision reduces the risk of numerical instability while still obtaining most of the memory reduction associated with weight-only quantization. The resulting compressed model is referred to as `UniNER-W4A16`[1].

### 5.2.3 Evaluation datasets

The effect of quantization was evaluated on seven English-language test sets covering both document-style text and short user queries. This selection was intended to test whether compression affects performance consistently across different domains, entity ontologies, and linguistic styles.

Five test sets originate from the CrossNER benchmark [84]. CrossNER was introduced for the evaluation of cross-domain NER and contains fully annotated data from five specialized domains: artificial intelligence, literature, music, politics, and natural science. Unlike broad-domain benchmarks that primarily use generic labels such as person, organization, or location, CrossNER includes categories specific to each domain. It therefore provides a suitable test of whether a model can interpret fine-grained entity labels and recognize entities whose meanings depend on specialized subject matter.

The individual CrossNER subsets used in the evaluation have the following characteristics:

**`CrossNER AI`** contains text related to artificial intelligence and includes entities such as algorithm, task, metric, researcher, and research field;

**`CrossNER literature`** focuses on literary works and their context, including writer, book, poem, award, and related organization;

**`CrossNER music`** contains music-related text and categories such as musician, band, albums, song, musical genre, instrument, and award;

**`CrossNER politics`** covers political discourse and includes politician, political party, election, country, governmental organization, and political event;

**`CrossNER science`** corresponds to the natural-science domain and contains scientific discipline, scientist, chemical compound, astronomical object, protein, theory, and other specialized concepts.

[1] The model is available at `https://huggingface.co/daisd-ai`.

59:3385955744

The remaining two test sets, `MIT Movie` [81] and `MIT Restaurant` [81], originate from spoken-language-understanding and slot-filling research. Although these corpora were originally formulated as slot-filling datasets, their annotations define contiguous semantic spans and can consequently be evaluated using an NER-style protocol.

The `MIT Movie` dataset contains natural-language queries requesting movie information. Its annotations cover categories such as actor, director, genre, film title, character, plot, rating, and release-related information. The `MIT Restaurant` dataset contains queries related to restaurant search, with categories including restaurant name, cuisine or dish, location, price range, rating, opening hours, and amenity-related information. Compared with CrossNER, these datasets contain shorter and more conversational inputs. They therefore test the model under a different linguistic distribution, including incomplete sentences, search-style expressions, and compact user requests.

Taken together, the selected datasets provide variation along three important dimensions: subject domain, entity-type granularity, and input style. The CrossNER subsets assess domain transfer and the recognition of specialized entities in relatively conventional text, whereas MIT Movie and MIT Restaurant assess semantic span extraction from short conversational queries.

### 5.2.4 Evaluation results

To assess the impact of quantization on model quality, the full-precision `UniNER-7B-all` model and the quantized `UniNER-W4A16` model were evaluated using entity-level F1 score. F1 is the harmonic mean of precision and recall and therefore reflects both the model's ability to avoid unsupported entity predictions and its ability to recover entities present in the reference annotations. The results are summarized in Table 5.1.

TABLE 5.1: F1 scores of `UniNER-7B-all` and quantized `UniNER-W4A16` models across test datasets.

| Dataset | UniNER-7B-all | UniNER-W4A16 |
|---|---|---|
| CrossNER AI | 62.21 | 61.16 |
| CrossNER literature | 66.59 | 66.36 |
| CrossNER music | 69.20 | 69.46 |
| CrossNER politics | 66.64 | 66.35 |
| CrossNER science | 70.19 | 66.33 |
| MIT Movie | 59.77 | 57.82 |
| MIT Restaurant | 36.70 | 36.92 |

Quantization introduces a modest reduction in aggregate performance, but the effect is not uniform across datasets. On `CrossNER AI`, the F1 score decreases by 1.05 points, while the differences on `CrossNER literature` and `CrossNER politics` are only 0.23 and 0.29 points, respectively. The quantized model also records a reduction of 1.95 points on `MIT Movie`.

The largest decline is observed on `CrossNER science`, where the F1 score decreases from 70.19 to 66.33, corresponding to a difference of 3.86 points. This result suggests that the effect of quantization may depend on the domain and entity inventory. One possible explanation is that the science subset contains a comparatively high proportion of specialized or infrequent terms whose recognition depends on fine distinctions in the model's learned representations. However, establishing the cause of this dataset-specific difference would require a more detailed error analysis, including separate measurements of precision, recall, entity length, and performance by entity type.

In two cases, the quantized model slightly exceeds the score of the original model. Its result is 0.26 points higher on `CrossNER music` and 0.22 points higher on `MIT Restaurant`. These small

60:9998692647

improvements should not be interpreted as evidence that quantization systematically increases model quality. They are more plausibly attributable to minor variation in generative decoding, rounding effects in the aggregate metric, or a regularizing effect that happens to benefit particular examples.

Across all seven datasets, the unweighted mean F1 score decreases from approximately 61.61 for `UniNER-7B-all` to 59.77 for `UniNER-W4A16`, a reduction of approximately 1.84 points. Excluding the larger decrease on `CrossNER science`, the difference is substantially smaller. The results therefore indicate that four-bit weight quantization retains most of the predictive capability of the original model across heterogeneous domains and input styles.

Overall, `UniNER-W4A16` preserves performance broadly comparable to that of the full-precision `UniNER-7B-all` model while substantially reducing the memory required for storing its weights. The findings support the use of W4A16 quantization when deployment constraints make the original model impractical, although they also demonstrate that compression can have a dataset-dependent effect. For applications involving highly specialized scientific terminology, a separate validation or entity-level error analysis is therefore advisable before replacing the full-precision model.

## 5.3 ner-on model

While the `UniversalNER` model provides a simple yet effective strategy for the NER task, its inference paradigm is suboptimal for large-scale or realistic use cases. By design, `UniversalNER` can return entities for only a single entity type per query. Consequently, when processing large text corpora with many entity types, multiple forward passes are required—one for each type—which significantly increases inference time.

In addition, `UniversalNER` outputs only the extracted entity strings, without their corresponding character spans. This limitation becomes critical in cases where the same surface form corresponds to different entities. For example, consider the sentence about Paris Hilton visiting capital of France: "Paris visited Paris." For the entity type *person*, `UniversalNER` would return "Paris," and for the entity type *city*, it would again return "Paris." Without span information, it is impossible to disambiguate these two mentions in the final output. Although this example is deliberately simple, it highlights a broader issue: natural language is inherently ambiguous, and the absence of span-level predictions raises a significant obstacle for reliable entity recognition. To address these limitations, the `ner-on` models were introduced.

The `ner-on` training dataset was constructed as part of this research using knowledge distillation. First, 100,000 documents were randomly sampled from the `sample-10BT` subset of the `FineWeb` dataset [100], with document lengths restricted to between 500 and 1000 tokens. For each document, ChromaDB was then used to retrieve the 40 Wikidata entity types with the highest semantic similarity to the document content. This step produced pairs consisting of raw text and a set of probable entity types.

Subsequently, the `openai/gpt-oss-120b` [97] model was prompted with an instruction to return the original text annotated with entity spans corresponding to the supplied types. An illustrative example of this annotation process is shown in Figure 5.1.

As part of this research, two datasets were derived: (i) a dataset containing both the requested entity types and the corresponding span-level annotations, and (ii) a dataset containing only the annotations, without explicitly providing the entity types. Although prior work has shown that conditioning on entity types generally yields superior performance [147, 145, 125], such information may not always be available in downstream pipelines. To evaluate both settings, two models were

61:4740380093

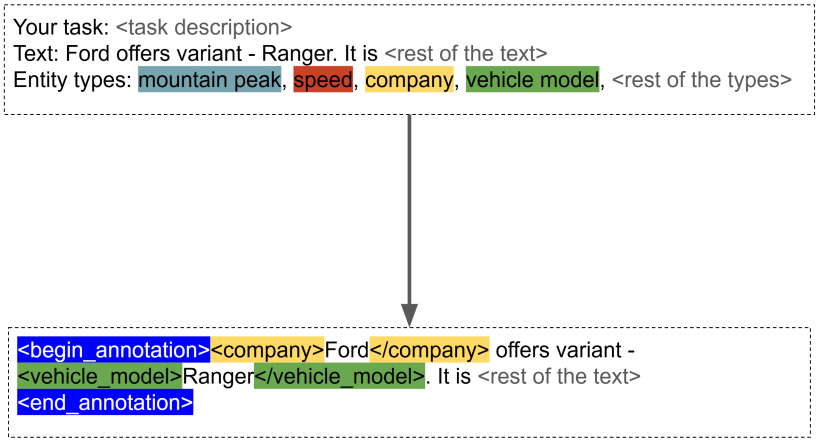



FIGURE 5.1: Example of the annotation process. The model receives, as input, an instruction, the original text, and a list of candidate entity types (upper box). As output (lower box), it returns the same text with each entity annotated using the corresponding type. Explicit begin and end annotation tags are included to mitigate the risk of endless repetition. Colors indicate the association between annotated entities and their respective types.

fine-tuned from `Qwen/Qwen3-4B-Base` [128] using the Alignment Handbook [132] and supervised fine-tuning: (i) `ner-on`, trained without explicit entity types, and (ii) `ner-on-types`, trained with the entity types provided as input.

LLMs are known to suffer from instability in text reproduction [8] and from pathological repetition behaviors [139]. To ensure generation quality, preliminary evaluations were conducted. While the `ner-on-types` model exhibited stable behavior, the `ner-on` model frequently entered an infinite loop, repeatedly generating the closing annotation tag. To mitigate this issue, model merging [3] was applied using mergekit [42]. The `ner-on` and `ner-on-types` models were combined using a linear merge strategy. This approach eliminated the repetition problem and produced a single unified model capable of operating both with and without supplied entity types, referred to as `ner-on-merged`[2].

Finally, the quality of the resulting model was evaluated and compared against `UniNER-W4A16` on three benchmark datasets: *RSS-500* [29], *Reuters-128* [112], and *TweekiGold* [48]. The results are presented in Table 5.2.

TABLE 5.2: Comparison of model precision, defined as the number of correctly recognized entities divided by the total number of predicted entities. The two `ner-on-merged` columns report results for the same fine-tuned model evaluated under two inference settings: without entity types supplied as input (no types sufix) and with the relevant entity types explicitly provided (types sufix).

| Dataset | UniNER-W4A16 | ner-on-merged (no types) | ner-on-merged (types) |
|---|---|---|---|
| RSS-500 | 72.04 | 78.14 | 80.47 |
| Reuters-128 | 79.40 | 81.37 | 83.10 |
| TweekiGold | 83.44 | 83.44 | 82.78 |

The most pronounced performance gain is observed on the *RSS-500* dataset, where both variants of `ner-on-merged` substantially outperform `UniNER-W4A16`. On *Reuters-128*, the improvement remains consistent but more moderate. In contrast, on *TweekiGold*, `UniNER-W4A16` achieves performance comparable to the proposed models, with the variant of `ner-on-merged` trained

[2]The model is available at `https://huggingface.co/daisd-ai`

62:2028849715

without entity types slightly outperforming its type-conditioned counterpart. Overall, these results demonstrate that the proposed approach not only addresses the fundamental limitations of prior methods but also achieves state-of-the-art performance across diverse benchmarks.

## 5.4 Chapter summary

This chapter investigated NER as a foundational component for transforming unstructured biomedical text into structured representations suitable for Petri net construction. The initial objective was to reliably identify substrates, modifiers, and products that could later be linked and incorporated into biomedical models.

Two complementary contributions were presented. First, practical improvements to the `UniversalNER` model demonstrated that aggressive weight quantization can substantially reduce computational requirements while preserving recognition quality. This result enabled large-scale processing and made the approach suitable for integration into interactive modeling workflows. Second, the `ner-on` model family introduced a multi-type, span-aware extraction paradigm capable of recognizing heterogeneous entity classes in a single inference pass. The model addressed key limitations of previous prompt-based approaches, particularly the absence of span information and the need for repeated inference per entity type.

Beyond empirical performance improvements, the experiments provided an important methodological insight. Reliable entity recognition alone does not guarantee coherent structured models: even highly accurate NER systems require substantial post-processing and disambiguation before they can support Petri net-based models. This observation motivated the development of dedicated entity linking methods described in the following chapter.

Finally, the evolution of reasoning-capable language models revealed that explicit intermediate extraction stages may, in some cases, be replaced by direct structured generation. Nevertheless, the techniques developed here remain valuable for scalable corpus processing, dataset construction, and interpretability, and they form the conceptual basis for the modular pipeline explored in subsequent chapters.

63:6257205187

# Chapter 6

# Entity Linking

## 6.1 Introduction

This chapter presents the entity linking work developed in the course of this research. It begins with a discussion of ED and then introduces `FELA`, an end-to-end entity linking framework.

Most existing EL systems are designed for single-use scenarios: they perform well within a specific domain or over a fixed KB, but adapting them to new domains or alternative KBs is often costly. Such adaptation typically requires extensive data annotation, the creation of new datasets, and substantial computational resources. Moreover, rapid advances in model architectures frequently render existing systems obsolete, further limiting their long-term usability.

The central idea of this work is to address these limitations by proposing a flexible framework that can operate with any text-to-text model and any KB. Within the scope of this dissertation, the primary objective of entity linking is to de-duplicate entities extracted during the NER step. While this objective is conceptually straightforward, it is challenging in practice due to the limited granularity of most available KBs.

To illustrate this issue, consider the sentence "HDL undergoes modifications." In Petri net terms, this sentence would give rise to two places: *HDL* and *modified HDL*, with an transition connecting them. Although the entity *HDL* is well represented in existing KBs, the more specific entity *modified HDL*, particularly with a concrete type of modification, is typically absent. In such cases, a conventional EL system would link *modified HDL* to the generic *HDL* entity, leading to incorrect linking and the propagation of errors in downstream tasks.

One possible solution would be the construction of a highly specialized KB with finer-grained entities. However, developing such a resource lies beyond the scope of this dissertation. Instead, this work focuses on designing robust linking and disambiguation mechanisms that operate in a general setting, achieve state-of-the-art performance, and provide a modular, easily extensible pipeline.

Despite these challenges, both the proposed ED framework and the end-to-end EL approach achieve state-of-the-art performance on standard benchmarking datasets, demonstrating the effectiveness and generality of the proposed methods.

## 6.2 Entity Disambiguation

This section describes the proposed ED pipeline. First, a high-level overview of the pipeline architecture is presented. Next, the datasets used for fine-tuning are introduced, followed by a detailed description of the fine-tuning strategies and evaluated models. Finally, the experimental evaluation is discussed, including a comparison with state-of-the-art approaches.



64:3917624469

### 6.2.1 Overview

The proposed pipeline[1], illustrated in Figure 6.1, consists of three core components: (i) an LLM, (ii) an embedding model, and (iii) a vector store.

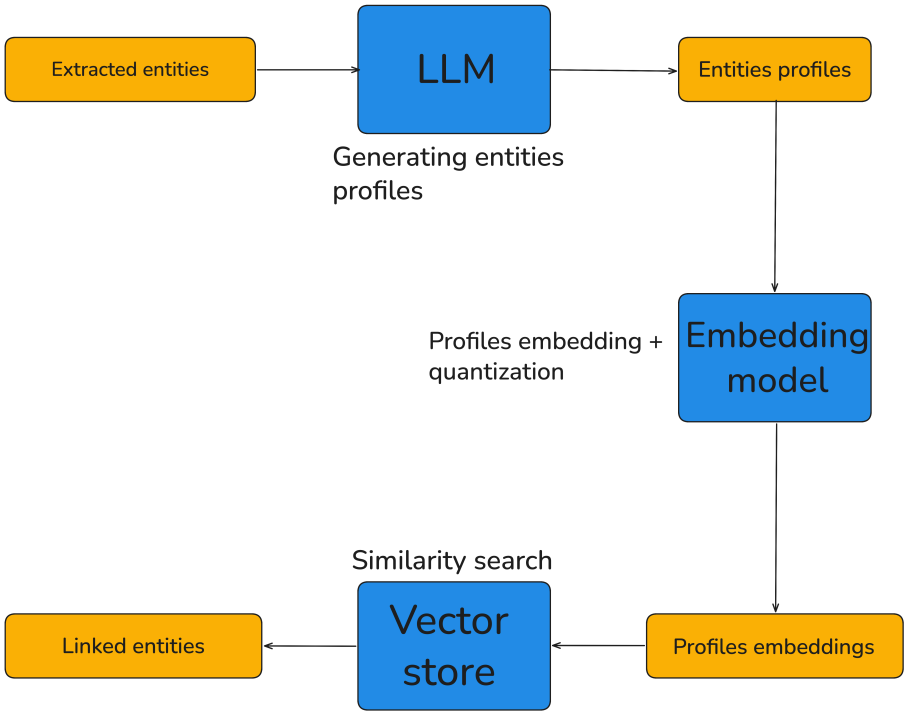



FIGURE 6.1: Overview of the proposed EL pipeline. The process begins with input text containing named entities. These are passed to the LLM (`anydef`), which generates structured entity profiles. The profiles are then converted into 1024-dimensional vectors via the embedding model and stored in a binary-quantized vector store. During inference, query vectors are matched against this store using `Faiss` to retrieve the nearest entity candidates. Each element of this pipeline, that is LLM, embedding model and vector store, may be updated in the future for newer and better solutions.

The central idea of the approach is to leverage an LLM to generate an *entity profile* for each entity recognized in the text corpus together with its surrounding context. Each profile comprises three elements: a disambiguated entity identifier, a textual entity description, and an entity type. An example of such an entity profile is shown in Figure 6.2.

**Entity:** Boston
**Disambiguated entity:** Boston Red Sox
**Entity type:** baseball team
**Definition:** baseball team and Major League Baseball franchise in Boston, Massachusetts, United States

FIGURE 6.2: An example of an entity profile generated for the entity **Boston** in the sentence: "*Boston manager Bobby Valentine said simply, Kelly was claimed by the Mets.*" The profile includes a disambiguated name, entity type, and definition, demonstrating the model's ability to interpret context and distinguish between ambiguous references (in this particular example distinguishing the city from a sports team).

[1] The implementation is available at `https://github.com/daisd-ai/universal-el`

65:2446785967

Each generated entity profile is subsequently embedded into a 1024-dimensional vector using a general-purpose embedding model. These embeddings are intended to capture the semantic information of entities and their contexts, enabling efficient similarity-based retrieval. To reduce both memory consumption and query latency, all vectors, including those stored in the vector store, are quantized to binary precision. Binary quantization converts each floating-point vector component into a single binary value, typically according to whether the original value lies above or below a selected threshold. As a result, every dimension can be represented using one bit rather than a 16- or 32-bit floating-point number. Although this transformation discards part of the numerical information contained in the original embeddings, it substantially reduces storage requirements and allows similarity calculations to be implemented using efficient bitwise operations. In the case of Wikidata, this optimization reduced the storage requirements from approximately 120 GB to only 5 GB, making large-scale deployment feasible.

In the final step, the generated embeddings are used to query a `Faiss` [61] vector store to retrieve the most similar vectors. Each vector is associated with a unique KB identifier, which corresponds to the disambiguated entity. The retrieved candidates are then used to resolve entity ambiguity in the input text.

### 6.2.2 Datasets

Several datasets were introduced or adapted for the ED task:

- `UniversalNER` [147] derived dataset - the original dataset consists of recognized entities extracted from nearly 50,000 text passages. All available instances selected for this experiment were retained.
- `ned graphs` [17] derived dataset - the source dataset contains more than 100,000 pairs of text segments and entities linked to Wikipedia. It is the largest publicly available dataset of this type. All available instances selected for this experiment were retained.
- `KILT Benchmark` [102] derived dataset - a high-quality benchmark composed of multiple sub-datasets. An initial set of approximately 45,000 text - entities pairs was retrieved from the benchmark. Because the original annotations refer to Wikipedia entities, the identifiers were mapped to Wikidata. Instances for which no valid mapping could be established were removed, resulting in a final training set of 31,500 text - entity pairs.

Both the `ned graphs` and `KILT Benchmark` datasets link recognized entities to Wikipedia. While Wikipedia provides a solid foundation with approximately 6 million entities, Wikidata offers a significantly richer ontology with over 100 million entities. To leverage this richness, the `wikimapper` Python package was used to map Wikipedia identifiers to their corresponding Wikidata identifiers. Instances without an available Wikipedia - Wikidata mapping were excluded. This filtering step affected the KILT-derived dataset, reducing its size from approximately 45,000 to 31,500 final training instances.

For each entity instance, contextual information was constructed by extracting up to 250 words to the left and right of the entity mention. This contextual window was used during training to support accurate disambiguation.

### 6.2.3 Models

Using the datasets described above, several fine-tuning strategies were evaluated. It is important to note that although all approaches were benchmarked, models that exhibited performance sub-

66:1727213771

stantially below the state-of-the-art level were discarded together with their training logs. Only the best-performing model was retained for detailed evaluation and comparison with state-of-the-art methods.

All fine-tuning experiments were conducted using `Mistral-7B-v0.1` [60] as the base model. `Mistral-7B-v0.1` is a decoder-only large language model with approximately seven billion parameters, designed to provide strong language-modeling performance while maintaining relatively efficient training and inference. Its selection was guided by the balance between model size, language-modeling capabilities, and the GPU resources available for fine-tuning. The evaluated approaches were as follows:

1. **SFT:** When trained on the adapted `UniversalNER` dataset, the model achieved an average precision of approximately 20%. Due to this low performance, the approach was discarded.

2. **DPO:** Three configurations were evaluated: training on the `ned graphs` dataset only, on the `KILT Benchmark` dataset only, and on a combination of both. Despite the substantially larger size of `ned graphs`, the best results were obtained using only the `KILT Benchmark` dataset. This configuration achieved an average precision of approximately 40%, which remained below state-of-the-art performance and was therefore discarded.

3. **ORPO:** Insights gained from DPO experiments indicated that dataset quality was more important than dataset size. Consequently, only the `KILT Benchmark` dataset was used for ORPO fine-tuning. This approach outperformed all previous methods and resulted in the best overall model, named `anydef` [2], which was selected for comparison with state-of-the-art systems.

### 6.2.4 Evaluation

The proposed pipeline was evaluated on four established entity-linking benchmarks representing different document types and linguistic conditions: `ISTEX-1000` [29], `RSS-500` [29], `Reuters-128` [112], and `TweekiGold` [48]. `ISTEX-1000` contains 1,000 author-affiliation strings extracted from scientific publications and manually annotated with the corresponding entities. Its short, structured inputs primarily test the recognition and disambiguation of research institutions and geographical locations. `RSS-500` consists of 500 news excerpts annotated with entity links and provides broader topical and contextual diversity. `Reuters-128` contains 128 English economic news articles with annotated entity mentions and knowledge-base identifiers, making it suitable for evaluating entity linking in longer, professionally written documents. Finally, `TweekiGold` is a manually annotated collection of posts from Twitter, designed to evaluate entity linking in short and noisy user-generated text. It introduces challenges such as informal language, abbreviations, limited context, and non-standard spelling. Together, these datasets enable evaluation across scientific metadata, news content, economic reporting, and social-media text.

Two evaluation metrics were considered:

- **Precision:** the proportion of correctly linked entities among all evaluated entities.
- **Retrieval rate:** the proportion of cases in which the correct entity appears within the top-$k$ retrieved candidates from the vector store. A higher retrieval rate indicates more effective candidate generation.

[2] The model is available at `https://huggingface.co/daisd-ai`

67:3895785669

Precision results are reported in Table 6.1, while retrieval rates for $k = 25$ candidates are shown in Table 6.2. In both cases, the proposed pipeline was evaluated against the EPGEL system, which served as the reference baseline for assessing entity-linking precision and candidate-retrieval effectiveness.

TABLE 6.1: Precision (expressed as a percentage; with 95% Confidence Intervals (CIs)) of proposed `anydef` pipeline and `EPGEL` processing systems across test datasets.

| | Precision [%] | |
|---|---|---|
| Dataset | anydef | EPGEL |
| RSS-500 | 66.2 ± 7.9 | 72.9 ± 7.3 |
| ISTEX-1000 | 86.7 ± 2.8 | 92.8 ± 1.9 |
| Reuters-128 | 63.8 ± 4.1 | 66.7 ± 3.8 |
| TweekiGold | 75.2 ± 2.9 | 79.5 ± 2.7 |

TABLE 6.2: Retrieval rate (expressed as a percentage; with 95% CIs) of `anydef` and `EPGEL` processing systems across test datasets for 25 candidates.

| | Retrieval rate [%] | |
|---|---|---|
| Dataset | anydef | EPGEL |
| RSS-500 | 82.8 ± 6.0 | 86.8 ± 5.3 |
| ISTEX-1000 | 97.9 ± 1.2 | 98.5 ± 0.9 |
| Reuters-128 | 80.5 ± 3.0 | 77.6 ± 3.6 |
| TweekiGold | 89.9 ± 1.9 | 94.7 ± 1.5 |

In both metrics, `anydef` performs slightly below the state-of-the-art system `EPGEL`. However, this performance gap must be interpreted in light of the vastly different training regimes. The final training dataset for `anydef` consisted of only 31,500 samples, whereas `EPGEL` was trained on approximately 100 million samples. Despite this disparity, `anydef` achieves competitive results while offering a modular, flexible, and fully upgradable architecture.

Overall, these results demonstrate that the proposed approach - referred to more broadly as *universal entity linking* - provides a strong trade-off between performance, scalability, and adaptability, making it well suited for evolving domains and KBs.

## 6.3 FELA – End-to-end Entity Linking

This section introduces `FELA` (*Flexible Entity Linking Approach*), an end-to-end entity linking pipeline that integrates the previously presented NER and ED research into a unified architecture. The section first presents the design principles and workflow of the pipeline, followed by an empirical evaluation and comparison with state-of-the-art systems.

### 6.3.1 FELA overview

The primary design objective of `FELA` [3] is long-term modularity. In modern machine learning practice, model architectures evolve rapidly, often rendering tightly coupled pipelines obsolete after only a few years. `FELA`, illustrated in Figure 6.3, is therefore constructed around a minimal set of assumptions: any text-to-text model may be used as long as it produces outputs conforming

[3] The implementation is available at `https://github.com/daisd-ai/FELA`

68:1143747537

to the expected interface. Consequently, individual components of the system can be replaced without requiring changes to the remaining pipeline.

The only component that cannot be replaced transparently is the embedding model. Since vector representations depend on the embedding space in which they were created, changing the embedding model requires re-embedding the entire KB. This limitation is inherent to all embedding-based retrieval systems rather than specific to `FELA`.

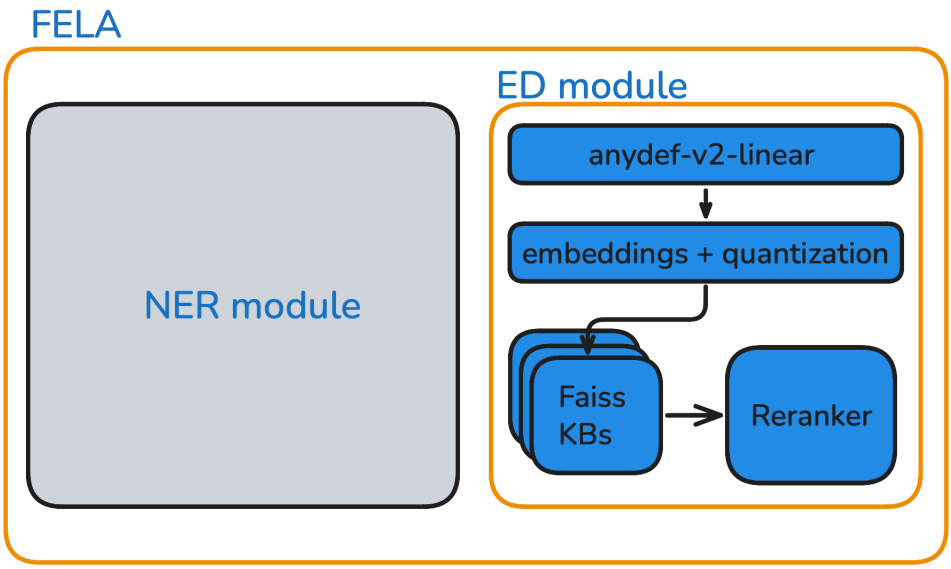



FIGURE 6.3: Overview of the `FELA` pipeline.

Conceptually, `FELA` extends the previously described ED pipeline by introducing an upstream Named Entity Recognition stage. Instead of assuming pre-identified entities, the system extracts, represents, and links them in a single workflow. The processing steps are as follows:

1. The input text is divided into smaller fragments (chunks) to reduce memory usage during both fine-tuning and inference.
2. Named entities are extracted from each chunk using the NER model.
3. For each extracted entity, an entity profile is generated using the ED model.
4. Each profile is embedded and subsequently quantized to reduce storage requirements.
5. For every profile, the system retrieves the top-$k$ candidate entities from each available KB, producing $n \cdot k$ candidates, where $n$ is the number of KBs.
6. A reranking model selects the most plausible candidate among the retrieved entities.

In contrast to the standalone ED experiments, the reranking stage proved essential in the end-to-end setting. Errors introduced during the NER stage propagate to later components, increasing ambiguity in candidate generation. The addition of a reranker significantly improves robustness and is necessary for achieving competitive performance.

### 6.3.2 FELA evaluation

The `FELA` pipeline was evaluated on the same benchmarks used for the ED experiments: `RSS-500`, `Reuters-128`, and `TweekiGold`. The `ISTEX-1000` dataset was excluded because its instances con-

69:6386950421

tain isolated entity mentions without contextual information, which contradicts the design assumptions of context-aware end-to-end linking. For the NER stage, the quantized `UniNER-W4A16` model was employed, while the final candidate selection was performed using the `bge-reranker-v2-gemma` reranking model [75, 21]. This model is a reranker that jointly processes the input context and each candidate entity description to estimate their semantic relevance. In contrast to embedding-based retrieval, which compares independently encoded vectors, the reranker directly models interactions between the query and candidate text, allowing more accurate ordering of the retrieved candidates at the cost of additional computation.

Following the ED evaluation protocol, both precision and retrieval rate were measured. The results were compared against the state-of-the-art system `ReFinED` [6]. Precision scores are presented in Table 6.3, while retrieval rates for $k = 25$ candidates are shown in Table 6.4.

TABLE 6.3: Precision (expressed as a percentage) of `ReFinED` and `FELA` across test datasets.

| | Precision [%] | |
|---|---|---|
| Dataset | ReFinED | FELA |
| RSS-500 | 72.85 | 76.82 |
| Reuters-128 | 59.50 | 51.17 |
| TweekiGold | 65.05 | 71.18 |

TABLE 6.4: Retrieval rate (expressed as a percentage) of `ReFinED` and `FELA` across test datasets for 25 candidates.

| | Retrieval rate [%] | |
|---|---|---|
| Dataset | ReFinED | FELA |
| RSS-500 | 89.40 | 91.39 |
| Reuters-128 | 70.97 | 63.60 |
| TweekiGold | 82.18 | 84.72 |

The results show that `FELA` outperforms `ReFinED` on `RSS-500` and `TweekiGold` in both precision and retrieval rate, while performing worse on `Reuters-128`. The latter dataset contains shorter and less descriptive contexts, which amplifies error propagation from the NER stage and reduces the effectiveness of profile-based disambiguation.

Overall, the experiments demonstrate that a fully modular pipeline can achieve performance competitive with specialized state-of-the-art systems. Importantly, unlike tightly integrated architectures, `FELA` remains adaptable to future advances in NER, embedding models, or LLMs. The approach therefore provides not only strong empirical performance but also a future-proof entity linking framework suitable for evolving KBs and application domains.

## 6.4 Chapter summary

This chapter introduced a flexible approach to entity linking designed for evolving language models and heterogeneous KBs. The primary motivation was not merely achieving high benchmark accuracy, but enabling reliable entity deduplication within automatically constructed biological models, where incorrect or insufficiently specific entity links may cause distinct process elements to be merged or assigned to inappropriate components in the resulting Petri net representation.

The proposed ED method generates contextual entity profiles and performs similarity-based retrieval in a vector space rather than relying on fixed symbolic matching rules. Despite being

70:4142544064

trained on comparatively small datasets, the resulting `anydef` model achieved performance competitive with specialized state-of-the-art systems. These results indicate that high-quality contextual representations can compensate for substantially smaller training corpora when combined with semantic retrieval.

Building upon this component, the `FELA` pipeline integrated recognition, disambiguation, and reranking into a fully modular end-to-end entity linking system. The evaluation demonstrated that such a modular architecture can match or exceed tightly coupled state-of-the-art solution on several benchmarks while remaining adaptable to future advances in language models and embedding techniques. This flexibility is particularly important in long-term scientific workflows, where model components are expected to evolve more rapidly than curated KBs.

However, during integration with literature-derived biological modeling, an important limitation emerged. Biomedical texts frequently describe stateful biological objects whose meaning depends on context (e.g., modified forms, complexes, or functional states). Such entities do not correspond to stable identifiers in existing KBs and therefore cannot be reliably normalized through conventional linking without collapsing biologically distinct concepts into a single node. In practice, enforcing strict entity normalization degraded the structural correctness of generated Petri nets rather than improving it.

This observation revealed that, for systems biology modeling, relations and transformations often constitute the primary information, while entity identity is context-dependent. Consequently, the research direction shifted from strict ontology-based normalization toward relation-centered structured extraction, where entity identity is inferred from interactions rather than predefined identifiers.

Although the entity linking system was therefore not used as a core component of the final modeling pipeline, the work remains an important methodological result. It demonstrates both the feasibility of universal modular entity linking and, more importantly, clarifies the boundary between knowledge-base normalization and process-based biological representation. This insight directly motivated the modeling strategies developed in the subsequent chapter.

71:2790721600

# Chapter 7

# Semi-Automatic Construction of Petri Net-Based Models

## 7.1 Introduction

At the time when the idea for this dissertation was formulated, LLMs were only beginning to emerge as practical tools for NLP. Although the earliest LLMs demonstrated promising capabilities, their application to domain-specific scientific tasks, particularly in biomedical knowledge extraction, was limited and often unreliable. Their performance in structured information extraction required substantial task-specific engineering, including carefully designed pipelines and extensive post-processing.

Consequently, the initial concept for the automatic construction of Petri net-based biomedical models was grounded in a classical multi-stage information extraction pipeline. The envisioned workflow consisted of three principal components:

1. **NER** – identification of relevant biomedical entities (e.g., proteins, genes, biological processes),
2. **EL** – disambiguation of extracted entities and removal of redundancies through alignment with reference ontologies and databases,
3. **RE** – identification of interactions and dependencies between entities, enabling the construction of a structured model.

Such a pipeline reflected the prevailing state-of-the-art methodology in information extraction at that time. It decomposed the problem into manageable sub-tasks, each optimized independently. However, this modular architecture introduced cumulative error propagation, required significant maintenance effort, and demanded large, well-annotated datasets for each stage.

The rapid evolution of LLM architectures and training paradigms fundamentally altered this landscape. The emergence of advanced reasoning-capable models, including DeepSeek R1 and related open-source systems, significantly improved zero-shot and few-shot performance in complex extraction tasks. In parallel, increased accessibility to high-performance computing infrastructure enabled the generation of high-quality synthetic and semi-synthetic training datasets at scale.

These developments motivated a paradigm shift in the methodological approach adopted in this dissertation. Instead of a sequential, multi-stage pipeline, attention moved toward a unified, single-step information extraction strategy. In this approach, an LLM simultaneously identifies entities and the relations between them, effectively producing a structured representation in a single



72:2242066893

inference pass. Preliminary experiments demonstrated that this method can achieve promising results, particularly in terms of coherence between extracted entities and relations.

Nevertheless, it became evident that, despite their remarkable capabilities, LLMs should not be treated as fully autonomous, end-to-end systems for constructing Petri net-based biomedical models. Rather, they function most effectively as intelligent assistants that support domain experts by proposing candidate entities and interactions, which are subsequently validated and refined. This human-in-the-loop paradigm ensures both interpretability and biological plausibility of the resulting models.

In order to ensure methodological rigor and enable systematic evaluation, the scope of automatic model construction was intentionally restricted to a narrowly defined biomedical subdomain. The metabolism and functional role of HDL was selected as the target domain. This choice was motivated by two complementary factors. First, the available computational resources constrained the feasible scale of document processing, model inference, and iterative validation, making full-scale coverage of heterogeneous biomedical pathways impractical. Second, the author and his supervisor possessed prior domain familiarity with HDL biology, which facilitated verification of extracted relations and assessment of biological plausibility of generated Petri nets. Consequently, HDL served as a controlled yet biologically meaningful testbed, allowing the proposed methodology to be evaluated under realistic conditions while maintaining tractable complexity.

Section 7.2 presents the initial implementation of this concept in the form of `Watson`, a platform designed for document-based information extraction and structured representation. Section 7.3 discusses the limitations identified in the first implementation and outlines subsequent refinements.

## 7.2 Watson – Initial Implementation

This section introduces `Watson`, the first prototype platform developed to support semi-automatic construction of Petri net-based biomedical models from textual sources. The primary objective of Watson was to allow for the single-step extraction paradigm while preserving traceability and expert supervision.

The section is structured as follows. First, a general system overview is presented, describing the architectural assumptions and design goals. Next, the datasets used for training and evaluation are characterized. This is followed by a description of the fine-tuned language models employed within the system. Next, processing pipeline is described in detail. Finally, experimental results are presented and discussed.

### 7.2.1 Watson Overview

`Watson` [1] was designed as an end-to-end platform enabling transformation of unstructured biomedical text into structured representations suitable for Petri net construction. The system integrates LLMs with domain-specific post-processing modules and a user interface that facilitates expert validation.

The core design principles of Watson were:

- **Single-step structured extraction** - simultaneous identification of entities and relations,
- **Human-in-the-loop verification** - expert supervision to ensure biological correctness,
- **Modular extensibility** - separation of extraction, validation, and visualization layers,

[1] The platform code is available at `https://github.com/daisd-ai/Watson`

73:5065372197

- **Traceability** - preservation of links (evidence spans) between extracted elements and source text.

`Watson` treats the LLM as a structured reasoning engine capable of producing graph-like representations directly from raw text. The output format was constrained to a predefined schema compatible with Petri net formalism, including places, transitions, and arcs corresponding to biological entities and interactions. The general platform overview is presented in Figure 7.1.

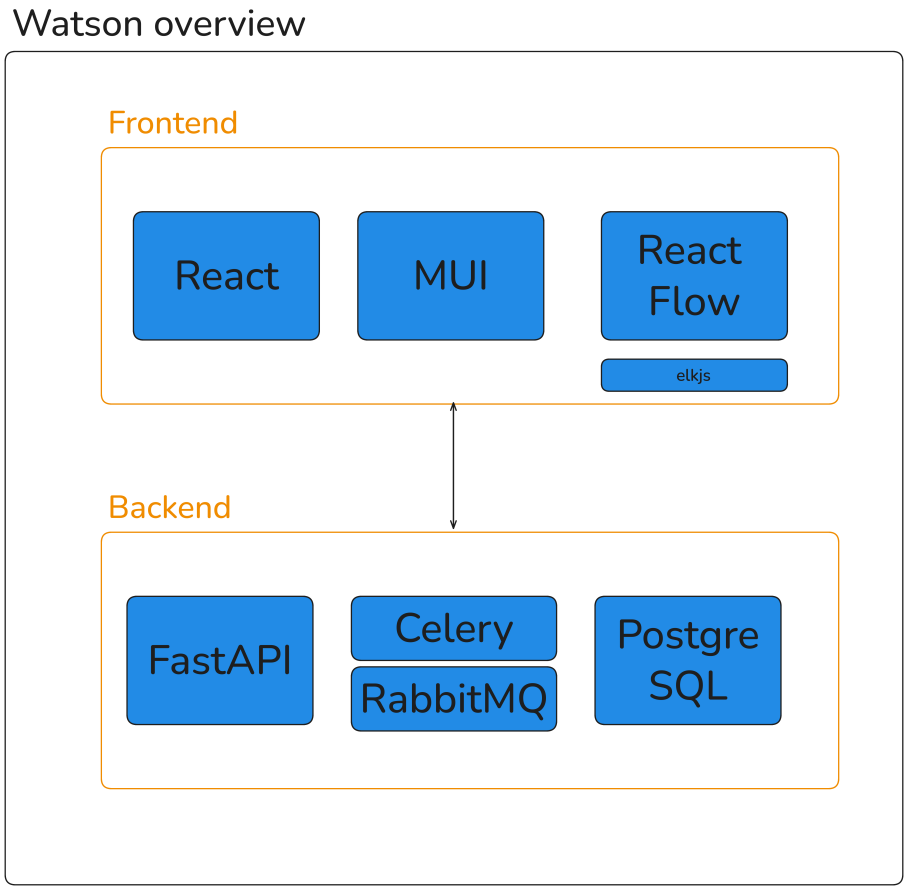



FIGURE 7.1: `Watson` platform architecture. The frontend provides interactive visualization and model inspection, while the backend performs asynchronous document processing and information extraction.

The system follows a client - server architecture and is divided into two major layers: the *frontend* responsible for user interaction and visualization, and the *backend* responsible for document processing, model inference, and persistence.

**Frontend.** The frontend [2] is implemented as a web application built using the React framework [89]. The graphical user interface is constructed using the Material UI (MUI) component library [93], ensuring consistency and usability. Interactive model visualization is handled by the React Flow library [109], while the ELKJS layout engine [32] is used to automatically arrange graph structures corresponding to Petri net elements. This allows the user to inspect extracted entities and relations in a structured graphical form rather than raw textual output.

[2]The frontend component was not implemented by the author of this dissertation and does not constitute part of the dissertation's original contribution. Accordingly, it is described only briefly to provide the necessary context for the overall system architecture.

74:2292749353

**Backend.** The backend exposes a REST API implemented in FastAPI [120]. Computationally expensive tasks, such as document parsing and large language model inference, are executed asynchronously using a Celery task queue [16]. RabbitMQ [11] acts as the message broker coordinating task scheduling and worker execution. Extracted entities, relations, intermediate representations, and user tasks are stored in a PostgreSQL relational database [103], ensuring persistence and reproducibility of results.

**Processing workflow.** The interaction between the frontend and backend proceeds as follows:

1. **Task creation.** The user submits a task through the web interface by providing a task name, a short description of the modeling objective, and one or more PDF documents containing scientific literature.

2. **Asynchronous execution.** The request is registered in the database and dispatched to the task queue. Backend workers retrieve the task, preprocess the documents, execute the information extraction pipeline, and construct an intermediate representation compatible with Petri net formalism.

3. **Result inspection.** After processing is completed, the results become available in the frontend. The user can interactively explore the extracted graph structure, verify correctness, and iteratively refine the model.

This architecture separates interactive visualization from computational processing, enabling responsive user interaction while supporting long-running extraction tasks executed on dedicated computational resources.

### 7.2.2 Datasets

This subsection describes the datasets constructed for (i) model fine-tuning, (ii) evaluation and (iii) evidence attribution. Because no publicly available corpora existed that were directly aligned with the task of extracting Petri net-compatible biological relations from literature, all datasets were generated specifically for this dissertation.

#### Fine-tuning datasets

The first fine-tuning datasets were created at an early stage of the project, when reasoning-capable open-source language models were only beginning to appear. At that time, the only openly available reasoning-oriented model was `Qwen/QwQ-32B` (`QwQ`) [129, 143], which was therefore selected as the primary data generation model. QwQ is a 32-billion-parameter model from the Qwen family, optimized for tasks requiring multi-step reasoning and the generation of intermediate reasoning traces before producing a final answer. These capabilities made it suitable for generating structured training examples that required interpretation, inference, and consistency across multiple output elements. For comparison purposes, the proprietary `gpt-4o` model was also employed.

Scientific articles were retrieved from PubMed using the Medical Subject Heading (MeSH) term “HDL”. From this collection, a random subset of articles was selected and divided into fixed-length fragments of approximately 500 tokens (around 375 words). To avoid constructing a dataset dominated by biologically irrelevant fragments, the `QwQ` model was first used as a heuristic filter. Chunks that did not contain meaningful biological relations were discarded. After filtering, approximately 90,000 text fragments were retained.

75:2026352137

Each fragment was then processed by `QwQ` and `gpt-4o` language models to generate structured extraction outputs. This procedure produced three distinct datasets:

- **Reasoning dataset** – containing the full reasoning trace generated by the `QwQ` model together with the final structured answer,
- **Only-answers dataset** – containing only the final structured outputs produced by `QwQ`, without intermediate reasoning,
- **GPT dataset** – containing structured outputs generated by the `gpt-4o` model.

The release of `DeepSeek R1` [27] represented a major advancement in open-source reasoning models. However, its computational requirements made large-scale dataset generation impractical, particularly when thousands of relation extraction examples were required.

Subsequently, `gpt-oss` open-source model family was released, including the `gpt-oss-120b` variant [97]. Owing to efficient quantization, this model could be executed on a single NVIDIA H100 GPU, making it suitable for large-scale annotation generation. Following the same procedure as for `QwQ-32B`, two additional datasets were constructed:

- outputs containing only final answers,
- outputs containing full reasoning traces.

#### Evaluation datasets

While `DeepSeek R1` was impractical for large-scale dataset creation, it was used to construct a high-quality evaluation benchmark. A separate subset of PubMed articles associated with the "HDL" MeSH term, not overlapping with the fine-tuning corpus, was randomly selected. From these articles, 1,000 fixed-length fragments were generated.

Each fragment was processed using `DeepSeek R1` to obtain high-confidence structured outputs. Because expert manual annotation at this scale was infeasible, these outputs served as a *pseudo-gold standard* against which other models were benchmarked. The use of a stronger model as a reference allowed consistent, repeatable quantitative evaluation while maintaining practical feasibility.

In addition, two manually curated Petri net-based models describing HDL-related biological processes were selected as true gold standards. This choice was motivated by the absence of publicly available benchmark datasets directly linking literature-derived knowledge to Petri net representations. These curated models enabled qualitative validation of biological correctness and structural plausibility of generated networks.

#### Evidence attribution dataset

In addition to extraction and evaluation corpora, an auxiliary *evidence attribution* dataset was constructed to provide provenance information for extracted biomedical relations. The objective of this dataset was to associate each predicted interaction with an explicit fragment of source text supporting its existence, thereby improving traceability and enabling expert verification.

To construct this dataset, the `gpt-oss-120b` model was prompted using the previously generated fine-tuning dataset. For each extracted relation, the model was instructed to identify and return the minimal text span directly supporting the interaction. This process produced approximately 90,000 relation - evidence pairs.

The resulting dataset enabled training and evaluation of attribution mechanisms within the `Watson` platform, allowing users not only to inspect predicted relations but also to immediately

76:6500561521

verify them against the original literature context. This functionality is essential in biomedical modeling, where interpretability and verifiability are required for domain expert acceptance.

### 7.2.3 Fine-Tuned Models

Based on the constructed datasets, six models were fine-tuned using the *Alignment Handbook* framework [132], an open-source collection of training recipes and tools for aligning LLMs. SFT was employed as the primary training paradigm. To enable efficient large-scale training within hardware constraints, the Liger kernel [55], a collection of optimized GPU kernels designed to reduce memory consumption and accelerate LLM training, was utilized for improved computational efficiency.

All training runs were executed on a single HPC node equipped with four NVIDIA H100 GPUs and 768 GB of system RAM. This configuration allowed stable multi-GPU training while maintaining reasonable throughput for large-scale datasets.

For relation extraction, the base model selected for fine-tuning was `Mistral-Nemo-Base-2407`, a general-purpose pretrained LLM from the Mistral-Nemo family. This model provided a favorable trade-off between parameter count, performance, and computational feasibility. For evidence attribution, a significantly smaller base model, `Qwen3-0.6B-Base` [128], a compact 0.6-billion-parameter model from the Qwen3 family, was chosen. The reduced model size was motivated by the narrower objective of attribution (text span localization rather than structured reasoning), as well as the goal of minimizing inference latency within the `Watson` platform.

All fine-tuned models were assigned the prefix *cannary-*. Models fine-tuned using outputs generated by `gpt-oss` were additionally marked with the suffix *-v2*, indicating the second iteration of dataset construction and training. The evidence attribution model was designated as `be-0.6b` (biological evidence, 0.6B parameters).

Rather than constructing a fully manually annotated benchmark for evidence attribution, an undertaking that would require extensive domain expert involvement, an automated evaluation protocol was adopted. A held-out set of 10,000 previously unseen examples generated by `gpt-oss-120b` was used as a reference dataset. Model outputs were compared against reference evidence spans using the `RapidFuzz` library, a high-performance Python library for approximate string matching, employing Indel similarity as the primary string-level evaluation metric.

This evaluation yielded an average Indel similarity score of 74.5 between text spans returned by the `be-0.6b` model and the reference outputs. While string-level similarity does not fully capture semantic equivalence, the results indicate that compact attribution models can reliably approximate high-capacity teacher models in evidence localization tasks.

### 7.2.4 Processing pipeline

As mentioned previously, the document processing pipeline, responsible for transforming unstructured scientific literature into Petri net–compatible structures, is executed asynchronously using the Celery task queue. An overview of the processing workflow is presented in Figure 7.2.

The pipeline begins with conversion of PDF documents into Markdown format using the `Docling` [126] library, an open-source toolkit for parsing and converting complex documents into structured, machine-readable representations. This step preserves the logical document structure, including section headers and figure captions, while removing layout-specific formatting that is irrelevant for language model processing.

Next, each document is segmented into fixed-length fragments of approximately 512 tokens using `LlamaIndex` [80], a framework for building language-model applications that provides tools

77:7578070828

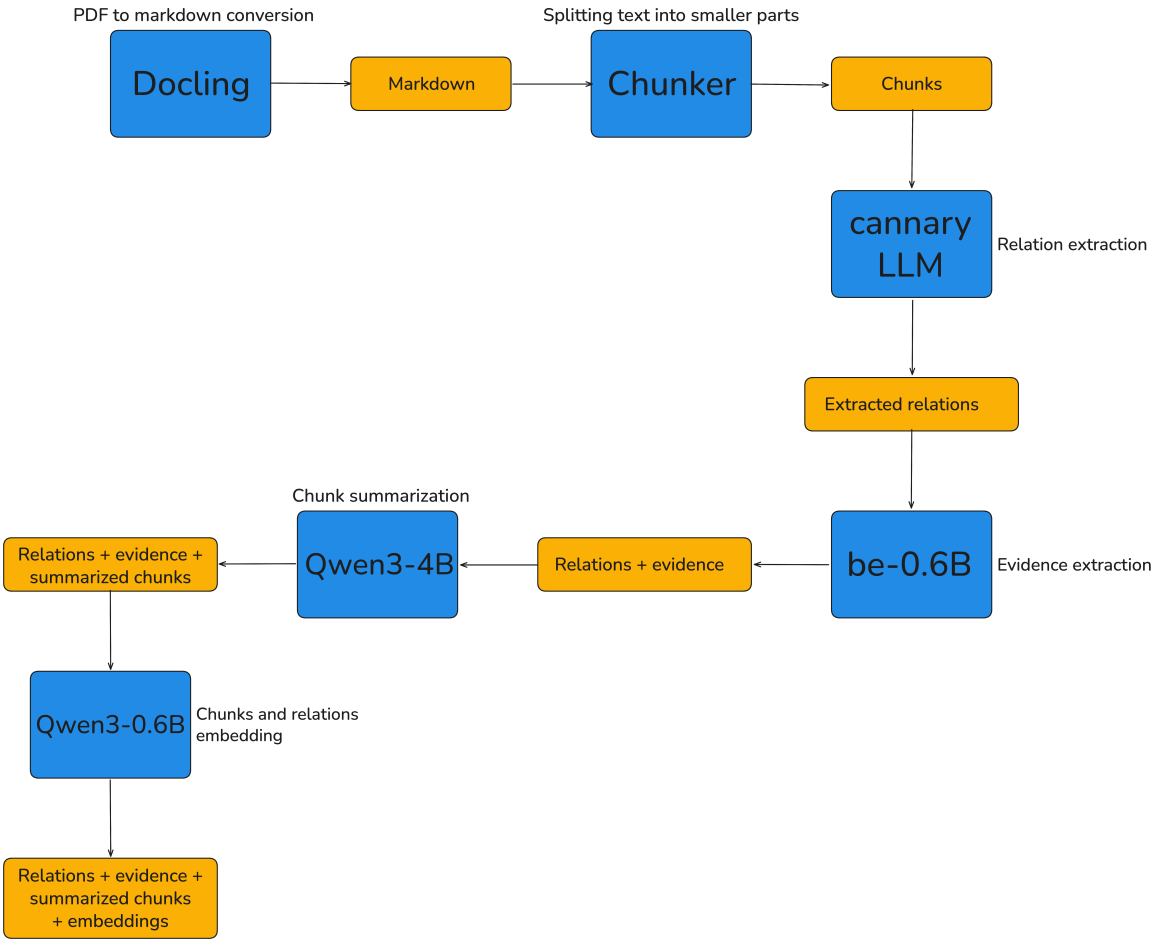



FIGURE 7.2: `Watson` asynchronous task processing overview.

for document ingestion, indexing, and text chunking. No overlap between fragments is introduced in order to avoid duplicate relation extraction and reduce downstream post-processing complexity. The selected fragment length provides a balance between contextual completeness and computational efficiency during inference.

Relation extraction is then performed using the `cannary` model. For each fragment, the model produces a structured representation of biological interactions in a single inference pass. Each extracted relation is represented as a tuple consisting of:

- substrates,
- modifiers,
- products,
- relation description.

This stage produces pairs of text fragments and their corresponding extracted relations.

In the subsequent step, the `be-0.6b` model identifies textual evidence supporting each extracted relation. The attribution step is essential for interpretability, allowing domain experts to directly verify whether a predicted interaction is supported by the source literature.

To improve usability, each fragment is then summarized using the `Qwen3-4B` model, a 4-billion-parameter language model from the Qwen3 family capable of efficient text understanding and

78:6156607916

generation. The generated summaries provide a concise description of fragment content and assist users in rapidly navigating large document collections.

Finally, each extracted relation is embedded into a high-dimensional vector representation using the `Qwen3-0.6B` embedding model. These embeddings enable semantic similarity search and relation retrieval within the `Watson` platform.

After processing, the system stores for each fragment in the database:

- extracted relations,
- supporting evidence spans,
- fragment summary,
- relation embeddings.

Together, these artifacts form a traceable intermediate representation linking the original literature to structured biological interactions suitable for subsequent Petri net construction.

### 7.2.5 Results

The evaluation was conducted using two complementary approaches. First, a manual expert assessment was performed in order to determine how effectively the `Watson` platform supports realistic modeling workflows. This evaluation focuses on biological plausibility and completeness of reconstructed pathways. Second, an automated LLM-as-a-judge evaluation was carried out, providing a scalable and reproducible quantitative assessment.

#### Manual evaluation

Manual inspection remains one of the most informative evaluation strategies for systems intended to assist domain experts. It allows identification of both correctly extracted interactions and biologically meaningful omissions, while simultaneously approximating real-world usage conditions of the platform. The primary limitation of this approach is its substantial time cost, which restricts evaluation throughput. Despite this, it provides an interpretable performance estimate.

Two independently developed, manually curated Petri net-based models were selected as references. The first model describes mechanisms associated with atherosclerosis [114], while the second represents HDL metabolism [38]. The former extends partially beyond the primary training domain and therefore allows assessment of generalization capability.

To assess the performance of the platform, two primary metrics were utilized: Recall@5 and precision. The evaluation of recall was conducted via the semantic search interface, where each transition from the reference PN model served as a query to retrieve the five most semantically similar relations identified by `Watson`. Each set of candidate relations was then manually compared against the corresponding reference transition, with particular attention paid to the alignment of substrates, modifiers, and products.

To quantify the degree of correspondence, a three-tier scoring system was applied to the retrieved results:

- A score of **1** was assigned when an extracted relation provided a close or exact match to the reference transition, accurately capturing the involved biological entities and their roles.
- A score of **0.5** was given if the core interaction was correctly identified but the representation was incomplete - for instance, due to missing modifiers or products - provided that these

79:4494281746

elements remained inferable from the surrounding textual context of the original source fragment.

- A score of **0** indicated that no relevant relation was retrieved or that the candidates lacked the essential components required to represent the target biological interaction.

Furthermore, a precision analysis was performed on the relations extracted from the HDL metabolism and atherosclerosis datasets to estimate the frequency of accurate extractions relative to potential model hallucinations. Given the high volume of data - approximately 3,000 relations per network - a comprehensive manual audit of all entries was not feasible. Consequently, a randomized sampling strategy was chosen, whereby 10% of the relations from each source document were selected. These samples were subsequently subjected to expert review to verify that each interaction was both biologically plausible and explicitly supported by the underlying source text.

The inclusion of a precision metric serves as a necessary complement to the Recall@5 assessment. During the construction of a PN model, it is observed that not every biologically valid relation extracted from the literature is inherently relevant to the specific process being modeled. Numerous interactions identified by the system may describe accurate biological phenomena that occur within a broader physiological context or in related pathways not explicitly defined in the reference PN model. While such relations may not contribute directly to the target network's structure, they remain valid from a biological standpoint. Therefore, measuring precision facilitates a more nuanced evaluation by distinguishing these out-of-scope but correct extractions from actual model hallucinations or erroneous interactions. This approach allows for a reliable estimate of the system's ability to extract grounded biological knowledge while quantifying the frequency of incorrect outputs within the generated networks.

#### Atherosclerosis

The first evaluation used a curated Petri net describing biological processes involved in atherosclerotic plaque formation [114]. The model consists of 42 transitions representing interactions among lipoproteins, immune cells, and vascular wall components.

`Watson` identified 10 transitions with full correspondence to the reference model (score **1**) and additional 10 transitions with partial agreement (score **0.5**). In the partially recovered cases, the direction and biological role of the interaction were correct, although certain details were simplified or omitted. The remaining 22 transitions were not detected (score **0**), resulting in a Recall@5 of 20 out of 42 transitions (48%).

The reconstructed network is presented in Figure 7.3. The missing transitions, marked by yellow thunderbolt symbols, occur both as isolated reactions and as groups forming coherent subprocesses. Because atherosclerosis was not part of the primary target domain, these omissions likely reflect domain shift rather than purely extraction failure.

In addition to the recall assessment, the precision of the extracted relations for the atherosclerosis corpus was evaluated. This analysis involved a manual verification of a randomized sample comprising 10% of the relations from each of the 13 input files. Of the 282 relations subjected to expert review, 241 were confirmed to be biologically accurate and explicitly supported by the source literature, while 41 were identified as incorrect or representative of model hallucinations. These findings yield an overall precision of 85.46%, demonstrating that the substantial majority of interactions identified by the system represent valid biological knowledge. Such results indicate that even when extracted relations are not directly incorporated into the specific Petri net model under construction, they remain grounded in the provided textual evidence.

80:5933174942

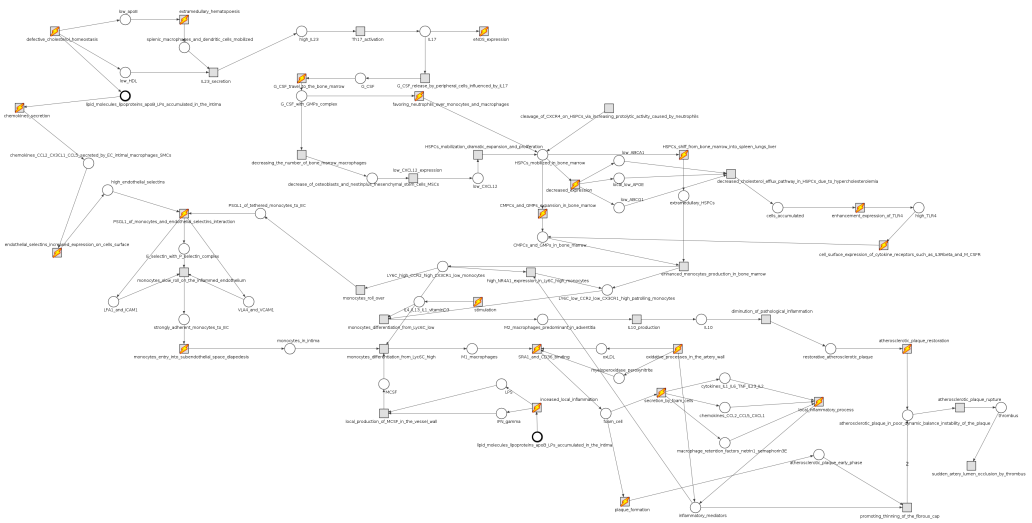


FIGURE 7.3: Reference atherosclerosis Petri net [114] with transitions not recovered by `Watson` marked by yellow thunderbolt symbols. Five of the 22 missing transitions appear as isolated reactions, whereas the others form clustered subprocesses.

### HDL metabolism

A second evaluation was performed using a curated Petri net describing HDL metabolism [38]. The reference network, shown in Figure 7.4, contains 122 transitions, of which 113 could be evaluated due to availability of supporting literature sources.

`Watson` reconstructed 24 transitions exactly (score **1**) and 32 transitions partially (score **0.5**). In the latter category, the extracted relations preserved the correct causal structure but lacked specific contextual or biochemical detail. The remaining 57 transitions were not identified (score **0**).

Among the unrecovered transitions, 37 correspond to biologically essential reactions, whereas 20 represent implicit input processes that are typically assumed in pathway models but rarely stated explicitly in textual sources. Under a strict evaluation requiring explicit textual evidence for all reactions, the system correctly identified 56 of 113 transitions (49%). When excluding implicit input transitions that cannot reasonably be expected to appear in literature descriptions, recall increases to 56 of 93 transitions (60%).

The precision of the relations extracted from the HDL metabolism literature corpus was also evaluated. Across 19 processed scientific articles, a random sample comprising 10% of the relations from each source document was manually verified. Of the 291 relations subjected to expert review, 241 were confirmed to be biologically accurate and supported by the source text, while 50 were identified as incorrect or represented hallucinated interactions. This yields an overall precision of 82.82%, suggesting that the majority of the relations identified by `Watson` represent valid biological knowledge. These results further demonstrate that the system consistently extracts grounded interactions, even if a portion of the retrieved data does not directly correspond to the specific transitions included in the curated Petri net model.

Missing transitions are highlighted with yellow thunderbolt symbols. Most omissions form contiguous multi-step subprocesses rather than isolated reactions. This pattern suggests that some details may be communicated through diagrams or assumed domain knowledge rather than stated explicitly in the text. However, this interpretation remains a hypothesis and requires further investigation, including a systematic comparison of textual descriptions, accompanying figures, and

81:4784558431

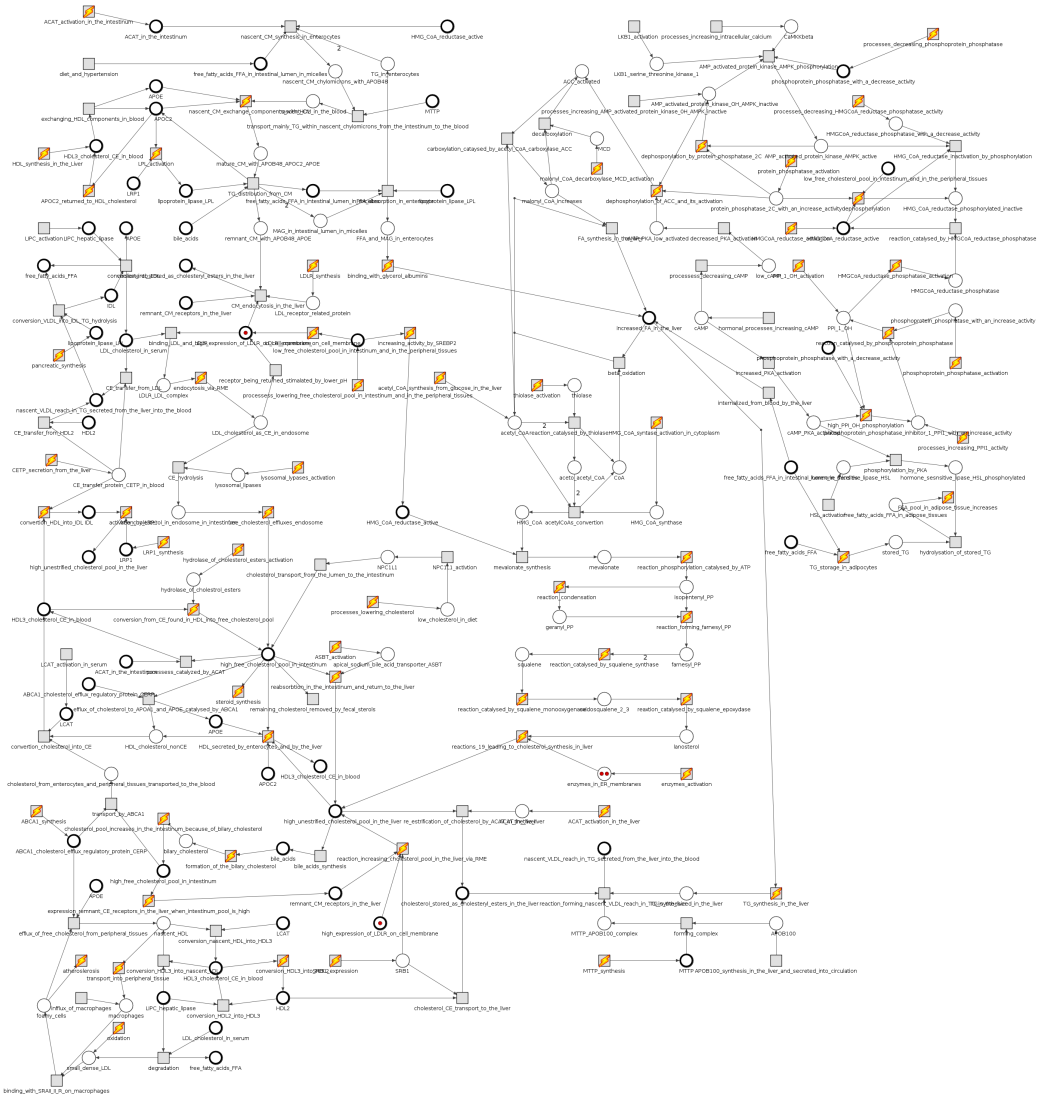


FIGURE 7.4: Reference HDL metabolism Petri net [38] with transitions not recovered by `Watson` indicated by yellow thunderbolt symbols. Missing elements frequently occur as multi-step subprocesses rather than isolated reactions.

domain-specific reporting conventions.

### Automatic evaluation

To complement the manual assessment, an automatic evaluation protocol was introduced. The primary advantage of this approach is scalability: once defined, the benchmark enables rapid comparison of multiple model variants and significantly accelerates iterative development.

The previously described evaluation dataset was used as the reference corpus. Since the dataset itself was generated using a strong teacher model, a small subset of samples (50 instances, representing 5% of the corpus) was manually inspected to estimate the reliability of the pseudo-gold standard produced by `DeepSeek R1`. The inspection showed that 34 cases conveyed the intended biological meaning with only minor imperfections, 10 contained partially missing details, and 6 differed substantially. This confirmed that the dataset was sufficiently reliable for comparative benchmarking, while still acknowledging that it does not represent a perfect ground truth.

Next, a diverse selection of open-source LLMs was subsequently applied to the same corpus to

82:2517129141

extract relations in the Petri net-compatible format. The resulting outputs were then evaluated by two independent judge models: `gpt-oss-120b` [97] and `Qwen3-8B` [128]. These judges compared model predictions against the pseudo-gold reference using a pairwise similarity scoring scheme, guided by a specialized scoring template, presented in Table 7.1, to ensure consistency across evaluations. The aggregated results of this automated assessment are detailed in Table 7.2 and Table 7.3.

TABLE 7.1: Scoring rubric utilized by the LLM-as-a-judge models (`gpt-oss-120b` and `Qwen3-8B`) to quantify the semantic correspondence between model-generated relations and the pseudo-gold standard.

| **Score Range** | **Description** |
|---|---|
| 90–100 | Almost perfect match; all key relations captured with accurate details and high semantic fidelity. No major errors. |
| 75–89 | Strong match; most relations correct with minor omissions, vague modifiers, or harmless paraphrasing. |
| 60–74 | Moderate match; several correct relations, but some are missing, merged improperly, or have unclear elements. |
| 40–59 | Weak match; many errors, missing or wrong relations, or confusion in roles (substrate vs product). |
| 20–39 | Poor match; few correct elements, most relations misrepresented or incorrect. |
| 0–19 | Completely incorrect; no meaningful overlap with the gold data. |

TABLE 7.2: LLM-as-judge evaluation using **gpt-oss-120b**, reporting mean scores and 95% CIs. Best performance is in **bold**, second-best is underlined. Models prefixed with `cannary` are trained in this study.

| **Model** | **Avg. score** | **95% CI** |
|---|---|---|
| DeepSeek-V3 | 55.75 | 1.35 |
| gpt-oss-120b | <u>56.50</u> | 1.52 |
| Qwen3-235B | 48.66 | 1.50 |
| Qwen2.5-32B | 47.15 | 1.60 |
| Qwen3-8B | 31.61 | 1.72 |
| `cannary-oa-v2` | **57.42** | 1.33 |
| `cannary-reasoning-v2` | 55.33 | 1.49 |
| `cannary-reasoning` | 43.31 | 1.86 |
| `cannary-oa` | 42.75 | 1.74 |
| `cannary-gpt` | 37.44 | 1.58 |
| Mistral-Nemo-Instruct-2407 | 35.96 | 1.44 |

Across evaluations performed by `gpt-oss-120b`, the 12B-parameter model `cannary-oa-v2` achieved the highest mean score (57.42), surpassing multiple substantially larger models. Under evaluation by `Qwen3-8B`, the same model ranked second overall (64.54), closely matching the top-performing architecture.

Importantly, both models developed in this work, `cannary-oa-v2` and `cannary-reasoning-v2`, consistently achieved competitive performance - moderate semantic relations match - across both judges. This result suggests that domain-targeted supervised fine-tuning and curated synthetic datasets can compensate for substantially smaller parameter counts. Consequently, the proposed training methodology enables efficient large-scale extraction of Petri net relations without relying on the largest available general-purpose models.

83:1574909220

TABLE 7.3: LLM-as-judge evaluation using **Qwen3-8B**, reporting mean scores and 95% CIs. Best performance is in **bold**, second-best is underlined. Models prefixed with `cannary` are trained in this study.

| Model | Avg. score | 95% CI |
|---|---|---|
| DeepSeek-V3 | **64.60** | 0.97 |
| gpt-oss-120b | 63.40 | 1.31 |
| Qwen3-235B | 62.29 | 1.18 |
| Qwen2.5-32B | 57.24 | 1.57 |
| Qwen3-8B | 41.40 | 2.06 |
| `cannary-oa-v2` | 64.54 | 0.93 |
| `cannary-reasoning-v2` | 62.17 | 1.29 |
| `cannary-reasoning` | 53.08 | 1.92 |
| `cannary-oa` | 50.44 | 1.79 |
| `cannary-gpt` | 49.65 | 1.72 |
| Mistral-Nemo-Instruct-2407 | 52.12 | 1.38 |

**Summary and implications**

In this section, a framework for extracting biomedical relations from scientific literature and translating them into Petri net representations was presented. The evaluation strategy combined detailed manual inspection with large-scale automatic benchmarking in order to assess both practical utility and methodological limitations.

The manual evaluations reveal several important characteristics of text-derived PN reconstruction. First, the system is capable of recovering a substantial portion of biologically meaningful interactions directly from literature, particularly when those interactions are explicitly described in textual form. In both the atherosclerosis and HDL metabolism case studies, a considerable fraction of transitions was reconstructed either exactly or with preserved causal semantics.

However, the experiments also highlight potential intrinsic limitations of text-only extraction. A substantial number of missing transitions correspond to details that are not explicitly described in the prose. One possible explanation is that, in biomedical publications, some essential steps are conveyed primarily through schematic diagrams or treated as assumed background knowledge rather than stated verbatim. Because the current pipeline intentionally processes only textual content and excludes figures, such information may remain inaccessible. This interpretation suggests that the observed limitation is at least partly methodological rather than purely model-specific; however, further studies comparing textual content with accompanying figures are required to confirm this hypothesis.

Additional constraints arise from architectural design choices. The preprocessing strategy segments documents into fixed-length fragments to ensure computational tractability. While this enables scalable inference, it may disrupt cross-paragraph dependencies and obscure relations that span multiple textual segments. Furthermore, in order to preserve transparency and expert control, the system deliberately avoids aggressive automatic graph merging across fragments or documents. Although this design enhances interpretability, it results in a proliferation of partially overlapping candidate relations, thereby increasing the burden of downstream expert curation.

Model capacity and domain specificity also influence performance. The primary relation extraction model contains 12 billion parameters and was fine-tuned predominantly on HDL-related material. This targeted training strategy improves performance within the intended domain while maintaining deployability on accessible hardware. However, it may reduce the model's ability to generalize to interactions that are weakly represented or absent in the training corpus.

The automatic LLM-as-a-judge evaluation provides complementary evidence. Domain-adapted

84:1151646524

models achieved performance comparable to, and in some cases exceeding, substantially larger general-purpose architectures. This finding suggests that carefully curated synthetic datasets and supervised fine-tuning can compensate for differences in parameter count. At the same time, reliance on pseudo-gold standards introduces unavoidable uncertainty, reinforcing the importance of combining automated metrics with expert-driven validation. However, it should be noted that the manual evaluation of recall and precision was conducted on a different dataset than the one used for the automatic assessment. As a result, the observed differences in performance between manual and automatic evaluation cannot be directly interpreted as inconsistencies between the two methodologies. Instead, they reflect variations in dataset characteristics, including differences in annotation granularity, textual complexity, and the degree of alignment with the target representation. Nevertheless, the fact that manual evaluation yielded more favorable results suggests that automated benchmarks may not fully capture domain-specific correctness and contextual validity. This observation underscores a broader limitation of LLM-based evaluation frameworks, which may be sensitive to dataset composition and evaluation design. Consequently, automated evaluation should be interpreted with caution and complemented by expert-driven analysis to obtain a more comprehensive and reliable assessment of system performance.

Taken together, the results indicate that fully automated reconstruction of complete biological pathways from text alone is neither currently feasible nor, in practice, a desirable objective. Human-curated models are expected to achieve high precision, and a fully automated pipeline would likely introduce sufficient uncertainty to require expert verification of nearly every extracted relation, thereby diminishing its practical utility. In this context, the goal of the proposed framework is not to replace expert-driven modeling, but to support it. The system reliably generates a high-volume set of biologically plausible candidate relations, substantially accelerating the early stages of Petri net construction while maintaining traceability to source literature. This design offers a balanced approach, in which automated methods assist domain experts without undermining confidence in the resulting models, ultimately facilitating a more efficient yet still rigorous modeling workflow.

The limitations identified in this section motivate the refinements introduced in the following chapter, which aim to improve structural consistency, reduce redundancy, and enhance the practical usability of the platform.

## 7.3 Watson – Improvements

The initial implementation of `Watson` revealed several methodological and practical limitations affecting both extraction quality and usability of the generated biological models. The purpose of this chapter is not to introduce a new foundation model, but to redesign the extraction procedure itself in order to reduce systematic errors originating from the processing strategy.

The following issues were identified:

- **Document segmentation.** The original pipeline relied on fixed-size document chunking to enable tractable inference. Although computationally efficient, this approach may disrupt semantic continuity and prevent the model from identifying relations spanning multiple paragraphs or sections.
- **Apparent false positives due to relation granularity.** The extraction models often produced a larger number of relations not because of increased hallucination, but due to a finer decomposition of interactions. In particular, compound statements such as “A interacts with B, C, and D” were systematically expanded into multiple pairwise relations (e.g., “A

85:5915471136

interacts with B", "A interacts with C", "A interacts with D"). As a result, the higher relation count reflects differences in representation granularity rather than a true increase in biologically unsupported interactions. This effect was further influenced by the segmentation strategy, where limited local context encouraged the model to express relations in a more explicit, atomized form.

- **Processing separation.** Relation detection and evidence attribution were executed as independent stages. Consequently, the evidence spans often differed from those that implicitly justified the extracted relation, reducing traceability and interpretability.
- **Incomplete relation attributes.** The initial implementation omitted biological qualifiers such as inhibition directionality or organism specificity required for correct Petri net semantics.
- **Evaluation mismatch.** The original automatic evaluation compared sets of extracted relations, whereas the practical use case of the platform involves incremental expert curation. Therefore, the metric did not accurately reflect real workflow utility.

The improvements presented in this chapter focus on modifying the inference procedure rather than increasing model scale. Two complementary strategies are introduced. First, Subsection 7.3.1 proposes *Iterative Information Extraction*, in which the document is treated as a persistent structured memory during processing. Second, Subsection 7.3.2 introduces a deliberative extractor - council mechanism that reduces unsupported relations through model-based verification. Finally, Subsection 7.3.3 presents an evaluation of the proposed methods using a newly constructed benchmark tailored to the intended usage scenario of the platform.

### 7.3.1 Iterative Information Extraction

The core idea of Iterative Information Extraction (IIE) is to treat the processed document as an evolving state representation rather than a static prompt. Instead of extracting all relations in a single pass, the system extracts one relation at a time, explicitly marks its supporting evidence in the text, and then restarts inference using the modified document.

In effect, the document becomes a persistent external memory that records extraction progress. Each iteration is independent from the model's previous reasoning, but dependent on the accumulated annotations present in the text.

The processing workflow is illustrated in Figure 7.5.

The procedure consists of the following steps:

1. **Provide document.** The first iteration receives the original document without annotations.
2. **Extract relation.** The model, `gpt-oss-120b`, analyzes the full document and proposes a single relation together with supporting evidence span.
3. **Mark evidence.** The identified span is enclosed within an `<extracted>` XML tag. When exact matching fails, approximate span localization is performed using the `rapidfuzz` library.
4. **Repeat.** Steps 2–3 are repeated until the model determines that no additional relations can be extracted from the input and outputs the termination signal "Extraction complete".

This design converts the extraction process from a stateless prediction task into a monotonic annotation procedure. The model no longer needs to implicitly remember previously generated

86:8420110406

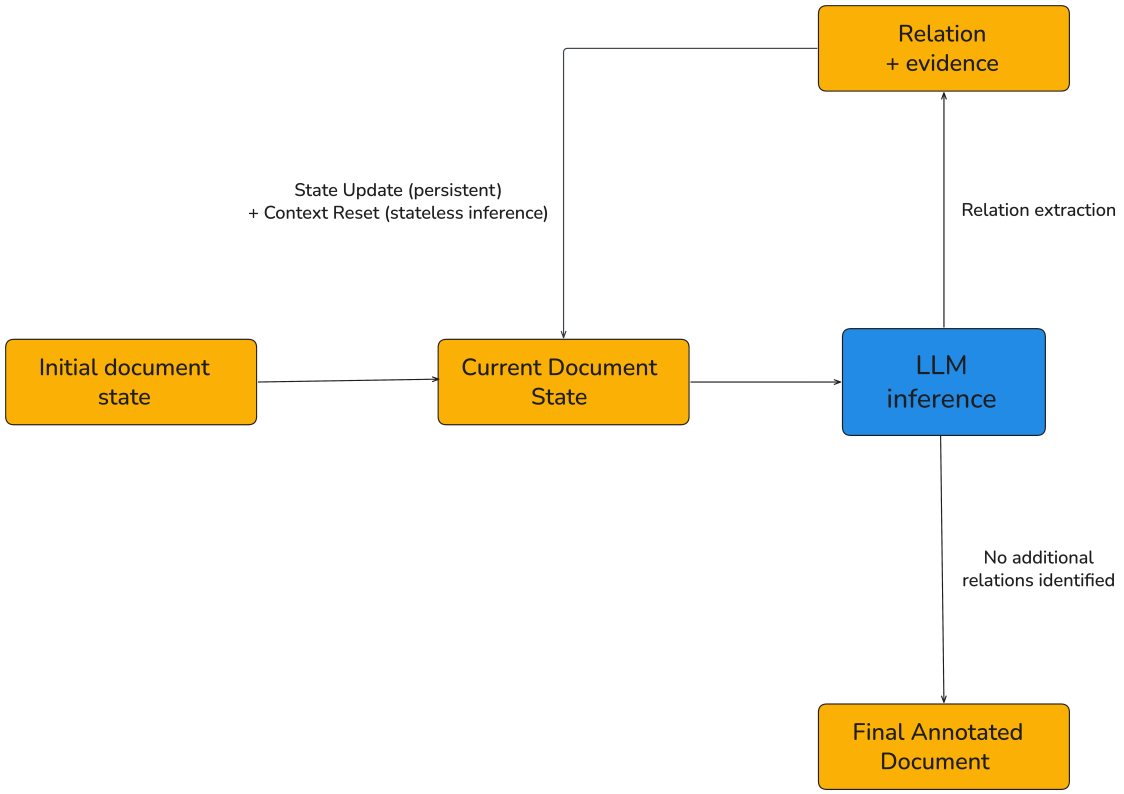



FIGURE 7.5: Iterative Information Extraction formulated as a monotonic state-transition process. The language model performs stateless inference at each iteration, while extraction progress is preserved through persistent evidence annotations embedded in the document state.

relations. Instead, prior decisions are embedded directly in the input context. The reset between iterations prevents error accumulation in hidden reasoning traces while preserving progress through explicit text markings. Furthermore, the procedure avoids progressively growing reasoning histories, thereby preventing the formation of excessively long input sequences and maintaining stable computational requirements across iterations.

**Expected effects.**

- **Improved global context.** The entire document remains available during each inference step, eliminating boundary effects caused by chunking and enabling detection of cross-paragraph relations.
- **Reduced speculative relations.** Because already-supported statements are marked, the model is discouraged from repeatedly generating similar relations and instead searches for genuinely new evidence.
- **Integrated evidence attribution.** Evidence identification becomes inseparable from extraction, ensuring that each produced relation is grounded in a specific textual fragment.
- **Deterministic progress.** The document evolves monotonically; previously validated information cannot be implicitly overwritten by later reasoning.

The quantitative evaluation of this approach is performed jointly with the agentic variant described in the following subsection. During preliminary experiments it was observed that the model occasionally attempted to annotate previously marked spans with additional `<extracted>` tags.

87:8007038618

This behavior indicated limitations of single-model self-regulation and motivated the introduction of an explicit verification mechanism in the subsequent agentic formulation.

**Vector database extension** An extension was investigated in which newly extracted relations were compared against previously stored relations in a vector database. The intention was to allow relation refinement rather than duplication, effectively introducing retrieval-guided behavior.

However, this approach required complex synchronization mechanisms to prevent concurrent updates in parallel execution settings. Empirical analysis on a sample of 100 documents showed that approximately 95% of extracted relations were novel, while only 5% represented potential updates to existing relations. Given the marginal redundancy reduction relative to implementation complexity, the mechanism was not integrated into the final system.

Consequently, the final Iterative Information Extraction method remains a controlled sequential workflow rather than a retrieval-driven adaptive process.

### 7.3.2 Extractor-critic-consensus

The purpose of Extractor-Critic-Consensus (ECC) extension is to verify each extracted relation before it becomes part of the persistent document state. Instead of immediately accepting the output of a single extraction model, the system introduces a deliberative evaluation stage performed by a council of one to $n$ `gpt-oss-120b` language models.

After the extractor proposes a relation together with supporting evidence, each council member independently evaluates the proposal using access to the full document. The models do not share intermediate reasoning and operate in parallel. A majority voting scheme is then applied to determine whether the relation satisfies the extraction criteria.

If the relation is accepted, the supporting evidence is marked in the document and the processing loop restarts from a clean context. If the relation is rejected, the extractor receives the rejection decision together with the critique provided by the council members and attempts to revise the extraction. The revised proposal is evaluated again by the council. If the revision is accepted, the relation is recorded; otherwise the iteration is discarded and the loop restarts without modifying the document.

The processing flow of the proposed procedure is shown in Figure 7.6.

Unlike IIE, where the model implicitly self-regulates extraction quality, the present design introduces an explicit verification stage. The council jointly determines both correctness of the proposed relation and whether further extraction attempts should continue. Consequently, the control flow is partially governed by model judgments rather than solely by predefined program logic, forming a deliberative multi-model inference procedure.

Preliminary experiments with the IIE approach revealed that a single model occasionally annotated already processed spans or produced relations that formally satisfied the output schema but were weakly supported by the text. The council-based evaluation enforces adherence to extraction constraints by requiring inter-model agreement before modifying the document state. The mechanism therefore prioritizes reliability and traceability over raw extraction throughput.

Although the ECC design is expected to improve precision and adherence to extraction constraints, it introduces a measurable increase in inference time. While document-level processing and council evaluations are executed in parallel, the additional verification and potential refinement stages necessarily extend the overall processing latency.

This computational overhead can be partially mitigated through horizontal scaling. In the experimental setup, individual council members were deployed on separate GPU devices and requests

88:4014161262

were distributed using a round-robin scheduling strategy. Such parallelization reduces wall-clock time but does not eliminate the intrinsic increase in total compute consumption associated with multi-model deliberation.

It should be noted that all experiments were conducted in a high-performance computing environment. In resource-constrained real-world deployments, the achievable degree of parallelization may be limited, and therefore the practical trade-off between extraction reliability and computational cost must be considered.

### 7.3.3 Evaluation

To enable comparison with the original `cannary` models, a new benchmark was constructed within the HDL domain. A set of 100 randomly selected PubMed full-text articles was collected, ensuring that none of the documents appeared in the original training corpus.

The following systems were evaluated:

- `cannary-reasoning-v2`,
- IIE using `gpt-oss-120b`,
- `Kimi-K2.5` [127],
- `DeepSeek-V3.2-Speciale` [28],
- ECC with 1, 3, 5, and 7 judges (all judges using `gpt-oss-120b`).

**Pseudo gold standard construction.** The `Kimi-K2.5` and `DeepSeek-V3.2-Speciale` models were used to construct pseudo gold standards. The motivation was to compare medium-scale domain-adapted models with large general-purpose open models and to avoid bias toward any single architecture.

Both models were applied to the HDL metabolism corpus used for Petri net construction. All extracted relations were manually curated and compared against the reference pathway representation. This resulted in a precision of 50% for `Kimi-K2.5` and 48% for `DeepSeek-V3.2-Speciale`. Since the difference was marginal, both were retained as independent pseudo gold standards in subsequent experiments.

**Evaluation scenarios.** Two complementary evaluation strategies were applied, using already established scoring system 7.1:

1. **Scenario 1 - global comparison.** All pseudo gold standard relations were compared with the complete set of extracted relations for each of the 100 documents.
2. **Scenario 2 - retrieval-oriented comparison.** For each pseudo gold standard relation, the five most semantically similar relations produced from the same document were selected and evaluated.

Scenario 1 measures aggregate extraction coverage, whereas Scenario 2 approximates practical platform usage in which a user inspects a limited set of candidate relations. Due to the substantial cost of expert annotation, manual evaluation was performed only for the most promising configuration identified in the automatic experiments, namely the ECC pipeline with one judge. In this setting, evaluation was restricted to the Recall@5 metric, as reliable estimation of precision would require exhaustive verification of whole texts, which was not feasible at scale.

89:2422500716

The results for Scenario 1 using `Kimi-K2.5` as pseudo gold standard are shown in Table 7.4; Scenario 1 using `DeepSeek-V3.2-Speciale` in Table 7.5; Scenario 2 using `Kimi-K2.5` in Table 7.6; and Scenario 2 using `DeepSeek-V3.2-Speciale` in Table 7.7. In all comparisons the evaluation judge was `gpt-oss-120b`.

TABLE 7.4: Scenario 1 (global comparison): agreement scores against the `Kimi-K2.5` pseudo gold standard. For each of 100 PubMed full-text articles, the complete set of relations extracted by a given system was evaluated by `gpt-oss-120b` against the reference set. Reported are the mean score, median score, and the 95% confidence interval of the mean.

| Model | Avg. score | Median score | 95% CI |
|---|---|---|---|
| Iterative Information Extraction | 45.08 | 48.00 | 4.34 |
| DeepSeek-V3.2-Speciale | 46.24 | 48.00 | 4.86 |
| Extractor-critic-consensus - 1 judge | 49.16 | 55.00 | 4.22 |
| Extractor-critic-consensus - 3 judges | 53.00 | 65.00 | 4.48 |
| Extractor-critic-consensus - 5 judges | 52.24 | 57.00 | 4.34 |
| Extractor-critic-consensus - 7 judges | 52.98 | 58.00 | 4.00 |
| cannary-reasoning-v2 | 51.71 | 58.00 | 4.51 |

TABLE 7.5: Scenario 1 (global comparison): agreement scores against the `DeepSeek-V3.2-Speciale` pseudo gold standard. For each of 100 PubMed full-text articles, the complete set of relations extracted by a given system was evaluated by `gpt-oss-120b` against the reference set. Reported are the mean score, median score, and the 95% confidence interval of the mean.

| Model | Avg. score | Median score | 95% CI |
|---|---|---|---|
| Iterative Information Extraction | 42.53 | 48.00 | 4.76 |
| Kimi-K2.5 | 49.98 | 52.00 | 4.91 |
| Extractor-critic-consensus - 1 judge | 47.11 | 52.00 | 4.98 |
| Extractor-critic-consensus - 3 judges | 46.34 | 52.00 | 4.80 |
| Extractor-critic-consensus - 5 judges | 45.97 | 52.00 | 4.94 |
| Extractor-critic-consensus - 7 judges | 46.61 | 48.00 | 4.71 |
| cannary-reasoning-v2 | 49.56 | 55.00 | 4.69 |

TABLE 7.6: Scenario 2 (retrieval-oriented comparison): agreement scores against the `Kimi-K2.5` pseudo gold standard. For each reference relation, the five most semantically similar candidate relations produced from the same document were selected and judged by `gpt-oss-120b`. Reported are the mean score, median score, and the 95% confidence interval of the mean across all evaluated candidates.

| Model | Avg. score | Median score | 95% CI |
|---|---|---|---|
| Iterative Information Extraction | 49.26 | 66.00 | 1.82 |
| DeepSeek-V3.2-Speciale | 45.21 | 55.00 | 1.86 |
| Extractor-critic-consensus - 1 judge | 56.61 | 68.00 | 1.70 |
| Extractor-critic-consensus - 3 judges | 57.62 | 68.00 | 1.65 |
| Extractor-critic-consensus - 5 judges | 57.91 | 68.00 | 1.64 |
| Extractor-critic-consensus - 7 judges | 59.89 | 68.00 | 1.57 |
| cannary-reasoning-v2 | 62.87 | 68.00 | 1.24 |

The experimental results reveal several consistent patterns across both evaluation scenarios and both pseudo gold standards.

First, the domain-adapted `cannary-reasoning-v2` model achieved competitive performance despite a substantially smaller parameter count and the absence of reinforcement learning–based post-training. In Scenario 1 (global comparison), its average score remained comparable to or higher than the general-purpose models and the proposed extraction procedures (Tables 7.4 and 7.5).

90:5324357366

TABLE 7.7: Scenario 2 (retrieval-oriented comparison): agreement scores against the `DeepSeek-V3.2-Speciale` pseudo gold standard. For each reference relation, the five most semantically similar candidate relations produced from the same document were selected and judged by `gpt-oss-120b`. Reported are the mean score, median score, and the 95% confidence interval of the mean across all evaluated candidates.

| Model | Avg. score | Median score | 95% CI |
|---|---|---|---|
| Iterative Information Extraction | 48.81 | 62.00 | 2.21 |
| Kimi-K2.5 | 54.50 | 68.00 | 2.12 |
| Extractor-critic-consensus - 1 judge | 53.68 | 68.00 | 2.10 |
| Extractor-critic-consensus - 3 judges | 54.65 | 68.00 | 2.10 |
| Extractor-critic-consensus - 5 judges | 55.39 | 68.00 | 2.04 |
| Extractor-critic-consensus - 7 judges | 55.81 | 68.00 | 2.05 |
| cannary-reasoning-v2 | 58.27 | 68.00 | 1.72 |

In Scenario 2 (retrieval-oriented comparison), the model obtained the highest average scores under both pseudo gold standards (Tables 7.6 and 7.7). This indicates that domain specialization can compensate for scale in structured biomedical relation extraction, particularly when the task requires adherence to well-defined biological semantics rather than broad linguistic generalization.

Second, both proposed procedures - IIE and ECC - produced substantially fewer relations than all baseline models (Table 7.8). The reduction was particularly pronounced relative to `cannary-reasoning-v2`, which generated approximately three times more relations than the largest ECC configuration. The decrease in output volume therefore reflects a shift in extraction behavior: instead of maximizing recall through exhaustive candidate generation, the modified procedures emphasize conservative acceptance criteria. Consequently, the methods act as precision-oriented filters rather than high-recall extractors.

TABLE 7.8: Extraction yield on the 100-document benchmark: total number of distinct relations produced by each system under the respective extraction procedure. The table illustrates the trade-off between conservative, verification-oriented pipelines (IIE/ECC) and high-recall extraction (`cannary-reasoning-v2`).

| Model | Number of retrieved relations |
|---|---|
| Iterative Information Extraction | 1295 |
| DeepSeek-V3.2-Speciale | 785 |
| Kimi-K2.5 | 1146 |
| Extractor-critic-consensus - 1 judge | 1753 |
| Extractor-critic-consensus - 3 judges | 1862 |
| Extractor-critic-consensus - 5 judges | 1936 |
| Extractor-critic-consensus - 7 judges | 2081 |
| cannary-reasoning-v2 | 5943 |

Third, the ECC pipeline consistently improved over single-model IIE in all evaluation configurations. In Scenario 1 using `Kimi-K2.5` as reference, the three-judge ECC variant achieved the highest agreement among the proposed methods (Table 7.4). In Scenario 2, performance monotonically increased with the number of council members, with the seven-judge configuration achieving the best results among the ECC variants (Tables 7.6 and 7.7). This pattern suggests that independent verification reduces acceptance of weakly supported relations and stabilizes extraction decisions.

Fourth, the improvement associated with increasing the number of judges was substantially stronger in Scenario 2 than in Scenario 1. Because Scenario 2 evaluates only the top five semantically closest relations, it approximates expert-assisted curation in which users inspect a limited candidate set. The stronger improvement therefore indicates that ECC primarily improves ranking

91:9165722035

reliability rather than global recall.

Finally, the two pseudo gold standards produced asymmetric agreement patterns. When one model served as the reference, the other frequently showed lower agreement than smaller domain-adapted systems. This suggests that both reference sets encode partially incompatible interpretations of relations. Consequently, absolute scores should be interpreted cautiously and primarily as relative indicators within each evaluation configuration rather than as objective correctness measures.

Complementary manual evaluation was conducted on two curated pathway reconstruction tasks, namely the atherosclerosis and HDL metabolism Petri nets. Using `DeepSeek-V3.2-Speciale` as the primary extractor within the ECC framework, the system achieved Recall@5 scores of 47.62% and 47.79%, respectively. When `Kimi-K2.5` was used as the extractor model, the results differed between datasets, yielding Recall@5 values of 38.1% for atherosclerosis and 50.44% for HDL metabolism. These findings suggest that, under expert evaluation, the overall retrieval effectiveness of the ECC pipeline remains competitive, while also being sensitive to both dataset characteristics and the choice of underlying extraction model.

### 7.3.4 Summary

This section addressed limitations identified in the initial `Watson` pipeline by redesigning the inference procedure. In particular, the proposed changes target (i) context fragmentation introduced by fixed-size chunking, (ii) over-generation of biologically plausible but textually unsupported relations, and (iii) the separation of relation prediction from evidence attribution, which undermined traceability and downstream usability.

IIE reframes relation extraction as a monotonic annotation process in which the full document remains available at every step and accepted evidence is explicitly marked in-place. This design converts the document into a persistent external memory, thereby avoiding boundary effects and reducing dependence on implicit conversational state. As a consequence, IIE primarily changes the operating point of the system: extraction becomes more conservative and more interpretable, producing fewer candidates that are directly grounded in localized evidence spans.

ECC extension further improves reliability by introducing explicit verification through independent model judgments. Across both evaluation scenarios and both pseudo gold standards, ECC outperformed single-model IIE, with the most consistent gains observed in the retrieval-oriented setting that approximates expert-in-the-loop curation. Increasing the number of judges generally improved performance, indicating that inter-model agreement serves as an effective safeguard against weakly supported relations, albeit at increased computational cost.

At the same time, the relation-count analysis highlights a structural trade-off. Compared to `cannary-reasoning-v2`, both IIE and ECC substantially reduced the number of extracted relations, reflecting a shift from high-recall candidate generation toward precision-oriented filtering. In the context of `Watson`, this shift is aligned with the intended usage scenario: users benefit more from a smaller set of well-supported, traceable candidates than from a large volume of noisy proposals that increase curation burden.

The asymmetry between pseudo gold standards suggests that the automated scores are sensitive to the particular reference construction and may encode different boundary decisions about what constitutes a “matching” relation. Therefore, the reported metrics should be interpreted primarily as comparative evidence about pipeline behavior (coverage versus conservativeness, and ranking reliability under retrieval) rather than as absolute correctness. Overall, the chapter demonstrates that modifying the inference strategy can substantially improve usability, relevant properties-

92:2520092507

grounding, controllability, and verification, while making explicit the compute-reliability trade-off introduced by deliberative multi-model evaluation.

Finally, the Recall@5 values obtained for `cannary-reasoning-v2`, `DeepSeek-V3.2-Speciale`, and `Kimi-K2.5` are broadly comparable, despite substantial differences in model size and extraction strategies. This observation suggests that a smaller, domain-adapted model is sufficient for the task of relation extraction from biomedical text, provided that it is appropriately fine-tuned. At the same time, the convergence of performance across models indicates that the available textual data may have been largely exploited, pointing to diminishing returns from further scaling within the same modality. Consequently, achieving additional improvements is likely to require incorporation of complementary information sources, such as figures or structured external databases. Moreover, the use of smaller, specialized models may offer practical advantages, including increased robustness and more predictable behavior in domain-specific settings.

93:1013534803

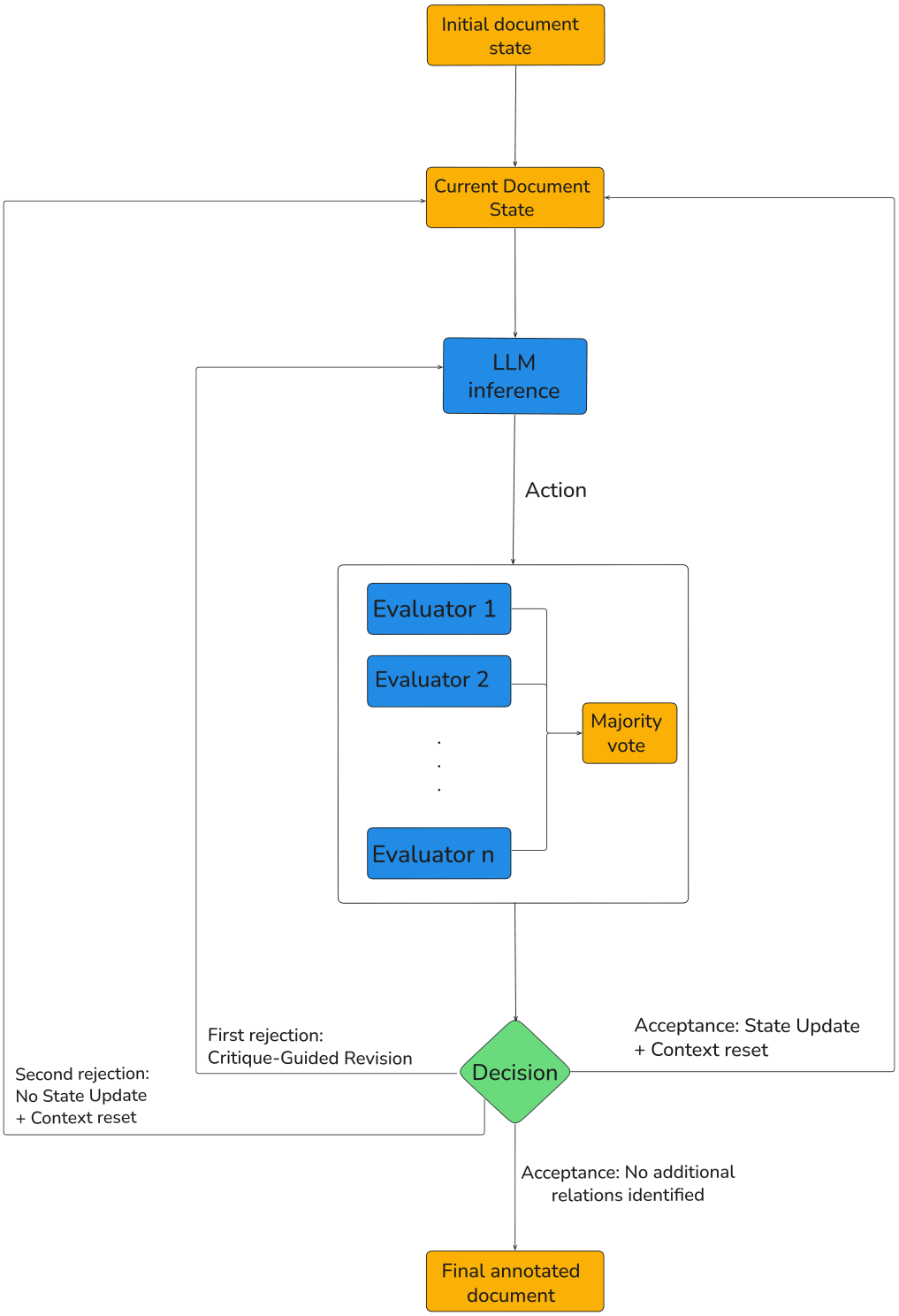



FIGURE 7.6: ECC as a nested deliberative state-transition process. The extractor proposes a relation–evidence pair which is independently evaluated by multiple agents. Only majority acceptance triggers the annotation and advances the persistent document state; rejection initiates a critique-guided revision loop.

94:4601671111

# Chapter 8

# Summary

This dissertation addressed the challenge of transforming unstructured biomedical literature into structured, computationally actionable models, with a particular focus on supporting the construction of Petri net representations. The work was motivated by the rapidly increasing volume of scientific publications and the resulting fragmentation of knowledge, which limits the efficient integration of findings into coherent system-level models.

The central objective was to develop methods that assist domain experts in extracting biologically meaningful entities and relations from text, while maintaining the precision, interpretability, and traceability required for scientific modeling. Rather than pursuing full automation, the dissertation adopted a semi-automated, human-in-the-loop paradigm, balancing computational efficiency with reliability.

The contributions of this work span multiple interconnected components of the information extraction and modeling pipeline.

First, NER was investigated as a foundational step in structuring biomedical text. The proposed improvements to the `UniversalNER` model demonstrated that aggressive weight quantization can significantly reduce computational requirements without sacrificing recognition performance, enabling scalable processing of large corpora. In parallel, the `ner-on` model family introduced a multi-type, span-aware extraction approach that overcomes key limitations of prompt-based methods. Importantly, the results showed that even highly accurate entity recognition is insufficient on its own for constructing coherent Petri net-based models, highlighting the need for further processing stages.

Second, a flexible approach to EL was developed, centered around contextual representation and similarity-based retrieval. The `anydef` model demonstrated that competitive performance can be achieved with relatively small training datasets when combined with semantic embeddings. The modular `FELA` pipeline further illustrated the benefits of integrating recognition, disambiguation, and reranking into an adaptable system. However, practical application to biological modeling revealed a fundamental limitation: many biologically relevant entities are inherently context-dependent and cannot be reliably mapped to static identifiers without loss of meaning. This finding led to a critical insight that strict normalization may degrade, rather than improve, the structural correctness of executable models.

Third, the dissertation examined the broader paradigm of information extraction in light of advances in LLMs. Initially conceived as a classical multi-stage pipeline (NER, EL, RE), the methodology evolved toward unified extraction approaches enabled by reasoning-capable models. These models allow simultaneous identification of entities and relations, improving coherence and reducing error propagation. Nevertheless, the research demonstrated that LLMs are most effective



95:2272593943

when used as assistive tools rather than fully autonomous systems, reinforcing the importance of expert validation.

Building on these insights, a framework for relation-centered extraction and Petri net construction was developed and implemented in the `Watson` platform. The system was evaluated both manually and automatically, revealing that it can recover a substantial portion of biologically meaningful interactions directly from text. At the same time, the analysis suggests several potential inherent limitations of text-based extraction, including the omission of implicit knowledge, restricted access to information contained in figures, and challenges arising from document segmentation. These observations should be treated as preliminary hypotheses that require further investigation.

To address these limitations, the inference procedure was redesigned. The IIE approach reframed extraction as a monotonic, evidence-grounded annotation process, preserving full document context and improving interpretability. The ECC extension further enhanced reliability by incorporating multi-model verification. Together, these methods shifted the system toward a precision-oriented operating regime, producing fewer but more reliable and traceable candidate relations. This trade-off was shown to better align with the needs of expert-driven modeling workflows.

An important overarching finding of the dissertation is that fully automated reconstruction of complete biological pathways from text alone is neither currently feasible nor practically desirable. Instead, the proposed approach emphasizes the generation of high-quality candidate relations that can be efficiently curated by experts. This paradigm preserves scientific rigor while reducing the manual effort required during the early stages of model construction.

The methodology was developed and evaluated within a focused biomedical domain, namely the metabolism and functional role of HDL. This restriction enabled controlled experimentation and domain-informed validation. A Petri net model of HDL-related processes was constructed as part of this work, to better grasp the biomedical context.

In summary, this dissertation makes the following key contributions:

- development of efficient and scalable NER methods tailored to biomedical text,
- introduction of a flexible, modular entity linking framework and identification of its limitations in process-based modeling,
- demonstration of a paradigm shift from pipeline-based to unified, relation-centered information extraction,
- design of the `Watson` platform for semi-automatic extraction and modeling,
- proposal of IIE and ECC as methods for improving reliability and interpretability,
- empirical analysis of the strengths and limitations of text-based biological model reconstruction.

The results highlight both the potential and the boundaries of current artificial intelligence methods in scientific knowledge extraction. Future work should focus on integrating additional information modalities, such as figures and structured databases, as well as improving methods for handling implicit knowledge and long-range dependencies. Further advances in this direction may enable more complete and accurate reconstruction of complex biological systems, while continuing to support expert-driven scientific discovery.

96:5876718374

## Acknowledgments

I gratefully acknowledge Poznan Supercomputing and Networking Center (PSNC) and Polish high-performance computing infrastructure PLGrid (HPC Center: ACK Cyfronet AGH) for providing computer facilities and support within computational grant.

97:7821042030

# Bibliography


[1] Monica Agrawal, Stefan Hegselmann, Hunter Lang, Yoon Kim, and David Sontag. Large language models are few-shot clinical information extractors. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1998–2022, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.

[2] Red Hat AI and vLLM Project. LLM Compressor, 8 2024.

[3] Anton Alexandrov, Veselin Raychev, Mark Niklas Mueller, Ce Zhang, Martin Vechev, and Kristina Toutanova. Mitigating catastrophic forgetting in language transfer via model merging. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 17167–17186, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[4] HASSANE ALLA and RENÉ DAVID. CONTINUOUS AND HYBRID PETRI NETS. *Journal of Circuits, Systems and Computers*, 08(01):159–188, February 1998.

[5] Isabelle Augenstein, Leon Derczynski, and Kalina Bontcheva. Generalisation in named entity recognition: A quantitative analysis. *Comput. Speech Lang.*, 44:61–83, July 2017.

[6] Tom Ayoola, Shubhi Tyagi, Joseph Fisher, Christos Christodoulopoulos, and Andrea Pierleoni. ReFinED: An efficient zero-shot-capable approach to end-to-end entity linking. In Anastassia Loukina, Rashmi Gangadharaiah, and Bonan Min, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Track*, pages 209–220, Hybrid: Seattle, Washington + Online, July 2022. Association for Computational Linguistics.

[7] John A Bachman, Benjamin M Gyori, and Peter K Sorger. Automated assembly of molecular mechanisms at scale from text mining and curated databases. *Mol. Syst. Biol.*, 19(5):e11325, May 2023.

[8] Andrew Blair-Stanek and Benjamin Van Durme. Llms provide unstable answers to legal questions, 2025.

[9] Christoph Brinkrolf, Lennart Ochel, and Ralf Hofestädt. VANESA: An open-source hybrid functional petri net modeling and simulation environment in systems biology. *Biosystems*, 210:104531, December 2021.

[10] David Guy Brizan and Abdullah Uz Tansel. A. survey of entity resolution and record linkage methodologies. *Communications of the IIMA*, 6(3):5, 2006.

[11] Broadcom. Rabbitmq one broker to queue them all, 2026.

[12] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott




98:7738343915

Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc.

[13] Finja Büchel, Nicolas Rodriguez, Neil Swainston, Clemens Wrzodek, Tobias Czauderna, Roland Keller, Florian Mittag, Michael Schubert, Mihai Glont, Martin Golebiewski, Martijn van Iersel, Sarah Keating, Matthias Rall, Michael Wybrow, Henning Hermjakob, Michael Hucka, Douglas B Kell, Wolfgang Müller, Pedro Mendes, Andreas Zell, Claudine Chaouiya, Julio Saez-Rodriguez, Falk Schreiber, Camille Laibe, Andreas Dräger, and Nicolas Le Novère. Path2models: large-scale generation of computational models from biochemical pathway maps. *BMC Systems Biology*, 7(1), November 2013.

[14] Gustavo Callou, Paulo Maciel, Dietmar Tutsch, Julian Arajo, Joo Ferreira, and Rafael Souz. *A Petri Net-Based Approach to the Quantification of Data Center Dependability.* InTech, August 2012.

[15] David Campos, Sérgio Matos, and José Luís Oliveira. Biomedical named entity recognition: a survey of machine-learning tools. In *Theory and applications for advanced text mining.* IntechOpen, 2012.

[16] Celery Team. Celery - distributed task queue, 2026.

[17] Alberto Cetoli, Stefano Bragaglia, Andrew D. O'Harney, Marc Sloan, and Mohammad Akbari. A neural approach to entity linking on wikidata. In *Advances in Information Retrieval*, page 78–86, Cham, 2019. Springer International Publishing.

[18] Alan Chait and Robert H Eckel. Lipids, lipoproteins, and cardiovascular disease: Clinical pharmacology now and in the future. *J. Clin. Endocrinol. Metab.*, 101(3):804–814, March 2016.

[19] Xiao Chang, Shen Yan, Yizheng Zhang, Yingchun Zhang, Luyang Li, Zhanyu Gao, Xuefei Lin, and Xu Chi. Ginv2.0: a comprehensive topological network integrating molecular interactions from multiple knowledge bases. *npj Systems Biology and Applications*, 10(1), January 2024.

[20] C. Chaouiya. Petri net modelling of biological networks. *Briefings in Bioinformatics*, 8(4):210–219, March 2007.

[21] Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation, 2024.

[22] Xiang Chen, Lei Li, Shumin Deng, Chuanqi Tan, Changliang Xu, Fei Huang, Luo Si, Huajun Chen, and Ningyu Zhang. LightNER: A lightweight tuning paradigm for low-resource NER via pluggable prompting. In Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na, editors, *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2374–2387, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics.

[23] Kyung-Hyun Cho. The current status of research on high-density lipoproteins (HDL): A paradigm shift from HDL quantity to HDL quality and HDL functionality. *Int. J. Mol. Sci.*, 23(7):3967, April 2022.

[24] Chroma Team. Chroma: The ai-native open-source embedding database. `https://github.com`, 2023.

99:2013865732

[25] Chase Cockrell, Scott Christley, and Gary An. Facilitating automated conversion of scientific knowledge into scientific simulation models with the machine assisted generation, calibration, and comparison (magcc) framework, 2022.

[26] W. de Haan, A. Bhattacharjee, P. Ruddle, M. H. Kang, and M. R. Hayden. ABCA1 in adipocytes regulates adipose tissue lipid content, glucose tolerance, and insulin sensitivity. *J Lipid Res*, 55(3):516–523, Mar 2014.

[27] DeepSeek-AI. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025.

[28] DeepSeek-AI. Deepseek-v3.2: Pushing the frontier of open large language models, 2025.

[29] Antonin Delpeuch. OpenTapioca: Lightweight Entity Linking for Wikidata, April 2019. journal-article.

[30] Kartik Detroja, C K Bhensdadia, and Brijesh S Bhatt. A survey on relation extraction. *Intell. Syst. Appl.*, 19(200244):200244, September 2023.

[31] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[32] Sören Domrös, Reinhard von Hanxleden, Miro Spönemann, Ulf Rüegg, and Christoph Daniel Schulze. The eclipse layout kernel, 2023.

[33] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, March 1990.

[34] Yasuhiro Endo, Masanori Fujita, and Katsunori Ikewaki. HDL functions-current status and future perspectives. *Biomolecules*, 13(1):105, January 2023.

[35] Kenneth R Feingold. *Introduction to lipids and lipoproteins*. MDText.com, Inc., South Dartmouth, 2000.

[36] Brian A Ference, Henry N Ginsberg, Ian Graham, Kausik K Ray, Chris J Packard, Eric Bruckert, Robert A Hegele, Ronald M Krauss, Frederick J Raal, Heribert Schunkert, Gerald F Watts, Jan Borén, Sergio Fazio, Jay D Horton, Luis Masana, Stephen J Nicholls, Børge G Nordestgaard, Bart van de Sluis, Marja-Riitta Taskinen, Lale Tokgözoglu, Ulf Landmesser, Ulrich Laufs, Olov Wiklund, Jane K Stock, M John Chapman, and Alberico L Catapano. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. evidence from genetic, epidemiologic, and clinical studies. a consensus statement from the european atherosclerosis society consensus panel. *Eur. Heart J.*, 38(32):2459–2472, 08 2017.

[37] G. Florin, C. Fraize, and S. Natkin. Stochastic Petri nets: Properties, applications and tools. *Microelectronics Reliability*, 31(4):669–697, January 1991.

[38] Dorota Formanowicz, Marcin Radom, Agnieszka Rybarczyk, Krzysztof Tanaś, and Piotr Formanowicz. Control of cholesterol metabolism using a systems approach. *Biology*, 11(3):430, 03 2022.

[39] Dorota Formanowicz, Adam Aron Rynkiewicz, and Piotr Formanowicz. Regulatory bottlenecks linking hdl dysfunction, inflammation, and oxidative stress: A petri net-based systems analysis. *International Journal of Molecular Sciences*, 27(18):8066, September 2026.

[40] Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, Shawn Presser, and Connor Leahy. The Pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*, 2020.

[41] Henry N Ginsberg. Lipoprotein physiology. *Endocrinol. Metab. Clin. North Am.*, 27(3):503–519, September 1998.

100:5680432339

[42] Charles Goddard, Shamane Siriwardhana, Malikeh Ehghaghi, Luke Meyers, Vladimir Karpukhin, Brian Benedict, Mark McQuade, and Jacob Solawetz. Arcee's MergeKit: A toolkit for merging large language models. In Franck Dernoncourt, Daniel Preoţiuc-Pietro, and Anastasia Shimorina, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 477–485, Miami, Florida, US, November 2024. Association for Computational Linguistics.

[43] Rita González-Márquez, Luca Schmidt, Benjamin M Schmidt, Philipp Berens, and Dmitry Kobak. The landscape of biomedical research. *Patterns*, 5(6):100968, June 2024.

[44] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[45] Rentong Guo, Xiaofan Luan, Long Xiang, Xiao Yan, Xiaomeng Yi, Jigao Luo, Qianya Cheng, Weizhi Xu, Jiarui Luo, Frank Liu, et al. Manu: a cloud native vector database management system. *Proceedings of the VLDB Endowment*, 15(12):3548–3561, 2022.

[46] Benjamin M Gyori, John A Bachman, Kartik Subramanian, Jeremy L Muhlich, Lucian Galescu, and Peter K Sorger. From word models to executable models of signaling networks using automated assembly. *Mol. Syst. Biol.*, 13(11):954, November 2017.

[47] Chang Yeop Han, Chongren Tang, Myriam E Guevara, Hao Wei, Tomasz Wietecha, Baohai Shao, Savitha Subramanian, Mohamed Omer, Shari Wang, Kevin D O'Brien, Santica M Marcovina, Thomas N Wight, Tomas Vaisar, Maria C de Beer, Frederick C de Beer, William R Osborne, Keith B Elkon, and Alan Chait. Serum amyloid a impairs the antiinflammatory properties of HDL. *J. Clin. Invest.*, 126(1):266–281, January 2016.

[48] Bahareh Harandizadeh and Sameer Singh. Tweeki: Linking named entities on Twitter to a knowledge graph. In Wei Xu, Alan Ritter, Tim Baldwin, and Afshin Rahimi, editors, *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 222–231, Online, November 2020. Association for Computational Linguistics.

[49] Simon O Haykin. *Neural Networks and Learning Machines*. Pearson, Upper Saddle River, NJ, 3 edition, November 2008.

[50] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. DeBERTa: Decoding-enhanced BERT with Disentangled Attention, 2021.

[51] Monika Heiner, David Gilbert, and Robin Donaldson. *Petri Nets for Systems and Synthetic Biology*, page 215–264. Springer Berlin Heidelberg, 2008.

[52] Monika Heiner, Ina Koch, and Jürgen Will. Model validation of biological pathways using petri nets—demonstrated for apoptosis. *Biosystems*, 75(1-3):15–28, 07 2004.

[53] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, November 1997.

[54] Jiwoo Hong, Noah Lee, and James Thorne. ORPO: Monolithic preference optimization without reference model. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 11170–11189, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[55] Pin-Lun Hsu, Yun Dai, Vignesh Kothapalli, Qingquan Song, Shao Tang, Siyu Zhu, Steven Shimizu, Shivam Sahni, Haowen Ning, Yanning Chen, and Zhipeng Wang. Liger-kernel: Efficient triton kernels for LLM training. In *Championing Open-source DEvelopment in ML Workshop @ ICML25*, 2025.

[56] Jiansheng Huang, Dongdong Wang, Li-Hao Huang, and Hui Huang. Roles of reconstituted high-density lipoprotein nanoparticles in cardiovascular disease: A new paradigm for drug discovery. *Int. J. Mol. Sci.*, 21(3):739, January 2020.

101:9229410481

[57] Ming-Siang Huang, Jen-Chieh Han, Pei-Yen Lin, Yu-Ting You, Richard Tzong-Han Tsai, and Wen-Lian Hsu. Surveying biomedical relation extraction: a critical examination of current datasets and the proposal of a new resource. *Brief. Bioinform.*, 25(3), March 2024.

[58] Fatima Iqbal, Wendy S Baker, Madiha I Khan, Shwetha Thukuntla, Kevin H McKinney, Nicola Abate, and Demidmaa Tuvdendorj. Current and future therapies for addressing the effects of inflammation on hdl cholesterol metabolism. *British Journal of Pharmacology*, 174(22):3986–4006, 2017.

[59] Kurt Jensen. *Coloured Petri Nets*, page 248–299. Springer Berlin Heidelberg, 1987.

[60] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mistral 7b, 2023.

[61] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3):535–547, 2019.

[62] Anne Jomard and Elena Osto. High density lipoproteins: Metabolism, function, and therapeutic potential. *Front. Cardiovasc. Med.*, 7:39, March 2020.

[63] Josep Julve and Joan Carles Escolà-Gil. High-density lipoproteins and cardiovascular disease: The good, the bad, and the future. *International Journal of Molecular Sciences*, 22(14):7488, July 2021.

[64] Josep Julve, Jesús M Martín-Campos, Joan Carles Escolà-Gil, and Francisco Blanco-Vaca. Chylomicrons: Advances in biology, pathology, laboratory testing, and therapeutics. *Clin. Chim. Acta*, 455:134–148, April 2016.

[65] Minoru Kanehisa, Miho Furumichi, Yoko Sato, Yuriko Matsuura, and Mari Ishiguro-Watanabe. Kegg: biological systems database as a model of the real world. *Nucleic Acids Research*, 53(D1):D672–D677, October 2024.

[66] I. Koch, W. Reisig, and F. Schreiber, editors. *Modeling in Systems Biology. The Petri Net Approach*. Springer London, 2011.

[67] Ina Koch and Bianca Büttner. Computational modeling of signal transduction networks without kinetic parameters: Petri net approaches. *American Journal of Physiology-Cell Physiology*, 324(5):C1126–C1140, May 2023.

[68] Ina Koch, Wolfgang Reisig, and Falk Schreiber, editors. *Modeling in systems biology. The Petri Net Approach*. Springer, London, England, October 2010.

[69] Nikolaos Kolitsas, Octavian-Eugen Ganea, and Thomas Hofmann. End-to-end neural entity linking. In Anna Korhonen and Ivan Titov, editors, *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 519–529, Brussels, Belgium, October 2018. Association for Computational Linguistics.

[70] F Kronenberg and G Utermann. Lipoprotein(a): resurrected by genetics. *J. Intern. Med.*, 273(1):6–30, January 2013.

[71] Tuan Lai, Heng Ji, and ChengXiang Zhai. Improving candidate retrieval with Entity Profile Generation for Wikidata Entity Linking. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 3696–3711, 2022.

[72] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

102:2994076636

[73] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, February 2020.

[74] Ulf Leser and Jörg Hakenberg. What makes a gene name? named entity recognition in the biomedical literature. *Brief. Bioinform.*, 6(4):357–369, December 2005.

[75] Chaofan Li, Zheng Liu, Shitao Xiao, and Yingxia Shao. Making large language models a better foundation for dense retrieval, 2023.

[76] Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. A survey on deep learning for named entity recognition. *IEEE transactions on knowledge and data engineering*, 34(1):50–70, 2020.

[77] Xuefei Lin, Xiao Chang, Yizheng Zhang, Zhanyu Gao, and Xu Chi. Automatic construction of Petri net models for computational simulations of molecular interaction network. *npj Systems Biology and Applications*, 10(1), November 2024.

[78] Fei Liu, Monika Heiner, and David Gilbert. Coloured petri nets for multilevel, multiscale and multidimensional modelling of biological systems. *Briefings in Bioinformatics*, 20(3):877–886, November 2017.

[79] Fei Liu, Monika Heiner, and David Gilbert. Hybrid modelling of biological systems: current progress and future prospects. *Briefings in Bioinformatics*, 23(3), March 2022.

[80] Jerry Liu. LlamaIndex, 11 2022.

[81] Jingjing Liu, Panupong Pasupat, Scott Cyphers, and Jim Glass. Asgard: A portable architecture for multilingual dialogue systems. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8386–8390, 2013.

[82] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.*, 55(9):1–35, January 2023.

[83] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

[84] Zihan Liu, Yan Xu, Tiezheng Yu, Wenliang Dai, Ziwei Ji, Samuel Cahyawijaya, Andrea Madotto, and Pascale Fung. Crossner: Evaluating cross-domain named entity recognition, 2020.

[85] María Luna-Luna, Eric Niesor, and Óscar Pérez-Méndez. HDL as bidirectional lipid vectors: Time for new paradigms. *Biomedicines*, 10(5):1180, May 2022.

[86] Christopher M. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, NY, 1 edition, August 2006.

[87] Alireza Mansouri, Lilly Suriani Affendey, and Ali Mamat. Named entity recognition approaches. *International Journal of Computer Science and Network Security*, 8(2):339–344, 2008.

[88] Rui Mao, Kai He, Xulang Zhang, Guanyi Chen, Jinjie Ni, Zonglin Yang, and Erik Cambria. A survey on semantic processing techniques. *Information Fusion*, 101:101988, 2024.

[89] Meta Platforms, Inc. React: A javascript library for building user interfaces, 2026.

[90] Marija Milacic, Deidre Beavers, Patrick Conley, Chuqiao Gong, Marc Gillespie, Johannes Griss, Robin Haw, Bijay Jassal, Lisa Matthews, Bruce May, Robert Petryszak, Eliot Ragueneau, Karen Rothfels, Cristoffer Sevilla, Veronica Shamovsky, Ralf Stephan, Krishna Tiwari, Thawfeek Varusai, Joel Weiser, Adam Wright, Guanming Wu, Lincoln Stein, Henning Hermjakob, and Peter D'Eustachio. The reactome pathway knowledgebase 2024. *Nucleic Acids Research*, 52(D1):D672–D678, November 2023.

103:1025565885

[91] Nikola Milošević and Wolfgang Thielemann. Comparison of biomedical relationship extraction methods and models for knowledge graph creation. *Web Semant.*, 75(100756):100756, January 2023.

[92] Sewon Min, Xinxi Lyu, Ari Holtzman, Mikel Artetxe, Mike Lewis, Hannaneh Hajishirzi, and Luke Zettlemoyer. Rethinking the role of demonstrations: What makes in-context learning work? In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 11048–11064, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.

[93] MUI Team. Material ui: Comprehensive react component library, 2026.

[94] T Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77:541–580, 1989.

[95] Sumra Nazir, Vera Jankowski, Guzide Bender, Stephen Zewinger, Kerry-Anne Rye, and Emiel P C van der Vorst. Interaction between high-density lipoproteins and inflammation: Function matters more than concentration! *Adv. Drug Deliv. Rev.*, 159:94–119, October 2020.

[96] Oecd. *Health at a glance 2025*. Health at a Glance. OECD, November 2025.

[97] OpenAI. gpt-oss-120b & gpt-oss-20b model card, 2025.

[98] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc.

[99] M. Peleg. Using Petri Net Tools to Study Properties and Dynamics of Biological Systems. *Journal of the American Medical Informatics Association*, 12(2):181–199, November 2004.

[100] Guilherme Penedo, Hynek Kydlíček, Loubna Ben allal, Anton Lozhkov, Margaret Mitchell, Colin Raffel, Leandro Von Werra, and Thomas Wolf. The fineweb datasets: Decanting the web for the finest text data at scale. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024.

[101] Carl Adam Petri. *Kommunikation mit Automaten*. PhD thesis, Universität Bonn, Bonn, Germany, 1962.

[102] Fabio Petroni, Aleksandra Piktus, Angela Fan, Patrick S. H. Lewis, Majid Yazdani, Nicola De Cao, James Thorne, Yacine Jernite, Vladimir Karpukhin, Jean Maillard, Vassilis Plachouras, Tim Rocktäschel, and Sebastian Riedel. KILT: a benchmark for knowledge intensive language tasks. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tür, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou, editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 2523–2544, Online, 2021. Association for Computational Linguistics.

[103] PostgreSQL Global Development Group. Postgresql: The world's most advanced open source relational database, 2026.

[104] Alexandra Pomares Quimbaya, Alejandro Sierra Múnera, Rafael Andrés González Rivera, Julián Camilo Daza Rodríguez, Oscar Mauricio Muñoz Velandia, Angel Alberto Garcia Peña, and Cyril Labbé. Named entity recognition over electronic health records through a combined dictionary-based approach. *Procedia Computer Science*, 100:55–61, 2016.

[105] Marcin Radom, Agnieszka Rybarczyk, Bartlomiej Szawulak, Hubert Andrzejewski, Piotr Chabelski, Adam Kozak, and Piotr Formanowicz. Holmes: a graphical tool for development, simulation and analysis of petri net based models of complex biological systems. *Bioinformatics*, 33(23):3822–3823, December 2017.

[106] Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: your language model is secretly a reward model. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.

[107] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1), January 2020.

[108] Eliot Ragueneau, Chuqiao Gong, Pierre Sinquin, Cristoffer Sevilla, Deidre Beavers, Alexander Grentner, Johannes Griss, Gregory F J Hogue, Nancy T Li, Lisa Matthews, Bruce May, Marija Milacic, Helia Mohammadi, Robert Petryszak, Karen Rothfels, Veronica Shamovsky, Ralf Stephan, Krishna Tiwari, Joel Weiser, Adam Wright, Marc Gillespie, Guanming Wu, Lincoln Stein, Henning Hermjakob, and Peter D'Eustachio. The reactome knowledgebase 2026. *Nucleic Acids Research*, 54(D1):D673–D681, November 2025.

[109] React Flow Team. React flow, 2026.

[110] V. N. Reddy, M. L. Mavrovouniotis, and M. N. Liebman. Petri Net Representations in Metabolic Pathways. In *Proceedings of the First International Conference on Intelligent Systems for Molecular Biology*, pages 328–336, 1993.

[111] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China, November 2019. Association for Computational Linguistics.

[112] Michael Röder, Ricardo Usbeck, Sebastian Hellmann, Daniel Gerber, and Andreas Both. $N^3$ - a collection of datasets for named entity recognition and disambiguation in the NLP interchange format. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3529–3533, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA).

[113] Robert S Rosenson, H Bryan Brewer, Jr, M John Chapman, Sergio Fazio, M Mahmood Hussain, Anatol Kontush, Ronald M Krauss, James D Otvos, Alan T Remaley, and Ernst J Schaefer. HDL measures, particle heterogeneity, proposed nomenclature, and relation to atherosclerotic cardiovascular events. *Clin. Chem.*, 57(3):392–410, March 2011.

[114] Agnieszka Rybarczyk, Dorota Formanowicz, and Piotr Formanowicz. The role of macrophage dynamics in atherosclerosis analyzed using a Petri net-based model. *Appl. Sci.*, 14(8):3219, April 2024.

[115] Andrea Sackmann, Monika Heiner, and Ina Koch. Application of Petri net based analysis techniques to signal transduction pathways. *BMC Bioinformatics*, 7:482, 2006.

[116] Yutaka Sasaki, Yoshimasa Tsuruoka, John McNaught, and Sophia Ananiadou. How to make the most of ne dictionaries in statistical ner. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 63–70, 2008.

[117] Megumi Sato, Ryunosuke Ohkawa, Akira Yoshimoto, Kouji Yano, Naoya Ichimura, Madoka Nishimori, Shigeo Okubo, Yutaka Yatomi, and Minoru Tozuka. Effects of serum amyloid a on the

105:3481807064

structure and antioxidant ability of high-density lipoprotein. *Biosci. Rep.*, 36(4):e00369–e00369, August 2016.

[118] Sonja Scharf, Jörg Ackermann, Leonie Bender, Patrick Wurzel, Hendrik Schäfer, Martin-Leo Hansmann, and Ina Koch. Holistic view on the structure of immune response: Petri net model. *Biomedicines*, 11(2):452, February 2023.

[119] Mirjam Schuchardt, Nicole Prüfer, Yuexing Tu, Jaqueline Herrmann, Xiu-Ping Hu, Sarah Chebli, Katja Dahlke, Walter Zidek, Markus van der Giet, and Markus Tölle. Dysfunctional high-density lipoprotein activates toll-like receptors via serum amyloid a in vascular smooth muscle cells. *Sci. Rep.*, 9(1):3421, March 2019.

[120] Sebastián Ramírez. Fastapi framework, high performance, easy to learn, fast to code, ready for production, 2026.

[121] Yufei Shang, Yanrong Guo, Shijie Hao, and Richang Hong. Biomedical relation extraction via adaptive document-relation cross-mapping and concept unique identifier, 2025.

[122] Wei Shen, Yuhan Li, Yinan Liu, Jiawei Han, Jianyong Wang, and Xiaojie Yuan. Entity linking meets deep learning: Techniques and solutions. *IEEE Transactions on Knowledge and Data Engineering*, 35(3):2556–2578, 2021.

[123] Wei Shen, Jianyong Wang, and Jiawei Han. Entity linking with a knowledge base: Issues, techniques, and solutions. *IEEE Transactions on Knowledge and Data Engineering*, 27(2):443–460, 2014.

[124] Lucian P. Smith, Rahuman S. Malik-Sheriff, Tung V. N. Nguyen, Henning Hermjakob, Jonathan Karr, Bilal Shaikh, Logan Drescher, Ion I. Moraru, James C. Schaff, Eran Agmon, Alexander A. Patrie, Michael L. Blinov, Joseph L. Hellerstein, Elebeoba E. May, David P. Nickerson, John H. Gennari, and Herbert M. Sauro. Verification and reproducible curation of the biomodels repository. *PLOS Computational Biology*, 21(12):e1013239, December 2025.

[125] Ihor Stepanov and Mykhailo Shtopko. Gliner multi-task: Generalist lightweight model for various information extraction tasks, 2024.

[126] Deep Search Team. Docling technical report. Technical report, IBM Research Zurich, 8 2024.

[127] Kimi Team. Kimi k2.5: Visual agentic intelligence, 2026.

[128] Qwen Team. Qwen3 technical report, 2025.

[129] Qwen Team. Qwq-32b: Embracing the power of reinforcement learning, March 2025.

[130] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models, 2023.

[131] Sotirios Tsimikas. A test in context: Lipoprotein(a): Diagnosis, prognosis, controversies, and emerging therapies. *J. Am. Coll. Cardiol.*, 69(6):692–711, February 2017.

[132] Lewis Tunstall, Edward Beeching, Nathan Lambert, Nazneen Rajani, Shengyi Huang, Kashif Rasul, Alvaro Bartolome, Carlos M. Patiño, Alexander M. Rush, and Thomas Wolf. The Alignment Handbook.

[133] Marco A Valenzuela-Escárcega, Özgün Babur, Gus Hahn-Powell, Dane Bell, Thomas Hicks, Enrique Noriega-Atala, Xia Wang, Mihai Surdeanu, Emek Demir, and Clayton T Morrison. Large-scale automated machine reading discovers new cancer driving mechanisms. *Database: The Journal of Biological Databases and Curation*, 2018.

106:8013309533

[134] Marco A. Valenzuela-Escárcega, Gustave Hahn-Powell, Thomas Hicks, and Mihai Surdeanu. A domain-independent rule-based framework for event extraction. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing: Software Demonstrations (ACL-IJCNLP)*, pages 127–132. ACL-IJCNLP 2015, 2015. Paper available at `http://www.aclweb.org/anthology/P/P15/P15-4022.pdf`.

[135] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc.

[136] Jianguo Wang, Xiaomeng Yi, Rentong Guo, Hai Jin, Peng Xu, Shengjun Li, Xiangyu Wang, Xiangzhou Guo, Chengming Li, Xiaohai Xu, et al. Milvus: A purpose-built vector data management system. In *Proceedings of the 2021 International Conference on Management of Data*, pages 2614–2627, 2021.

[137] Shuhe Wang, Xiaofei Sun, Xiaoya Li, Rongbin Ouyang, Fei Wu, Tianwei Zhang, Jiwei Li, and Guoyin Wang. Gpt-ner: Named entity recognition via large language models, 2023.

[138] Shuhe Wang, Xiaofei Sun, Xiaoya Li, Rongbin Ouyang, Fei Wu, Tianwei Zhang, Jiwei Li, Guoyin Wang, and Chen Guo. GPT-NER: Named entity recognition via large language models. In Luis Chiruzzo, Alan Ritter, and Lu Wang, editors, *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 4257–4275, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics.

[139] Weiwei Wang, Weijie Zou, and Jiyong Min. Solving llm repetition problem in production: A comprehensive study of multiple solutions, 2025.

[140] World Health Organization (WHO). The top 10 causes of death. `https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death`, 2024.

[141] Ledell Wu, Fabio Petroni, Martin Josifoski, Sebastian Riedel, and Luke Zettlemoyer. Scalable zero-shot entity linking with dense entity retrieval. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6397–6407, Online, November 2020. Association for Computational Linguistics.

[142] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, Rui Zheng, Xiaoran Fan, Xiao Wang, Limao Xiong, Yuhao Zhou, Weiran Wang, Changhao Jiang, Yicheng Zou, Xiangyang Liu, Zhangyue Yin, Shihan Dou, Rongxiang Weng, Wenjuan Qin, Yongyan Zheng, Xipeng Qiu, Xuanjing Huang, Qi Zhang, and Tao Gui. The rise and potential of large language model based agents: a survey. *Science China Information Sciences*, 68(2), January 2025.

[143] An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*, 2024.

[144] Xi Yang, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E Smith, Christopher Parisien, Colin Compas, Cheryl Martin, Anthony B Costa, Mona G Flores, Ying Zhang, Tanja Magoc, Christopher A Harle, Gloria Lipori, Duane A Mitchell, William R Hogan, Elizabeth A Shenkman, Jiang Bian, and Yonghui Wu. A large language model for electronic health records. *NPJ Digit. Med.*, 5(1):194, December 2022.

[145] Urchade Zaratiana, Nadi Tomeh, Pierre Holat, and Thierry Charnois. GLiNER: Generalist model for named entity recognition using bidirectional transformer. In Kevin Duh, Helena Gomez, and Steven Bethard, editors, *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 5364–5376, Mexico City, Mexico, June 2024. Association for Computational Linguistics.

[146] Guibin Zhang, Hejia Geng, Xiaohang Yu, Zhenfei Yin, Zaibin Zhang, Zelin Tan, Heng Zhou, Zhong-Zhi Li, Xiangyuan Xue, Yijiang Li, Yifan Zhou, Yang Chen, Chen Zhang, Yutao Fan, Zihu Wang, Songtao Huang, Francisco Piedrahita Velez, Yue Liao, Hongru WANG, Mengyue Yang, Heng Ji, Jun Wang, Shuicheng YAN, Philip Torr, and LEI BAI. The landscape of agentic reinforcement learning for LLMs: A survey. *Transactions on Machine Learning Research*, 2026. Survey Certification.

[147] Wenxuan Zhou, Sheng Zhang, Yu Gu, Muhao Chen, and Hoifung Poon. UniversalNER: Targeted distillation from large language models for open named entity recognition. In *The Twelfth International Conference on Learning Representations*, 2024.

[148] W.M. Zuberek. Timed Petri nets definitions, properties, and applications. *Microelectronics Reliability*, 31(4):627–644, January 1991.

108:3806960529