



**Universiteit
Leiden**
The Netherlands

Healthcare information system engineering: AI technologies and open source approaches

Shen, Z.

Citation

Shen, Z. (2025, December 3). *Healthcare information system engineering: AI technologies and open source approaches*. Retrieved from <https://hdl.handle.net/1887/4284431>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4284431>

Note: To cite this publication please use the final published version (if applicable).

Healthcare Information System Engineering: AI Technologies and Open Source Approaches

Zhengru Shen

Healthcare Information System Engineering: AI Technologies and Open Source Approaches

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 3 december 2025
klokke 13:00 uur

door

Zhengru Shen
geboren te Anhui, China
in 1987

Promotores: Prof.dr. M.R. Spruit
Prof.dr. S. Brinkkemper Utrecht University

Promotiecommissie: Prof.dr. M.M. Bonsangue
Prof.dr. S. Verberne
Prof.dr.ir. J.M.W. Visser
Prof.dr.ir. R.W. Helms Open University
Prof.dr. Y. Velegrakis Utrecht University

Copyright © 2025 Zhengru Shen. All rights reserved.

This thesis was accomplished with financial support from European Union's Horizon 2020 research and innovation programme under grant agreement No. 634238 (OPERAM).

Contents

1	Introduction	1
1.1	Healthcare Information System Engineering	1
1.1.1	Electronic Health Records	2
1.1.2	E-Health	2
1.1.3	Data Mining and Big Data	2
1.1.4	Trends and Challenges	3
1.2	Open Source Software (OSS)	4
1.3	OSS in HIS Engineering	4
1.4	Research Questions	6
1.5	Research Methods	10
1.5.1	Prototyping	10
1.5.2	Computational experiments	10
1.5.3	Quantitative analysis	11
1.6	Dissertation Outline	11
1.7	Dissertation outcomes	12
1.7.1	Data	13
1.7.2	Code and Prototypes	13
1.7.3	Publications	14
1.7.4	Other Publications by the Author	15
2	Federated Architecture for Multinational Clinical Trials	17
2.1	Introduction	18
2.2	Background	19
2.2.1	Status Quo of Data Integration	19
2.2.2	Design Objectives	20
2.3	Data Integration Component	21

2.3.1	Data Representation	21
2.3.2	Integration Architecture	24
2.3.3	ETL	25
2.4	Evaluation & Contribution	27
2.5	Conclusion	28
3	A Lightweight API-Based Approach for Clinical NLP	29
3.1	Introduction	30
3.2	Methods	33
3.2.1	Design Science	33
3.2.2	Evaluation Design	33
3.3	Results	35
3.3.1	A Lightweight NLP Architecture for Clinical NLP	35
3.3.2	Prototype: API-Based Clinical Concept Extraction	39
3.3.3	Evaluation Results	43
3.4	Discussion	45
3.4.1	Evaluation Results	46
3.4.2	Error Analysis	46
3.4.3	Limitations and Future Research	47
3.5	Conclusion	48
4	Automatic Extraction of Adverse Drug Reactions from SmPC	49
4.1	Introduction	50
4.2	Materials and Methods	51
4.2.1	Dataset	51
4.2.2	Scraping the Side Effects Section of SmPC	51
4.2.3	ADR Extraction	52
4.2.4	Evaluation	56
4.3	Results	58
4.3.1	Overview	58
4.3.2	The Manual Evaluation Results	58
4.3.3	Error Analysis	59
4.4	Discussion	62
4.5	Conclusions	63
4.6	Supplementary Materials	63

5	Big Data Framework for Biomedical Literature Mining	65
5.1	Introduction	66
5.2	Related literature	67
5.3	Framework	69
5.4	Evaluation	71
5.4.1	Datasets	71
5.4.2	Experiment Setup	72
5.4.3	Evaluation Metrics	72
5.5	Results	73
5.5.1	Topic Modelling on the PMC OA Subset	73
5.5.2	Cancer Articles Classification	75
5.6	Discussions	75
5.7	Conclusion	76
6	A Systematic Review of Open Source Clinical Software	77
6.1	Introduction	78
6.2	Materials and Methods	80
6.2.1	Data Collection and Processing	80
6.2.2	Descriptive Analysis	83
6.2.3	Generalized Additive Models	84
6.2.4	Topic Modeling	85
6.2.5	Interactive Data Visualization	86
6.3	Results	86
6.3.1	Current Status	86
6.3.2	Geographic Distribution	87
6.3.3	Key Success Factors of Open Source Clinical Software	87
6.3.4	Main Focus Areas	89
6.4	Discussion	92
6.5	Conclusions	93
7	LOCATE: A web application to link open-source clinical software with literature	95
7.1	Introduction	96
7.2	Methods	98
7.2.1	Design Science	98
7.2.2	Data Pipeline	99
7.3	LOCATE	102

7.3.1	Overview	102
7.3.2	Key Design Principles	103
7.3.3	Technical Details	104
7.3.4	Use Cases	104
7.4	Evaluation	106
7.5	Discussion	107
7.6	Conclusions	108
8	Conclusions	109
8.1	Answers to Research Questions	109
8.2	Reflection	115
8.2.1	Limitations	115
8.2.2	Future Work	118
	Bibliography	121
	Summary	139
	Samenvatting	141
	Curriculum Vitae	145
	Acknowledgements	147