

جامعة نايف العربية للعلوم الأمنية
Naif Arab University for Security Sciences



SECURE CHATBOTS AGAINST DATA LEAK AND OVER-LEARNING THREATS

A thesis submitted in partial fulfilment of the requirements for the degree of
Master of Science in Information Security

By:

Sara Khaled Tuza

Bachelor's degree, Arab Open University, 2014

Under Supervision of:

Dr. Maryem Ammi

Submitted to:

department of Information Security, Computer and Information Security College
Naif Arab University for Security Sciences

May 2019

Contents

List of Figures	7
List of Tables	8
1 Introduction	10
1.1 Overview	10
1.2 Objectives	12
1.3 Problem Statement	14
1.4 Contributions	16
1.5 Thesis Organization	16
1.6 Summary	17
2 Theoretical Background and Literature Review	18
2.1 Artificial Intelligence	18
2.2 Chatbot Overview	23
2.2.1 Chatbot Definition	23
2.2.2 Chatbot Brief History	23
2.3 Chatbot Types	25
2.3.1 Backend Perspective	26
2.3.2 Purpose Perspective	28
2.3.3 Audience Perspective	29
2.4 Chatbot Importance	29
2.4.1 Business Industries that can gain the most benefits from Chatbots	30
2.4.2 Chatbot Areas	32
2.5 Chatbot Security Challenges	34
2.6 Security Threats : Related works	34
2.6.1 Classical Threats	35
2.6.2 New Chatbots Threats	38
3 Proposed Methodology	40
3.1 Chatbot Security Framework Context: Technologies and Tools Overview	40
3.1.1 Python	41
3.1.2 Tensorflow	41
3.2 Chatbot Architecture	42
3.2.1 Chatbot Interface	43

3.2.2	Natural Language Processing Component	43
3.2.3	Dialog Management	43
3.2.4	Message Generator	44
3.3	Proposed Security Framework	44
3.3.1	Proposed Data Leak Protection Model	47
3.3.2	Proposed Over-learning Protection Model	49
3.3.3	Proposed Security Framework	52
4	Experimental Results	53
4.1	Introduction	53
4.2	Tests Methodology	53
4.3	Adopted Test Case	54
4.3.1	Brief Description	54
4.3.2	Chatbot Scope	54
4.3.3	HR Assistant Chatbot	56
4.3.4	Security Framework Dataset	57
4.4	Tests Results	57
4.4.1	First Round of Testing	58
4.4.2	Second Test	61
4.4.3	Third Test	63
4.4.4	Forth Test	65
4.4.5	Fifth Test	67
4.4.6	Sixth Test	68
4.4.7	Seventh Test	71
4.5	Summary	73
5	Conclusion and Future work	74
5.1	Conclusion	74
5.2	Future Works	74
5.2.1	Develop Module to feed the framework with new threats	77
	References	78

List of Figures

2.1	Artificial Intelligence VS Machine Learning	19
2.2	Support Vector Machine	21
2.3	Deep Learning Architecture	22
2.4	The History of The Chatbots	24
2.5	High-level Chatbot Architecture	25
2.6	Different Chatbots General Types	26
2.7	AI Chatbots Types	28
2.8	Oracle Survey Results	30
3.1	Intelligent Chatbot Architecture	42
3.2	Architecture of the Proposed Security Framework	45
3.3	Architecture of Data Leak Protection Model	48
3.4	Architecture of Over-learning Protection Model	50
4.1	Screenshot of Cornell Movies-Dialogs Corpus	55
4.2	Screenshot of conversation with chatbot	55
4.3	Sample of Employees data for HR assistance chatbot	56
4.4	Conversation with HR assistant chatbot	57
4.5	Sample of Security Framework Dataset	57
4.6	Neural network Architecture for First Round	59
4.7	Accuracy Report	60
4.8	Neural network Architecture for Second Round	61
4.9	Accuracy Report	62
4.10	Neural network Architecture for Third Round	63
4.11	Accuracy Report	64
4.12	Neural network Architecture for Fourth Round	65
4.13	Accuracy Report	66
4.14	Neural network Architecture for Fifth Round	67
4.15	Accuracy Report	68
4.16	Accuracy Report	70
4.17	Accuracy Report	72
4.18	Test Results	73

List of Tables

4.1	Test Parameters of First Round of Testing	58
4.2	Results of First Round of Testing	59
4.3	Test Parameters of Second Round of Testing	61
4.4	Results of Second Round of Testing	62
4.5	Test Parameters of Third Round of Testing	63
4.6	Test Results of Third Round of Testing	64
4.7	Test Parameters of Fourth Round of Testing	65
4.8	Test Results of Fourth Round of Testing	66
4.9	Test Parameters of Fifth Round of Testing	67
4.10	Test Results of Fifth Round of Testing	68
4.11	Test Parameters of Sixth Round of Testing	69
4.12	Test Results of Sixth Round of Testing	69
4.13	Test Parameters of Seventh Round of Testing	71
4.14	Test Results of Seventh Round of Testing	71