A Chatbot Framework for the Children's Legal Centre

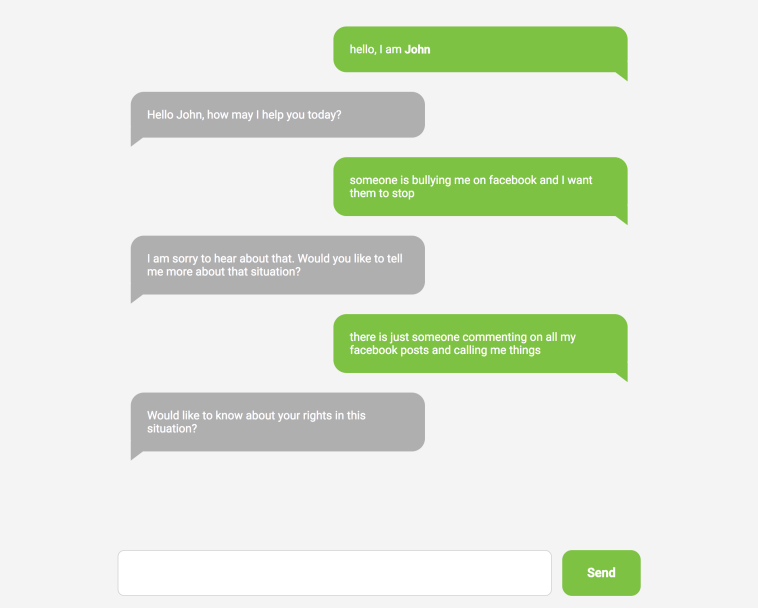
Jay Morgan[1], Adeline Paiement[1,3], Monika Seisenberger[1], Jane Williams[2], Adam Wyner[2]

[1] Department of Computer Science, Swansea Univeristy

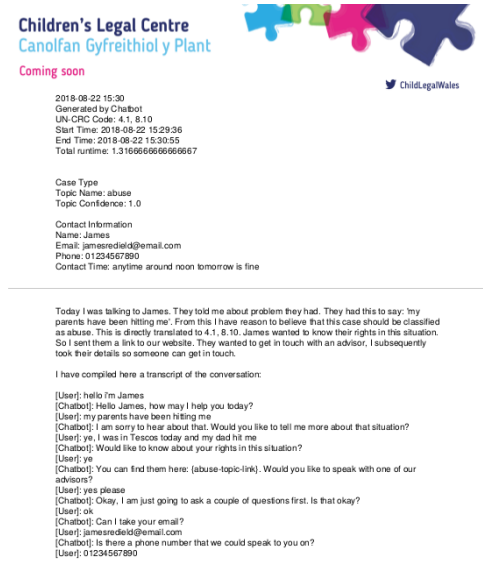
[2] College of Law, Swansea Univeristy

[3] Laboratoire LIS, Université de Toulon

Introduction



(a) Example conversation

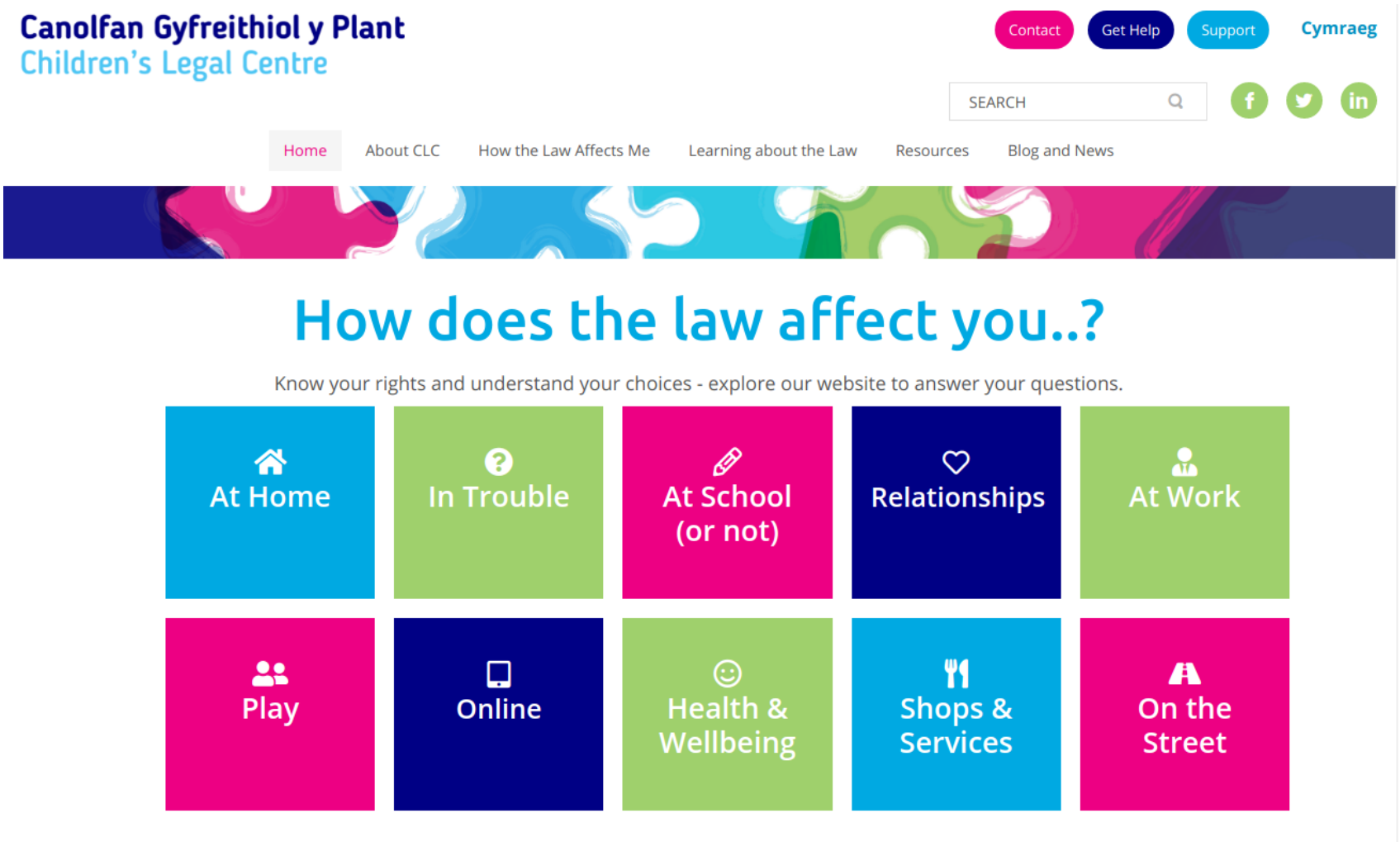


(b) Example Report

- A chatbot for the Children's Legal Centre Wales.
- Combination of machine learning and information extraction.

Introduction

Children's Legal Centre Wales[1]



[1] https://childrenslegalcentre.wales/

Overview

1. Construction of the corpus
   - Structure
   - Linguistic Analysis
2. Methodology:
   - Dialogue Graph
   - Classification of the messages with a Neural Network
   - Named Entity Recognition
3. Evaluation Methods
   - Neural Network Classification Performance
   - Explainability
   - User Studies
4. Concluding Remarks

Corpus

Layout of the Corpus

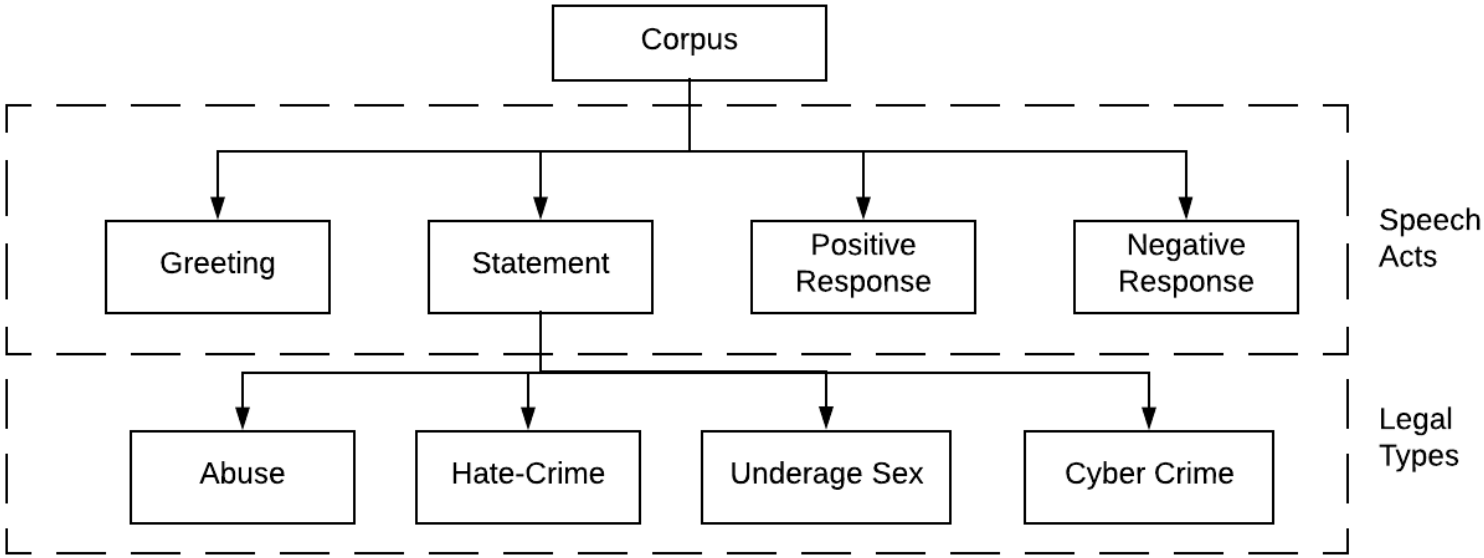


Figure: Hierarchical representation of the corpus

- With the CLC, a corpus of messages has been created.
- Annotated with a intent (speech act) of the message, and the legal situation being described.
- Made of artificial (made by the CLC), and real statements (extracted from a user study).

Corpus

Linguistic Analysis of the Corpus

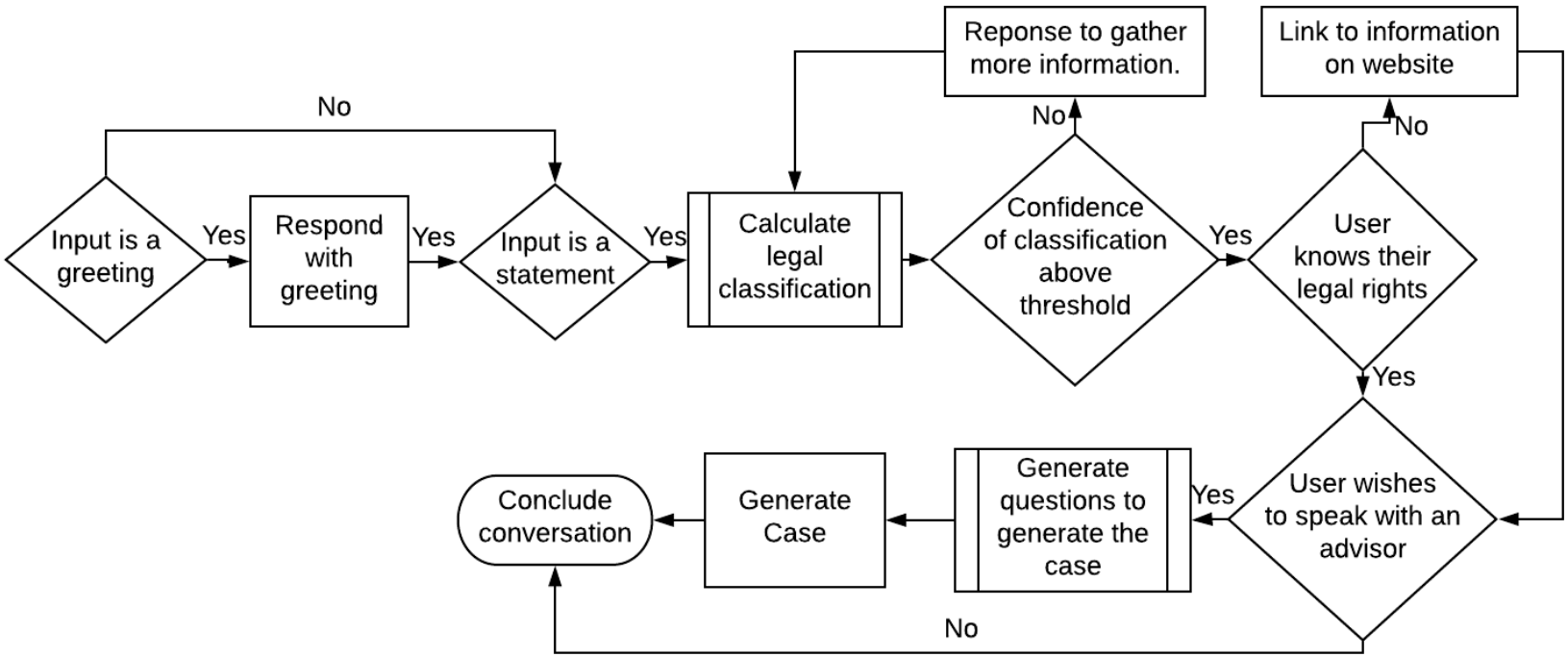
	NS		LR		NT		WT		AL (Std)	
Speech Act	**A**	**R**	**A**	**R**	**A**	**R**	**A**	**R**	**A**	**R**
Greeting	153	36	9.89	4.00	633	172	64	43	17.46 (13.31)	19.14 (9.42)
Statement	467	148	12.22	6.30	4642	2174	380	345	45.35 (16.63)	65.53 (42.68)
Positive	151	24	8.23	1.82	255	60	31	33	6.97 (4.15)	9.52 (11.81)
Negative	143	17	7.26	2.56	385	92	53	36	11.11 (7.63)	20.59 (15.94)
Legal Type	**A**	**R**	**A**	**R**	**A**	**R**	**A**	**R**	**A**	**R**
Abuse	187	46	9.43	3.57	1706	550	181	154	40.72 (15.88)	56.09 (38.84)
Hate-crime	78	19	7.08	2.80	913	227	128	81	55.23 (14.79)	53.00 (41.48)
Cyber-Crime	105	15	5.84	2.80	998	319	171	114	45.33 (16.90)	102.20 (35.50)
Underage Sex	97	54	7.59	6.32	1025	973	135	154	46.35 (15.48)	76.91 (41.33)

NS Number of Statements - Total number of statements for the classification type.
LR Lexical Richness - Number of unique tokens that occur throughout the classification type.
NT Number of Tokens - Number of non-unique tokens that occur.
WT Word Types - Number of unique tokens that occur.
AL Average Sentence Length - Average sentence length by token (standard deviation of average).

Table: Analysis of artificial and real statements for both speech acts and legal types.

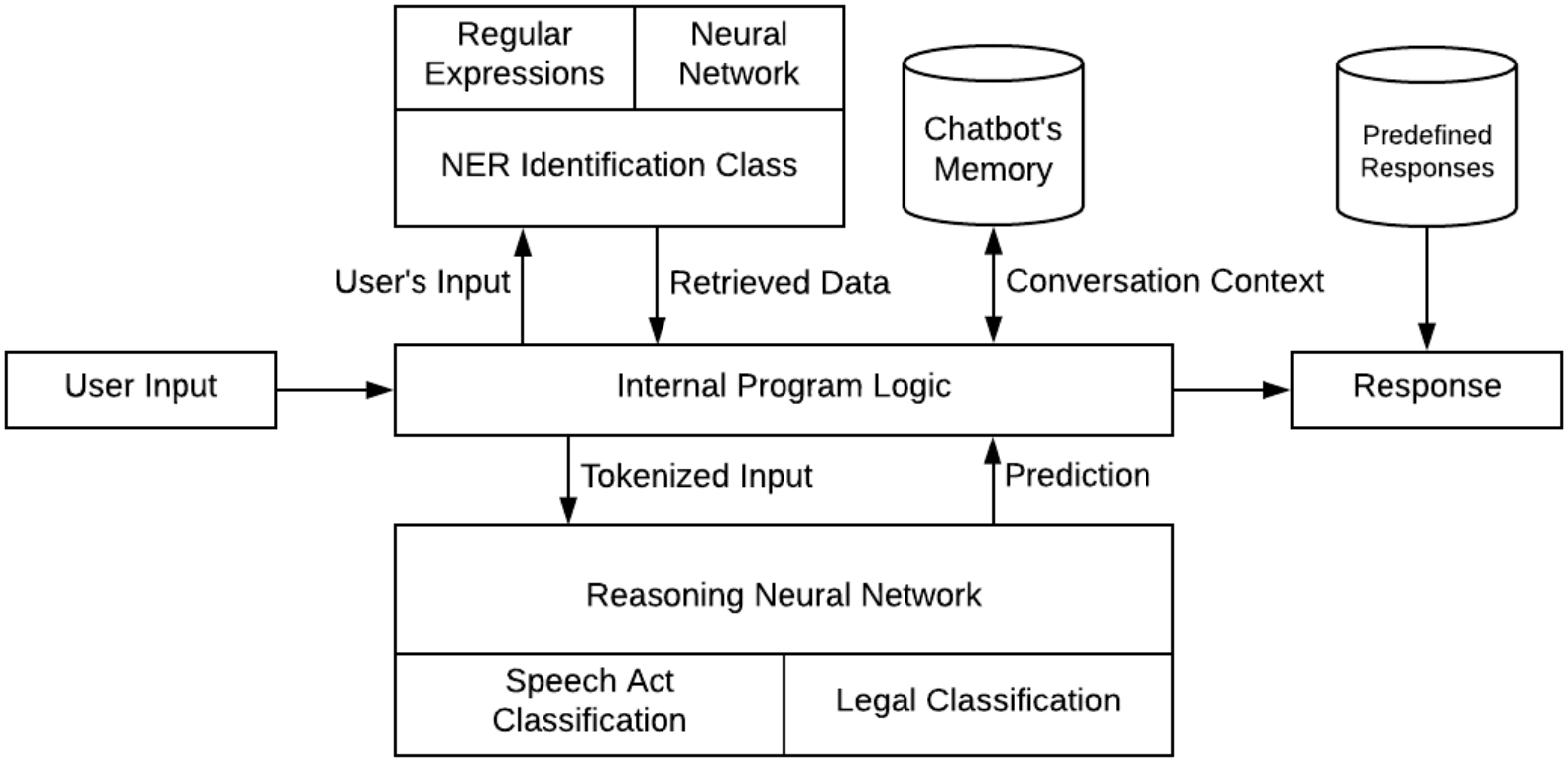
Methodology

Overview - Dialogue Graph



Methodology

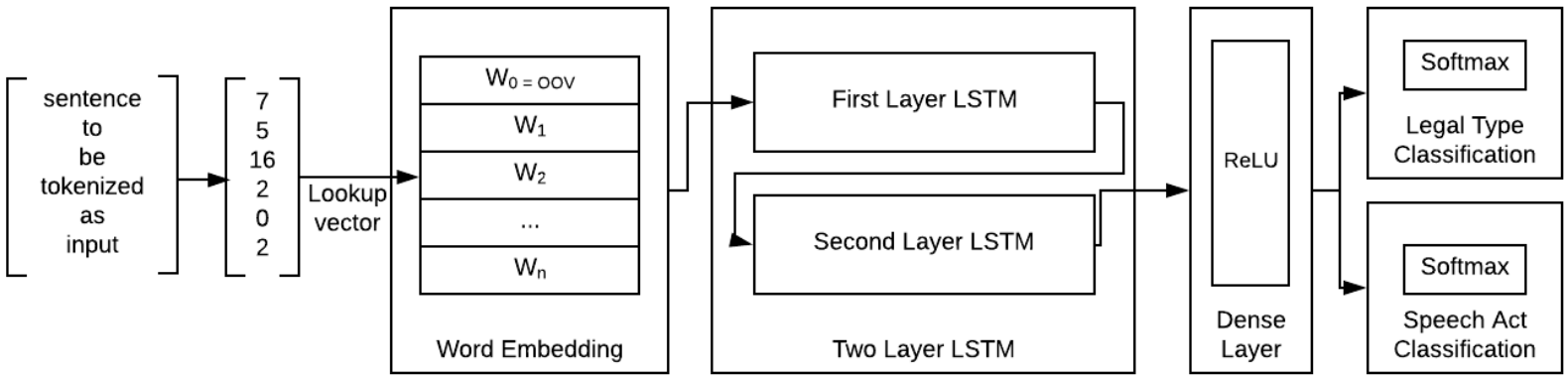
Overview - System Architecture



- Internal Programming Logic consults:
  - Prediction of Speech Act, Legal Type made by the Neural Network.
  - Information that has been found naturally through the conversation.

Methodology

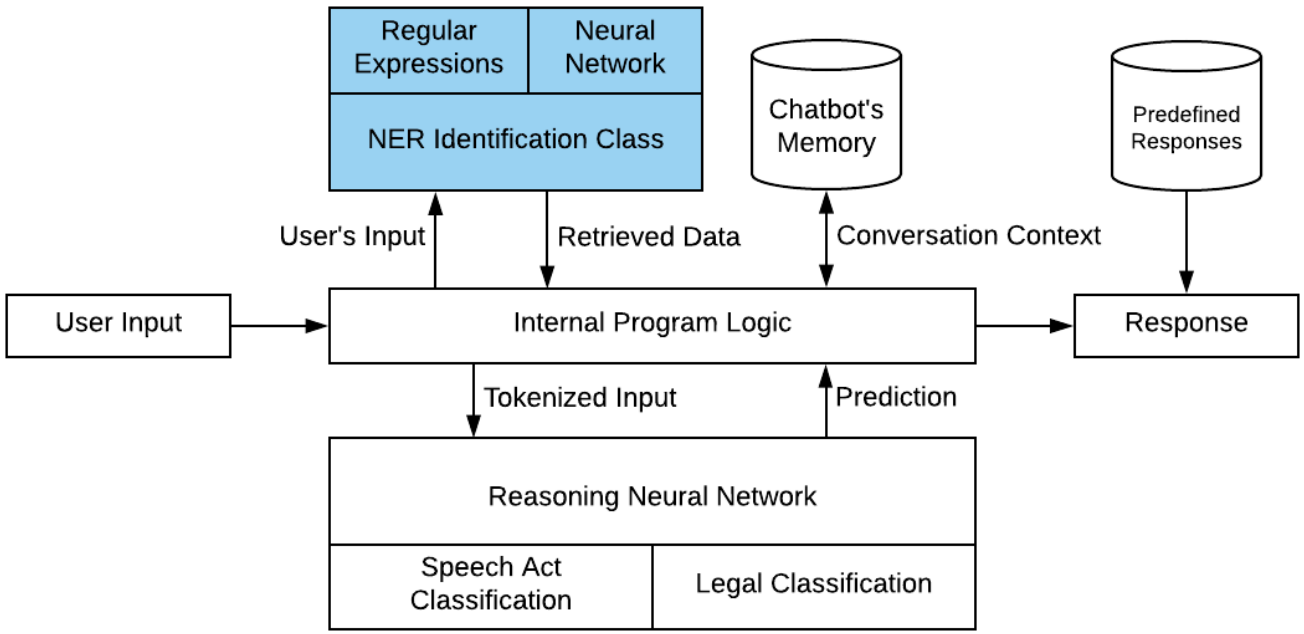
Classification of the User's Message



- Deep Learning:
  - Messages are tokenized and converted to word vectors.
  - Passed through two recurrent layers.
  - Classified by two parallel dense layers as a speech act and legal type.

Methodology

Named Entity Recognition



- Discover information naturally through the conversation before the case needs to be created.
- Two levels of information extraction:
  - Regular Expressions
  - Pretrained neural network

Evaluation Methods

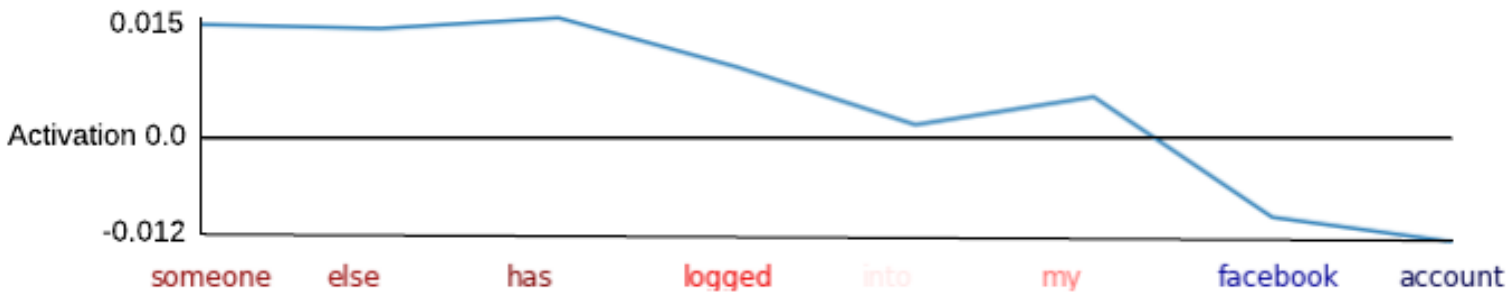
Neural Network Classification Performance

Network Type	**Speech Act F_1 Score**	**Legal Type F_1 Score**	**Avg F_1 score**
Dense Neural Network	97.36% (+/- 0.57%)	93.93% (+/- 1.21%)	95.65%
2 Layer RNN	95.16% (+/- 1.73%)	93.23 (+/- 1.24%)	94.20%
Pre-trained Embedding	98.41% (+/- 0.90%)	98.06% (+/- 0.35%)	98.24%
Training Corpus	**Speech Act F_1 Score**	**Legal Type F_1 Score**	**Avg F_1 Score**
Artificial Data Only	92.00%	92.88%	92.44%
Real Data Only	88.00%	83.99%	86.00%
Artificial & Real Data	98.41%	98.06%	98.24%

Table: Comparison between Dense and Recurrent Neural Network classification scores. Tests consist of a dense network with an untrained embedding layer; replacing the dense layers for a 2 recurrent layers; using a pre-trained embedding.

Evaluation Methods

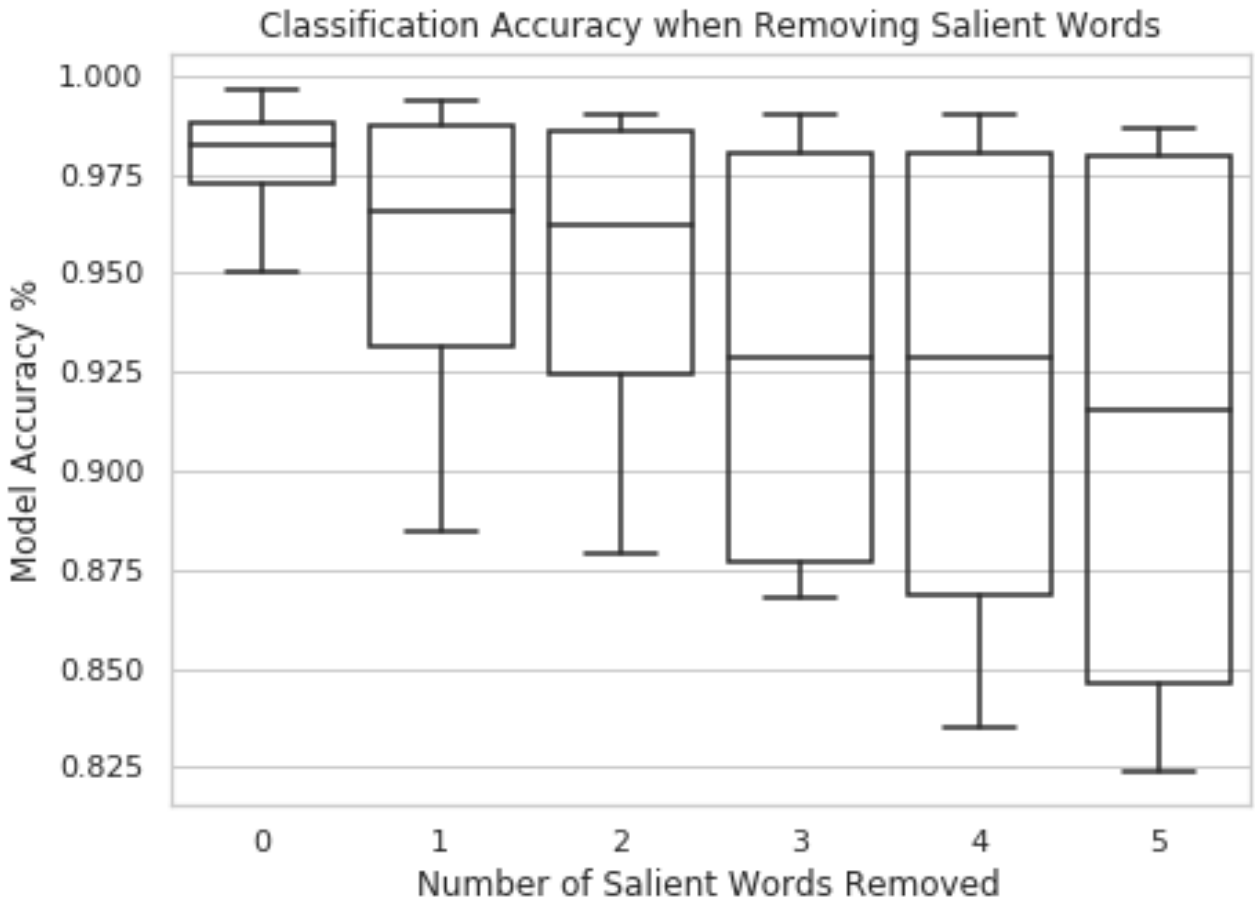
Explainability



- Defining saliency through LSTM cell state activation.
- Removal of salient words to attempt to reduce accuracy.

Evaluation Methods

Explainability



Evaluation Methods

User Studies

User Study Measure	Minimum	Maximum	Average
Ease of Use (Q1, Q2)	6	10	7.42
Interaction Performance (Q4, Q8)	6	9	6.71
Politeness & Responses (Q6, Q7)	7	10	7.57
Perceived Understanding (Q3, Q5)	3	10	5.00
Future Use (Q9, Q10)	4	10	8.29

Table: User study questionnaire responses.

- Group of 14 partipants interacted with the chatbot describing a relevant legal situation.
- After 3 conversations, a questionnaire was completed.

Concluding Remarks

- Created a chatbot framework.
- Prototyped developed with the Children's LegaL Centre Wales.
- Using a neural network that classifies both speect act and legal type.
- Extraction of information through the conversation to create a legal case with an advisor.
- Evaluated through both classification metrics and user studies.

Future Works

- Expansion into a larger set of legal types.
- Improvements in information extraction and representations.
- Rephrasing the user's response to present a more prosocial chatbot.