



# Smart Platforms for Science

*CERN openlab Technical Workshop 2019*

Taghi Aliyev

24/01/2019

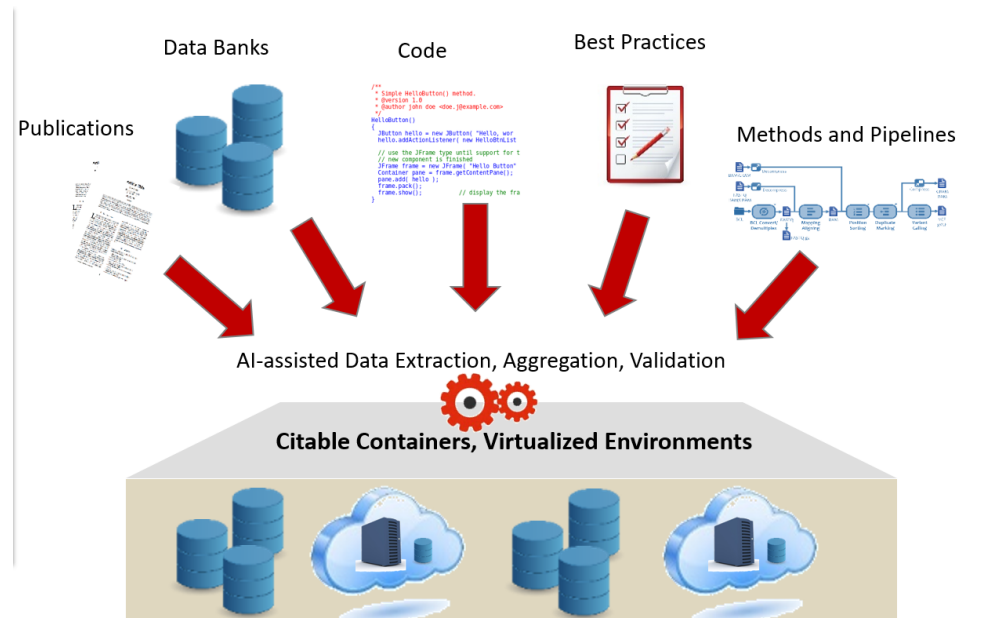
# Background and Motivation

## *Natural Language Processing Tools*

- A lot of replicative work in any scientific field
  - Non-reproducible research
  - Many different data structures and conventions --> Need for parsers...
- High barriers to enter the research fields
- Lack of common ground, all-in-one environments
- Sparked out off discussion with the members of Medical Community
  - Genomics Analysis Experts, Professors in Bio-Informatics, personal experiences

# Introduction to the Platform

- Large-scale collaborative research platform
- Main focus on ease-of-use, reproducibility of research
- Use of Machine Learning for Narrative interfaces
  - Information Retrieval
  - Natural Language Processing (Chatbots)
- Provide and host in-house solutions and projects



# Natural Language Processing

## *Chatbots and Information Retrieval*

- Lower the barriers for junior researchers
- Enhance the way research is done for everyone
- Chatbots as Personal Assistants
- Information Retrieval and Question Answering:

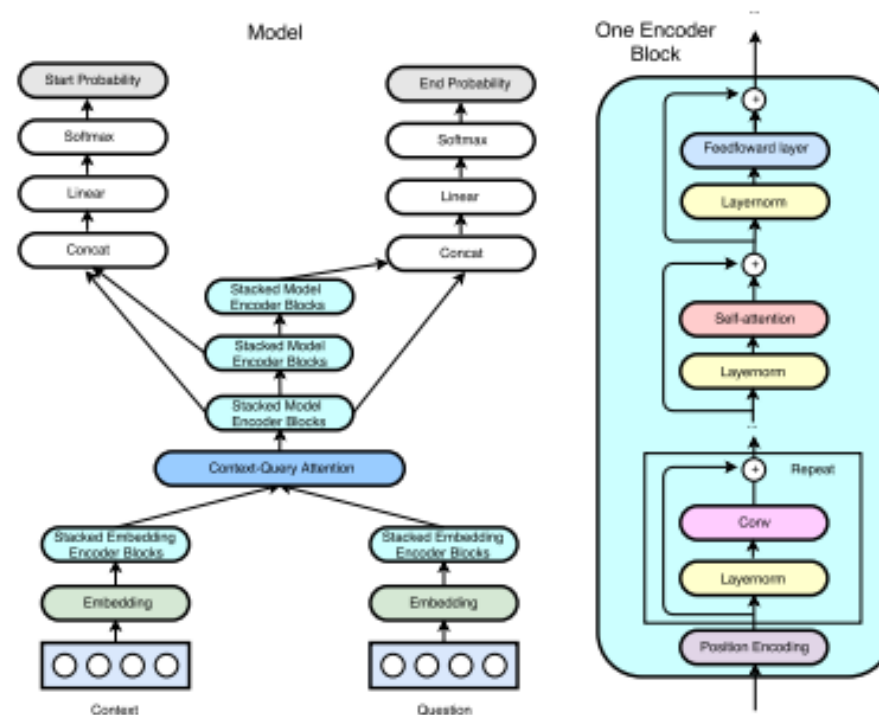
# Natural Language Processing

## *Models and Frameworks*

- Models being tested:
  - QANet
  - DSSM (Deep Semantic Similarity Models)
  - Recently released: BERT (Bidirectional Encoder Representations from Transformers)
- Framework to host the models:
  - RASA

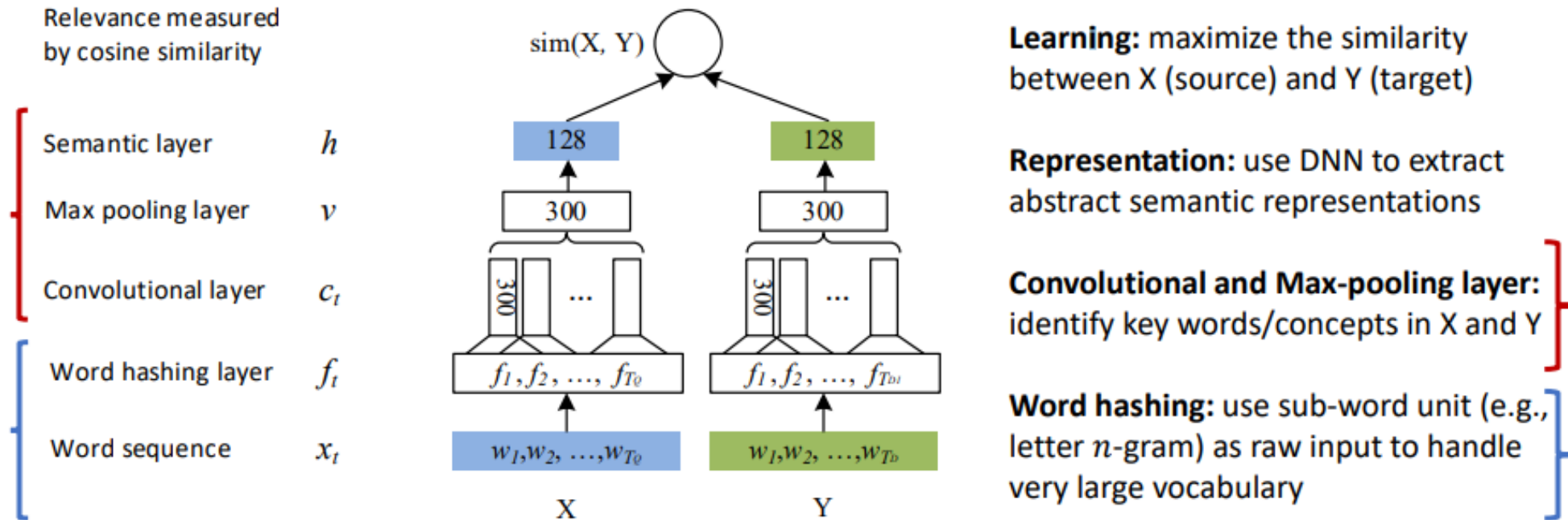
# Natural Language Processing

*Models – QANet; Combining local conv with global self-attention*



# Natural Language Processing

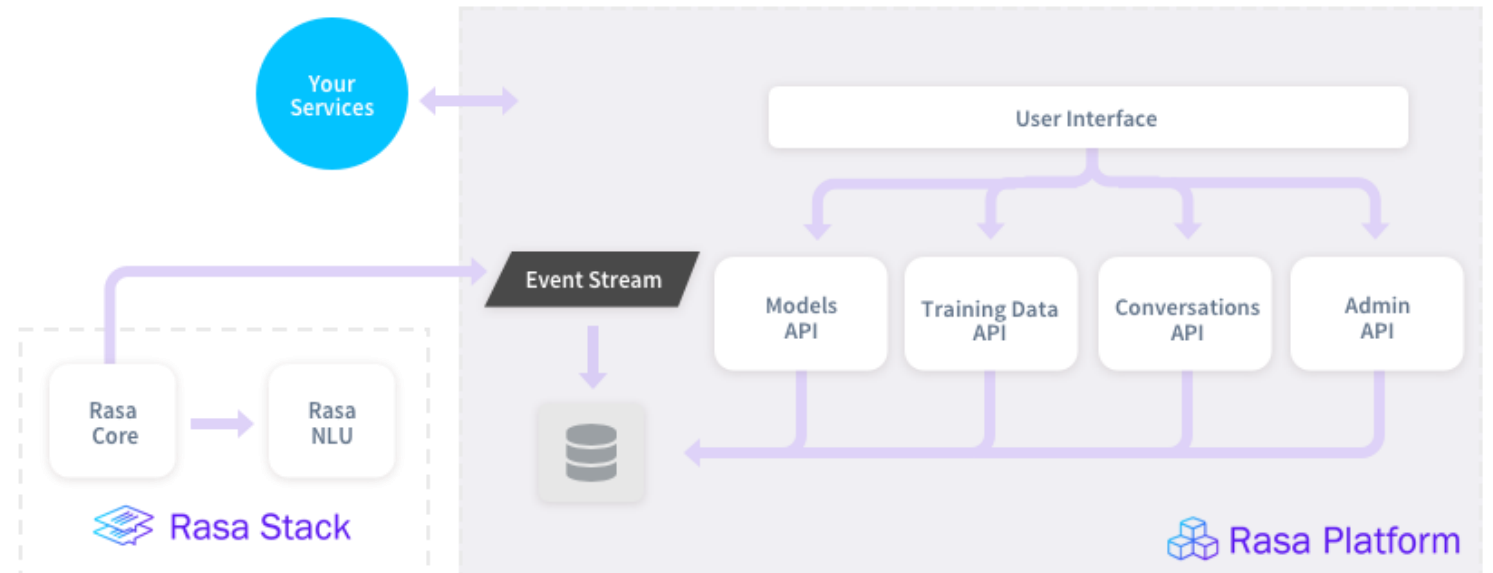
Model – DSSM; Deep Semantic Similarity Model



# Natural Language Processing

*Hosting Tool – RASA; Open Source tools for contextual AI Assistants*

- Python-based tool
- Allows for custom actions
  - Easing the integration of pre-trained models





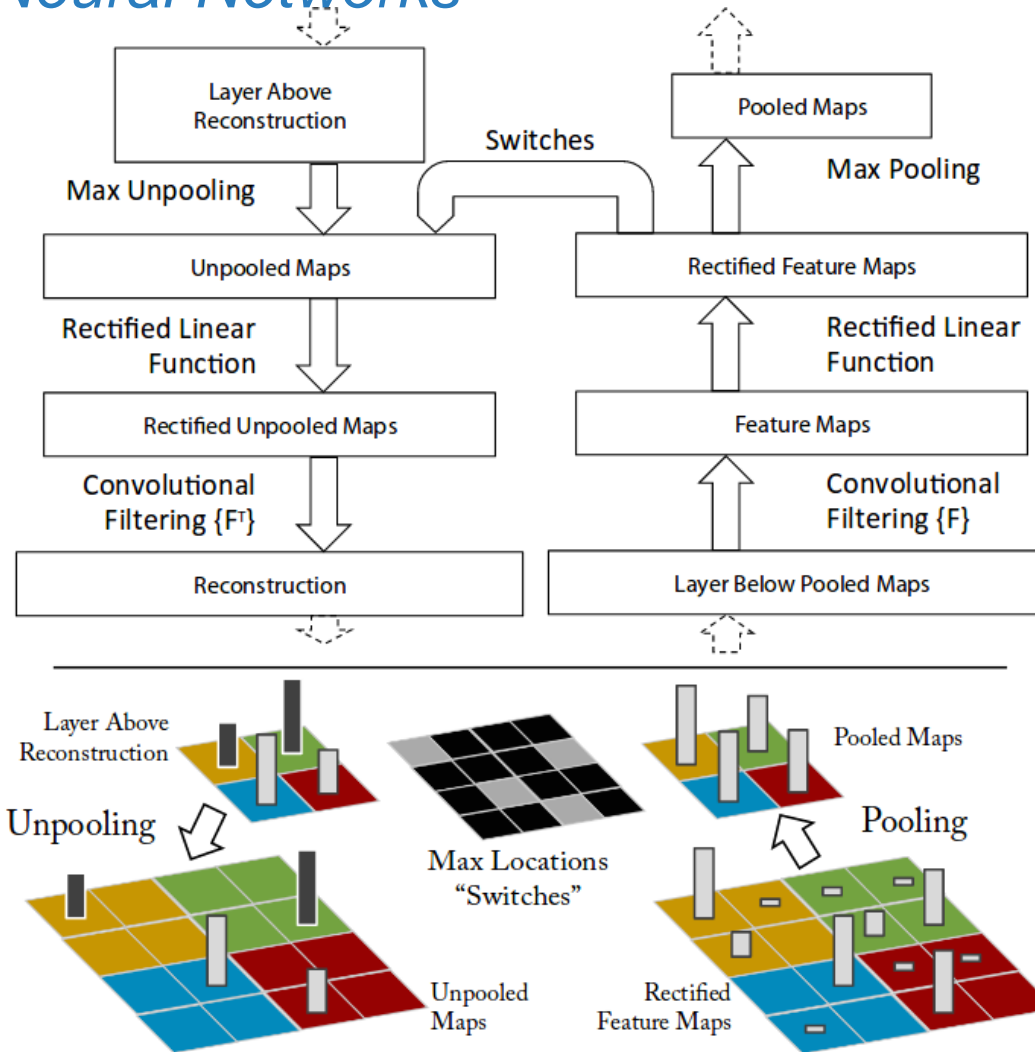
# Natural Language Processing

*Holding the models accountable and explainability*

- Understanding the reasoning and decision-making is crucial
- Not very straight-forward for deep neural networks
- More relevant for a conversational bot
  - Holding the model responsible when leads to accidents
  - Ability to back trace the effects and the outcome
- Initial test case:
  - TwinsUK with KCL for feature extraction in heritability studies
  - Pre-trained CNN

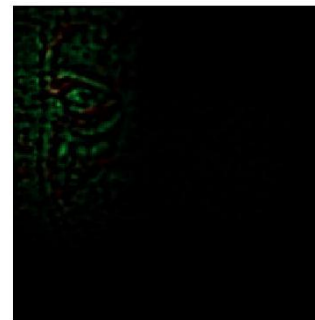
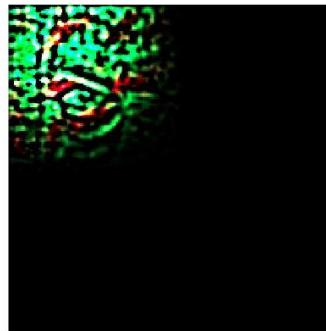
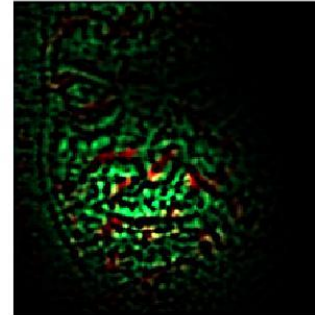
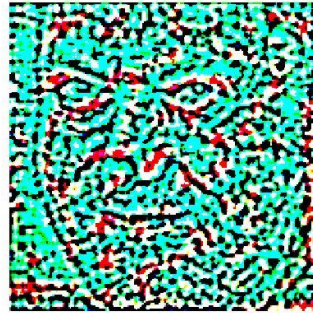
# Natural Language Processing

## Deconvolutional Neural Networks



# Deconvolution

*Some initial results*

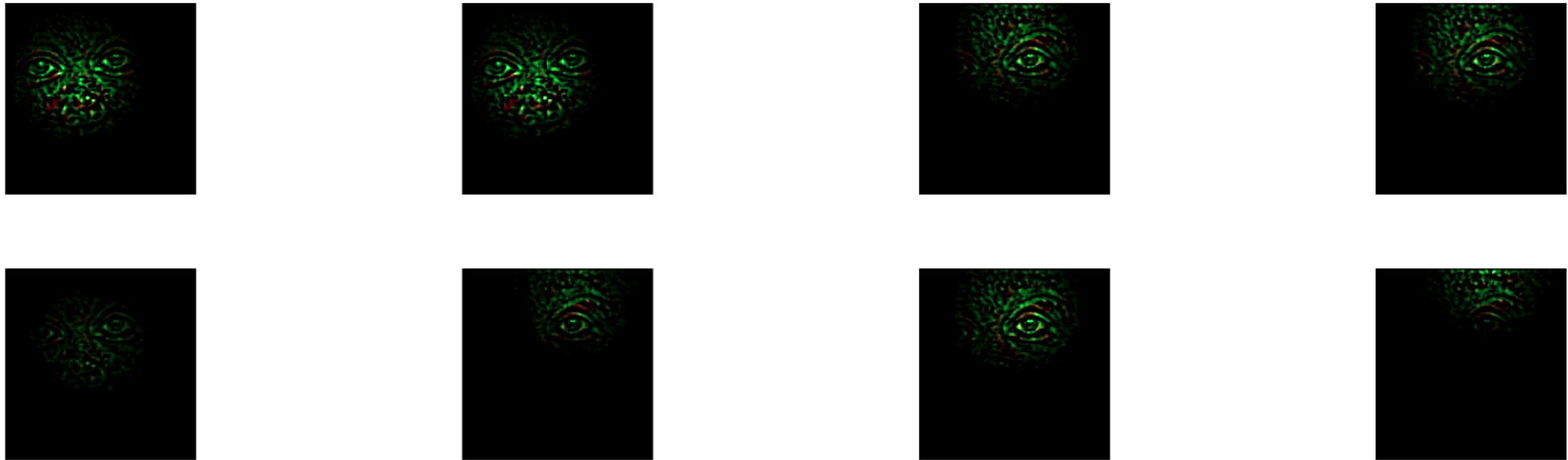


# Deconvolution

*Perturbation on input image and correlation*

- Results of initial tests on 2 twins
  - With 2 different ways to compute correlations

Results on 2 twins



# Deconvolution

*Where do we stand now?*

- Last touches for the convolutional neural network
- Next: Generalization to different network architypes
  - Especially for the textual cases
- Not an investigated problem
  - Even more true in Medical Informatics

# Natural Language Processing

## Application Areas and Use Cases

- Public/Social
  - GENIAL, Geneva Responsive City Camp
- Research
  - SQuAD 2.0 Challenge
  - Vignette extraction and analysis
- Education
  - Training tools/Personal Assistant
  - Still looking for partners and use cases



# Conclusion

- Deconvolution:
  - An interesting idea that can be incorporated to the platform to provide insights
- Conversational bots:
  - BERT proposes a generic and interesting approach
  - DSSM and QANet are proven to be of decent quality
  - Improvements are still required
- Use Cases:
  - GENIAL case being presented upcoming Monday at AMLD
  - Has interest of Canton of Geneva and a dedicated testing group