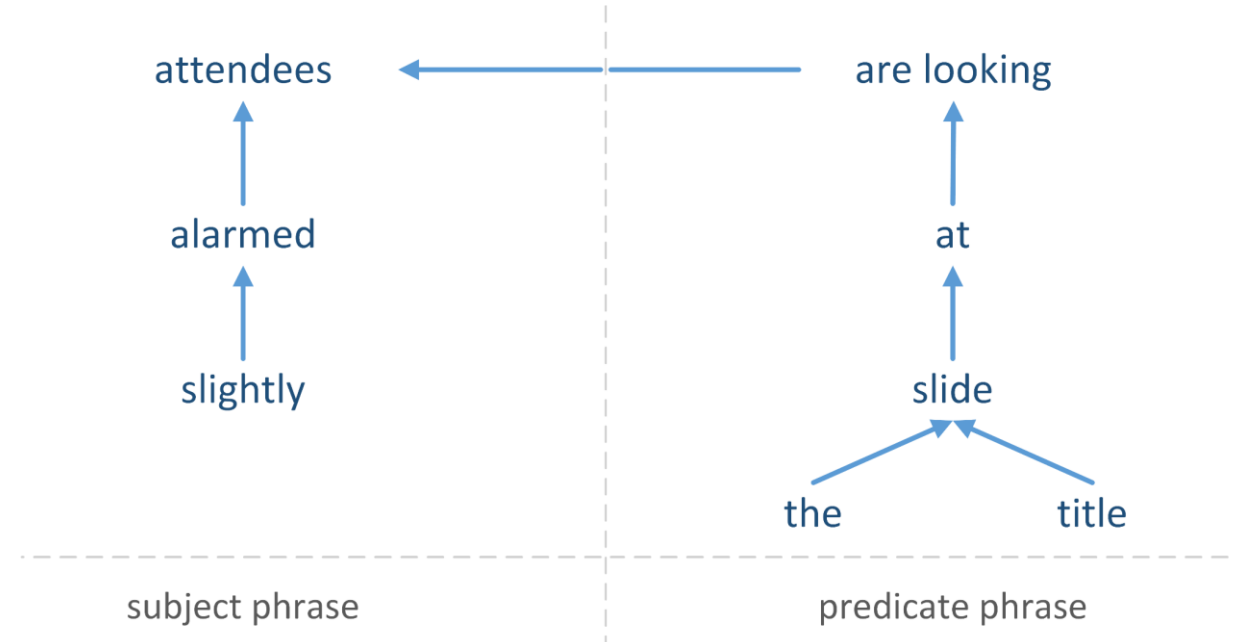


# Do androids read about electric sheep?

Machine reading comprehension algorithms

Marek Grzenkowicz

Roche Global IT Solution Centre



## About me



### Marek Grzenkiewicz

C64 BASIC → ... → ETL → Hadoop → **Python / Data analytics**

 [marek.grzenkiewicz@roche.com](mailto:marek.grzenkiewicz@roche.com)

 [@chopeen](https://twitter.com/chopeen)



### Roche Global IT Solution Centre (RGITSC)

Offices in Warsaw and Poznan

 [it.roche.pl](https://it.roche.pl)

 [instagram.com/it.roche.pl](https://instagram.com/it.roche.pl)

# Agenda



1. Natural language processing
  - What is it?
  - Why is it important?
  - Basic NLP operations [demo] 🙌
2. My vision
3. Do androids read?
  - Fundamental skills
  - The state of research
4. Summary

# **Reading between the lines**

# Reading between the lines

*How does it work?*



A sentence or a short paragraph of text.

Option 1

(raise your **left hand**)

Option 2

(raise your **right hand**)

## Reading between the lines



### *Bob? Who's Bob?*

My friend Alice is a great Java developer.  
Unfortunately, Bob isn't.



Bob is not my friend

A blue speech bubble with a tail pointing towards the bottom left.

Bob is not a great  
developer

An orange speech bubble with a tail pointing towards the bottom right.

## Reading between the lines

*Charlie is 6*

Roche

Charlie can already write, so he must be at least 6!

6 six years old

6 feet tall

## Reading between the lines

*Charlie is 6*

Roche

Charlie plays basketball, so he must be at least 6!

6 six years old

6 feet tall



## Reading between the lines



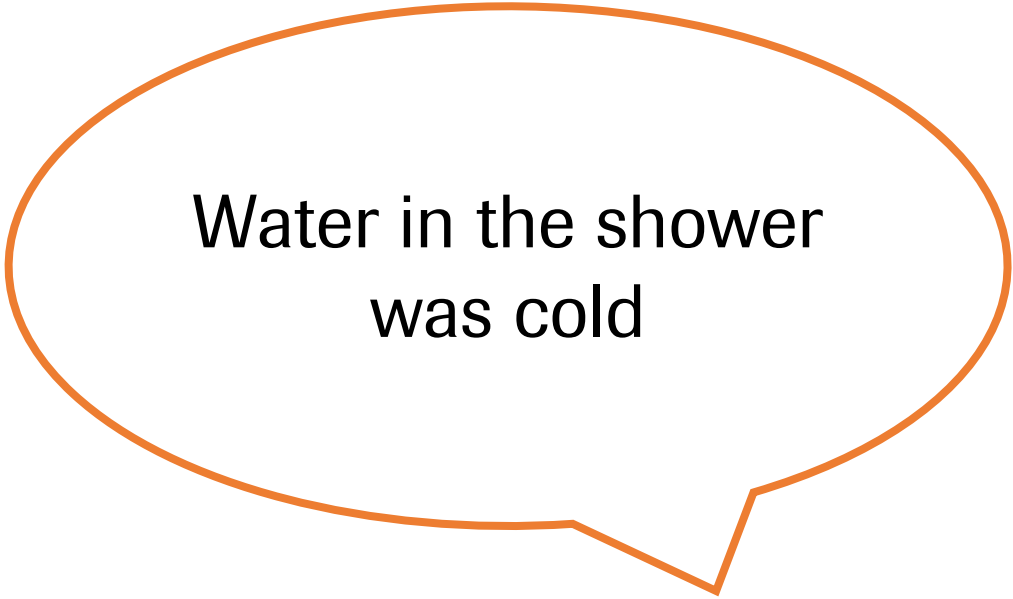
### *How's the water?*

**Chuck:** How did you like the Canary Islands?

**Frank:** It was nice – sunny, not too humid.  
The water was a bit cold though.



Water in the ocean  
was cold



Water in the shower  
was cold

## Reading between the lines

*Nothing is certain*



Let me tell you how it will be  
There's one for you, nineteen for me  
'Cause I'm the taxman, yeah, I'm the taxman  
The Beatles - Taxman

The Beatles were  
required to pay a  
95% tax

Before 1971, 1 pound = 20 shillings  
(and 1 shilling = 12 pence)

Altogether there  
were 20 forms to be  
filled

# Reading between the lines

*Album 1, side 1, track 1*



She was born in November 1963

The day Aldous Huxley died

Sheryl Crow – Run Baby Run

## Aldous Huxley

From Wikipedia, the free encyclopedia

**Aldous Leonard Huxley** (26 July 1894 – 22 November 1963) was an English writer and philosopher.<sup>[1][2][3][4]</sup> He wrote nearly fifty books<sup>[5][6]</sup>—both novels and non-fiction works—as well as wide-ranging essays, narratives, and poems.

## John F. Kennedy

From Wikipedia, the free encyclopedia

(Redirected from [John Fitzgerald Kennedy](#))

"*JFK*", "*John Kennedy*", and "*Jack Kennedy*" redirect here. For other uses, see *JFK (disambiguation)*, *John Kennedy (disambiguation)*.

**John Fitzgerald Kennedy** (May 29, 1917 – November 22, 1963), often referred to by initials **JFK** and **Jack**, was an American politician who served as the 35th [president of the United States](#) from January 1961 until his

# **Natural language processing**

# Natural language processing

## *Selected areas and tasks*



### Syntax

Lemmatization  
Stemming

POS tagging  
Text segmentation

### Speech

Speech-to-text  
(recognition)

Text-to-speech  
(generation)

### Semantics

Machine translation  
Optical character recognition (OCR)

Natural language understanding (NLU)  
Natural language generation (NLG)

Named entity recognition (NER)  
Sentiment analysis

Topic modelling  
Word sense disambiguation

# Natural language processing

## *Publication overload*



*Stud Health Technol Inform.* 2019 Aug 21;264:1586-1587. doi: 10.3233/SHTI190547.

### **Named Entity Recognition in Prehospital Trauma Care.**

Silverman GM<sup>1,2</sup>, Lindemann EA<sup>3</sup>, Rajamani G<sup>4</sup>, Finzel RL<sup>5</sup>, McEwan R<sup>1</sup>, Knoll BC<sup>2</sup>, Pakhomov S<sup>5</sup>, Melton GB<sup>2,3</sup>, Tignanelli CJ<sup>2,3,6</sup>.

#### **– Author information**

- 1 Academic Health Center - Information Systems, University of Minnesota, Minneapolis, Minnesota, USA.
- 2 Institute for Health Informatics, University of Minnesota, Minneapolis, Minnesota, USA.
- 3 Department of Surgery, University of Minnesota, Minneapolis, Minnesota, USA.
- 4 Stanford University, Stanford, CA, USA.
- 5 College of Pharmacy, University of Minnesota, Minneapolis, Minnesota, USA.
- 6 Department of Surgery, North Memorial Health Hospital, Robbinsdale, Minnesota, USA.

#### **Abstract**

Natural language processing (NLP) methods would improve outcomes in the area of prehospital Emergency Medical Services (EMS) data collection and abstraction. This study evaluated off-the-shelf solutions for automating labelling of clinically relevant data from EMS reports. A qualitative approach for choosing the best possible ensemble of pretrained NLP systems was developed and validated along with a feature using word embeddings to test phrase synonymy. The ensemble showed increased performance over individual systems.

**KEYWORDS:** Emergency medical services; Labeling; Natural language processing

PMID: 31438244 DOI: [10.3233/SHTI190547](https://doi.org/10.3233/SHTI190547)

[Indexed for MEDLINE]

Source: <https://www.ncbi.nlm.nih.gov/pubmed/31438244>

# Natural language processing


## *Publication overload*



**5 000+** life sciences and biomedical journals

**813 598** citations in FY 2017

**2 000+** articles daily



Only a fraction  
of available  
knowledge and  
data is used

### Sub-optimal results

- Research
- Drug development
- Patient treatment
- Software development

Source: [https://www.nlm.nih.gov/bsd/index\\_stats\\_comp.html](https://www.nlm.nih.gov/bsd/index_stats_comp.html)

# Natural language processing

*Basic NLP operations [demo]*



spaCy



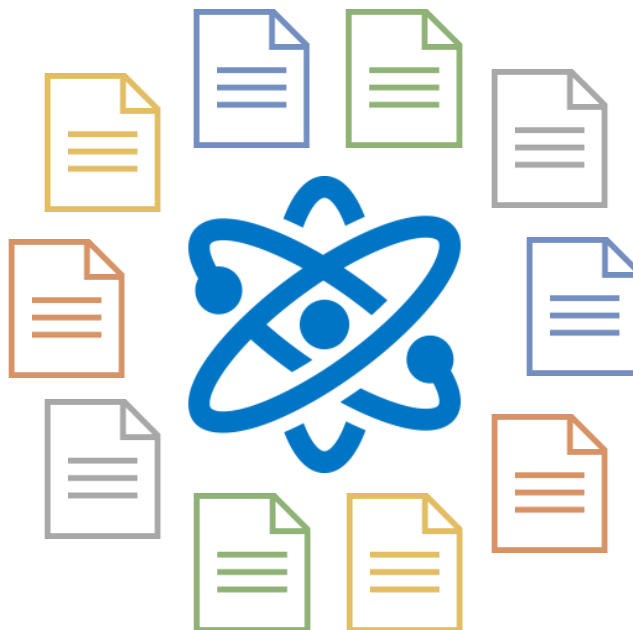
# **My vision**

# My vision

## *Tailored reading recommendations*



High-level  
requirements for  
a novel project



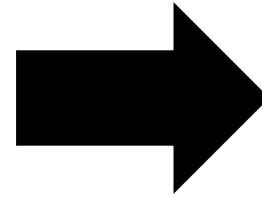
1000's of articles and  
research papers



Recommended  
materials

## My vision

### *Text summarization (Pharma)*



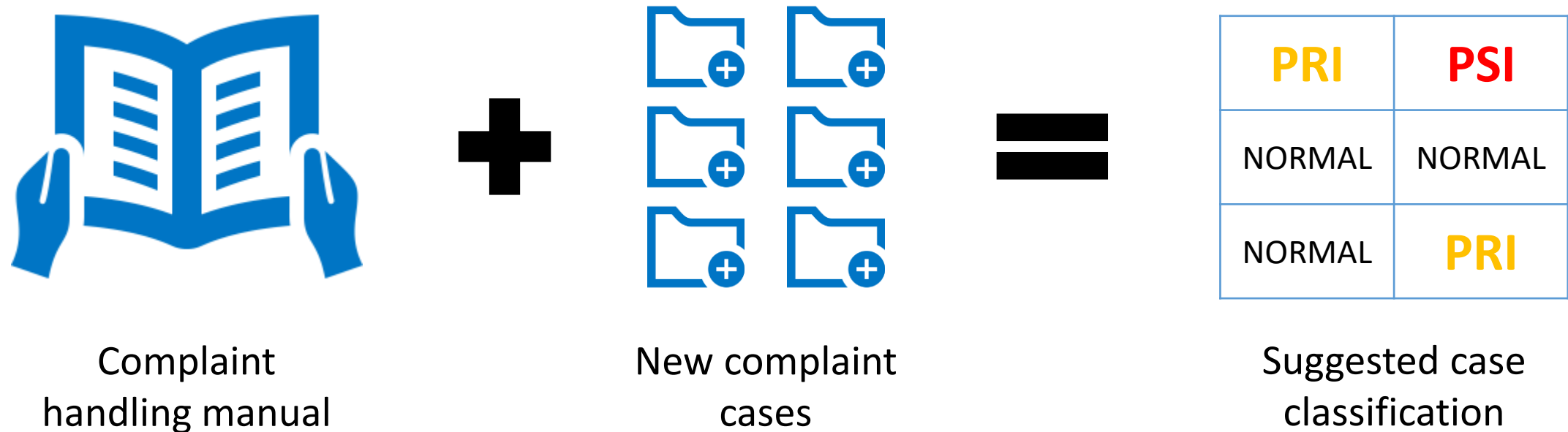
Executive summary slide

Audit protocols  
Observation notes  
FDA inspection forms

# My vision



## Decision support system for complaint handling (Diagnostics)



**PRI** = Potentially Reportable Incident, **PSI** = Potential Safety Issue

**Do androids read?**

# Reading comprehension

## *Definition*

Reading comprehension is the ability to:

1. process text,
2. understand its meaning,
3. integrate with what the reader already knows.

# Reading comprehension

## *Fundamental skills*



Knowing **meaning of words**

Understanding meaning of words from discourse context

Ability to identify the main thought of a passage

Ability to **answer questions** answered in a passage

Ability to recognize the **literary devices** used in a passage and determine its tone

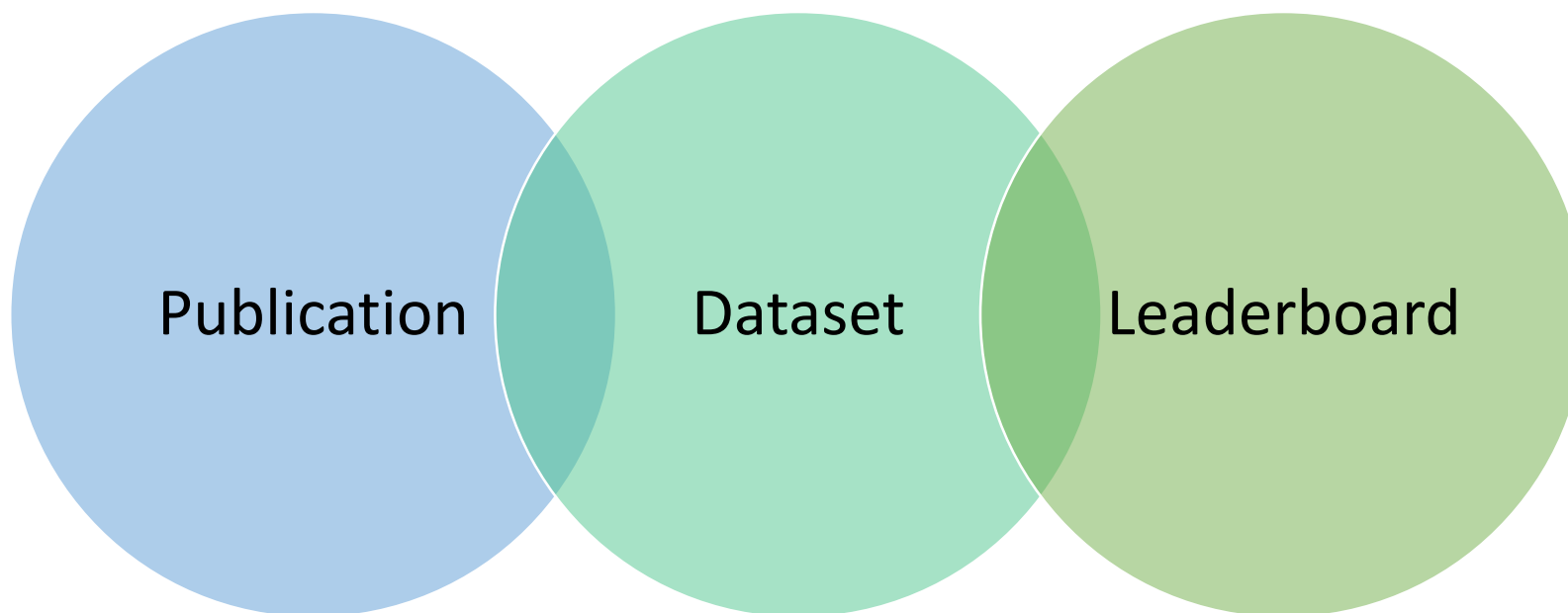
**Information integration**

- real-world knowledge
- common sense

# Reading comprehension

## *Current research*

1. Stanford Question Answering Dataset (SQuAD)
2. Reading Comprehension with Multiple Hops (QAngaroo)
3. NarrativeQA Reading Comprehension Challenge

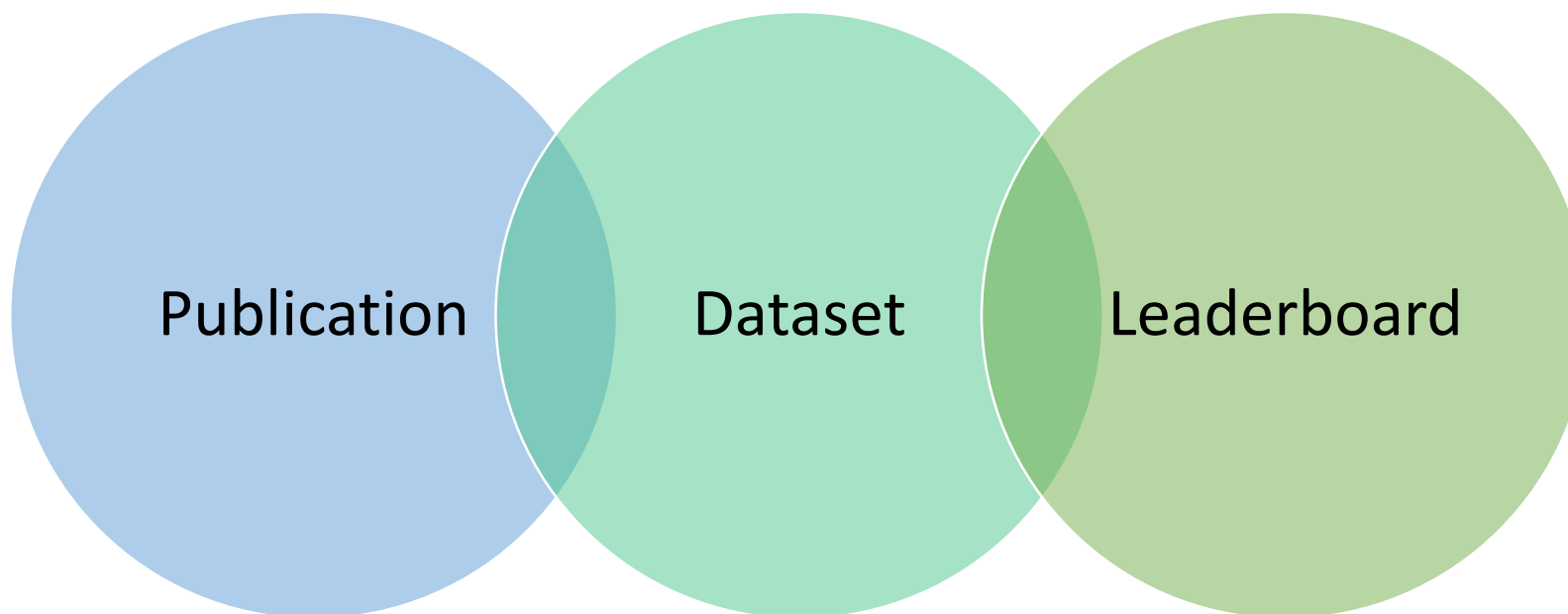




# Reading comprehension

## *Current research*

1. **Stanford Question Answering Dataset (SQuAD)**
2. Reading Comprehension with Multiple Hops (QAngaroo)
3. **NarrativeQA Reading Comprehension Challenge**



# SQuAD dataset



## **SQuAD v1.1**

- 100,000+ question-answer pairs
- extracted from 500+ Wikipedia articles

## **SQuAD v2.0**

- v1.1 combined with 50,000+ new, unanswerable questions

## *Example*

**Document:** The heat required for boiling the water and supplying the steam can be derived from various sources, most commonly from burning combustible materials with an appropriate supply of air in a closed space (called variously combustion chamber, firebox). In some cases the heat source is a nuclear reactor, geothermal energy, solar energy or waste heat from an internal combustion engine or industrial process. In the case of model or toy steam engines, the heat source can be an electric heating element.

**Question:** What is the usual source of heat for boiling water in the steam engine?

## *Example*

**Document:** The heat required for boiling the water and supplying the steam can be derived from various sources, most commonly from **burning combustible materials** with an appropriate supply of air in a closed space (called variously combustion chamber, firebox). In some cases the heat source is a nuclear reactor, geothermal energy, solar energy or waste heat from an internal combustion engine or industrial process. In the case of model or toy steam engines, the heat source can be an electric heating element.

**Question:** What is the usual source of heat for boiling water in the steam engine?

**Answer:** burning combustible materials



	Correct answers (exact match)
Human performance	86.83 %
Best model (Sep 2019)	89.73 %

Results for SQuAD v2.0

# SQuAD dataset



## *Limitations*

- ❌ Limited lexical and topical diversity
- ❌ The documents are short (i.e. paragraphs of text)
- ❌ Each answer is a span of text from a single document (→ QAngaroo)
- ❌ Shallow pattern matching may suffice to produce correct answers (→ NarrativeQA)

## NarrativeQA dataset



- 1,500+ documents
  - books (Project Gutenberg)
  - movie scripts (Internet Movie Script Database and other websites)
- 45,000+ question-answer pairs

# NarrativeQA dataset



## *Example*

### **Full story snippet:**

As he spins around one of the attackers jams all eight inches of his bayonet blade into Jacob's stomach. Jacob screams. It is a loud and piercing wail.

...

A doctor leans his head in front of the lamp and removes his mask. His expression is somber. He shakes his head. His words are simple and final.

DOCTOR: He's gone.

The doctor steps away. A nurse rudely pulls a green sheet over his head. The doctor turns to one of the aides and throws up his hands in defeat.

**Question:** What is the fatal injury that Jacob sustains which ultimately leads to his death?



# NarrativeQA dataset



## Example

### Full story snippet:

As he spins around one of the attackers **jams all eight inches of his bayonet blade into Jacob's stomach**. Jacob screams. It is a loud and piercing wail.

...

A doctor leans his head in front of the lamp and removes his mask. His expression is somber. He shakes his head. His words are simple and final.

DOCTOR: **He's gone.**

The doctor steps away. A nurse rudely pulls a green sheet over his head. The doctor turns to one of the aides and throws up his hands in defeat.

**Question:** What is the fatal injury that Jacob sustains which ultimately leads to his death?

**Answer:** a bayonet wound

# NarrativeQA dataset



## Results



	BLEU-1	BLEU-4	METEOR	ROUGE-L
Human performance	44.43	19.65	24.14	57.02
Best model (Dec 2017)	19.09	2.11	4.37	14.03

Results for models processing full stories

Results

		BLEU-1	BLEU-4	METEOR	ROUGE-L
✓ 🏆	Human performance	44.43	19.65	24.14	57.02
	Best model (Dec 2017)	19.09	2.11	4.37	14.03
✓ 🏆	Best model (Jan 2019)	48.70	20.98	21.95	54.74

Results for models processing full stories

# NarrativeQA dataset



## *Summary*

- ✗ No integration of external information required (every question can be answered using a single document)
- ✓ Decreasing performance gap between humans and machines – impressive progress made in less than two years

# Summary

## Reading comprehension

*Why is it hard?*

[...]

He was standing on a mountain top, it was freezing and he saw nothing but **whiteness**.

## Reading comprehension



*Why is it hard?*

[...]

He was standing on a mountain top, it was freezing and he saw nothing but whiteness.

The **teleportation** must have gone wrong! Was it because of the dream he had last night?

# Reading comprehension



## *Why is it hard?*

[...]

He was standing on a mountain top, it was freezing and he saw nothing but whiteness.

The teleportation must have gone wrong! Was it because of the dream he had last night? He dreamt about **climbing** the highest mountain in the **Alps**, wearing a **suit**. For the world, he could not remember the name of the mountain.



# Reading comprehension



## *Why is it hard?*

[...]

He was standing on a mountain top, it was freezing and he saw nothing but whiteness.

The teleportation must have gone wrong! Was it because of the dream he had last night? He dreamt about climbing the highest mountain in the Alps, wearing a suit. For the world, he could not remember the name of the mountain.

And then it **clicked in his head** - he knew where he was and it bad news!

## Reading comprehension



### *Why is it hard?*

[...]

He was standing on a mountain top, it was freezing and he saw nothing but whiteness.

The teleportation must have gone wrong! Was it because of the dream he had last night? He dreamt about climbing the **highest mountain in the Alps**, wearing a suit. For the world, he could not remember the name of the mountain.

And then it clicked in his head - he knew where he was and it bad news! The icy darkness around was the summit of              .

## Reading comprehension

*Do androids read about electric sheep?*

~~No.~~

Not yet.



Slides, code, papers, datasets, leaderboards, etc.

<https://go.roche.com/reading>

*Doing now what patients need next*