Machine Learning In Legal Tech - What's new?

Daniel Roythorne,
ThoughtRiver Ltd.

Outline

Outline

- What's old? - What are the origins of today's legal tech AI?

Outline

- What's old? - What are the origins of today's legal tech AI?
- What's new? - Is the hype around AI in legal tech well founded?

Outline

- What's old? - What are the origins of today's legal tech AI?
- What's new? - Is the hype around AI in legal tech well founded?
- How does it work in theory? - Machine learning (ML) techniques for contract review.

Outline

- What's old? - What are the origins of today's legal tech AI?
- What's new? - Is the hype around AI in legal tech well founded?
- How does it work in theory? - Machine learning (ML) techniques for contract review.
- What about in practice? - Engineering and integration.

What's old? - Neural Networks

Foundational idea, the nerual network, is from 1958.

Psychological Review
Vol. 65, No. 6, 1958

THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN[1]

F. ROSENBLATT

Cornell Aeronautical Laboratory

If we are eventually to understand the capability of higher organisms for perceptual recognition, generalization, recall, and thinking, we must first have answers to three fundamental questions:

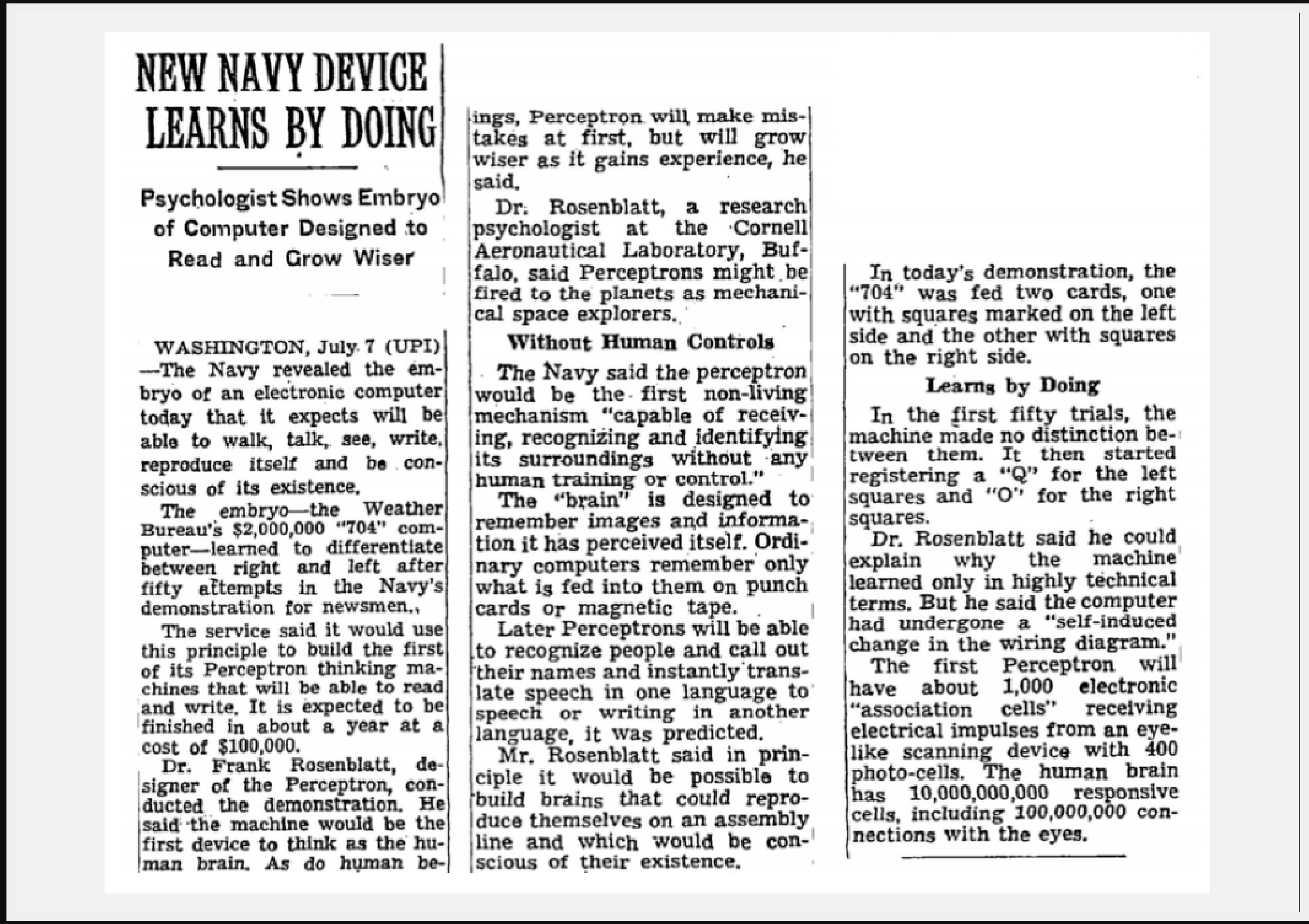
1. How is information about the

and the stored pattern. According to this hypothesis, if one understood the code or "wiring diagram" of the nervous system, one should, in principle, be able to discover exactly what an organism remembers by reconstructing the original sensory patterns from

Rosenblatt, F.(1958). 'The perceptron: A probabilistic model for information storage and organization in the brain', Psychological Review, 65(6), 386-408.

What's old? - Neural Networks

Grand promises being made in the same year.



NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI)—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's $2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human beings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

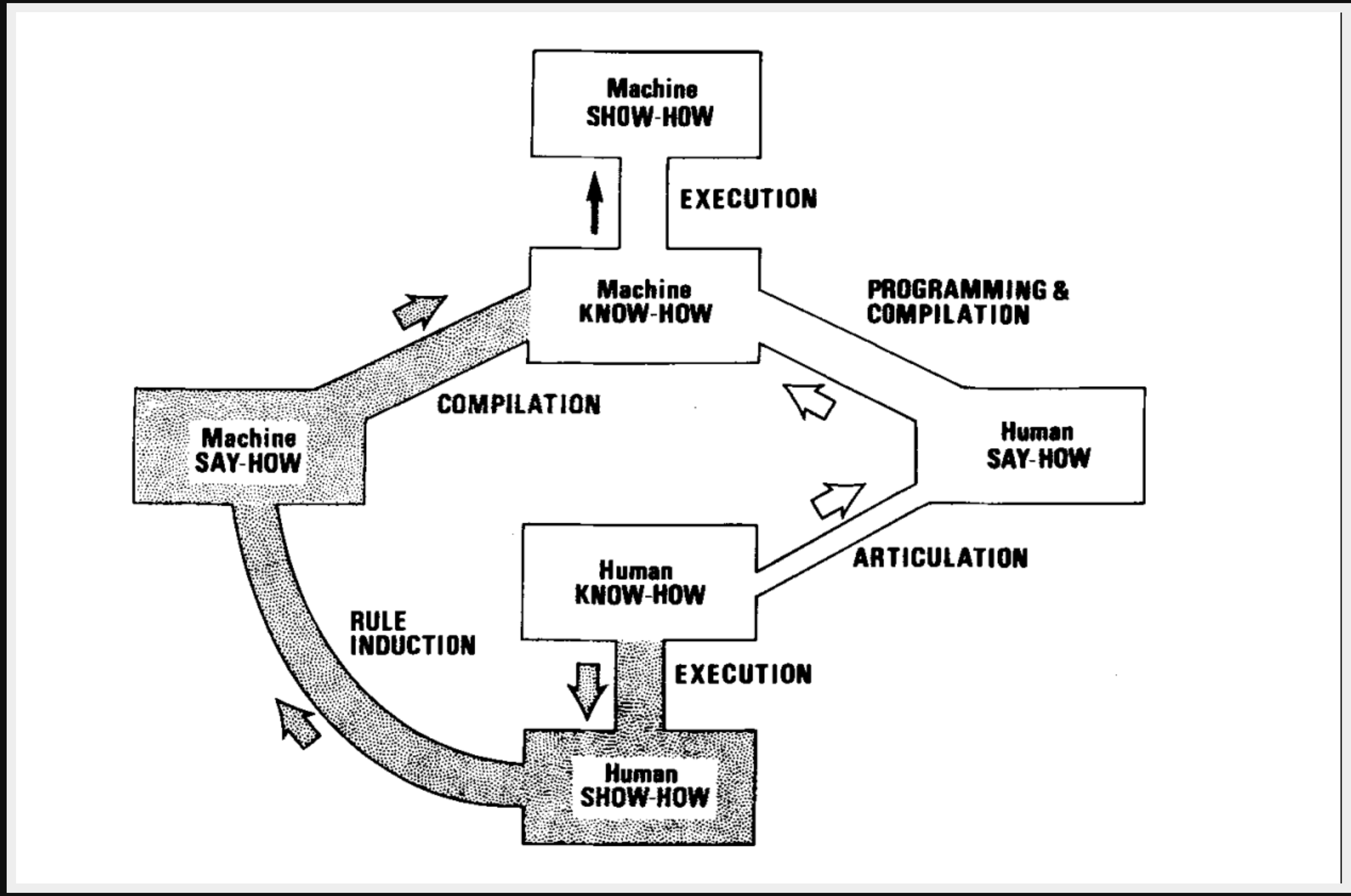
Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

(1958) 'New Navy Device Learns By Doing', New York Times, 8 July, p.25

What's old? - Expert systems

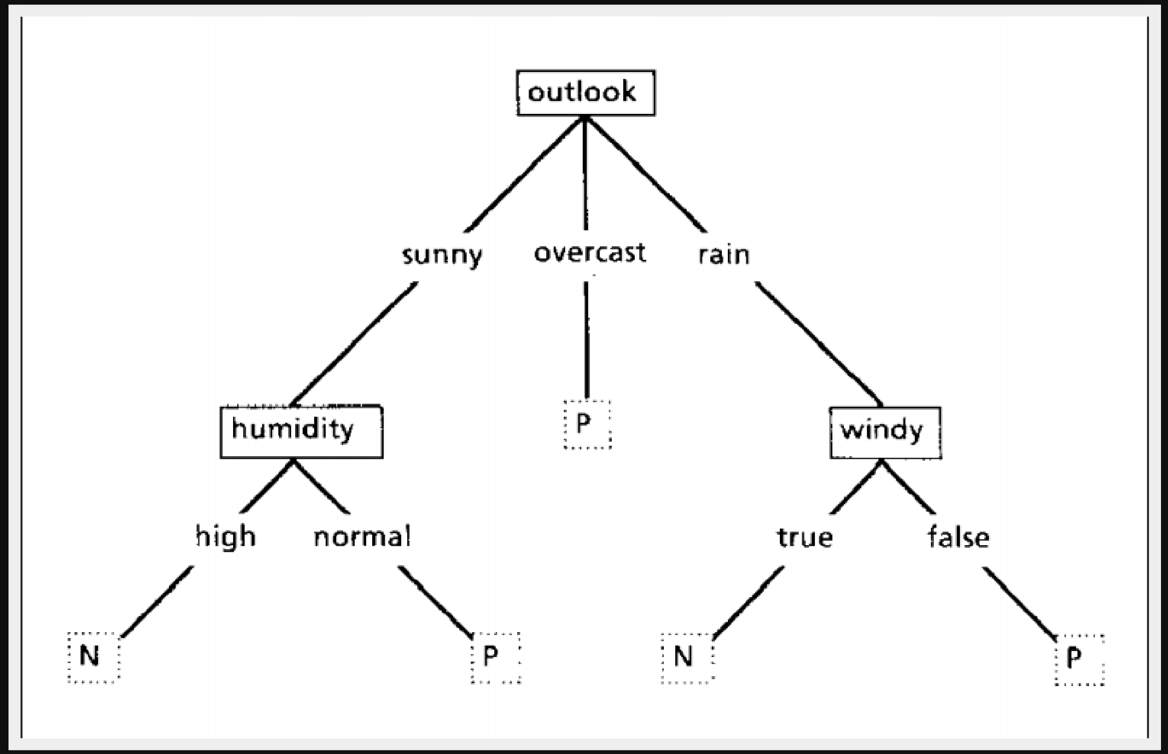
- 'capable not only of emulating the expert in the quality of decisions, but also in the ability to give reasons and justification' - 'bottleneck problem of artifical intelligence' (Feigenbaum, 1977)



Michie D. (1985). Current developments in artificial intelligence and expert systems. Zygon, 20(4), pp.375-389.

What's old? - Classification algorithms

Quinlan's 'Induction of Decision Trees' paper contains many of the concepts from modern inductive ML: the *training set*, generalisation through regularisation, noisy training sets, and missing data.



Quinlan, J.R. (1986&#41. Induction of Decision Trees. J.R. Mach Learn, 1, pp.81-106

More old ideas

More old ideas

- Back propogation (many earlier references, but Yann LeCun in 1987)

More old ideas

- Back propogation (many earlier references, but Yann LeCun in 1987)
- Automatic differentiation (by 1976, see Griewank, Andreas (2012))

More old ideas

- Back propogation (many earlier references, but Yann LeCun in 1987)
- Automatic differentiation (by 1976, see Griewank, Andreas (2012))
- Long short-term memory (LSTM) (Hochreiter S. and Schmidhuber J. (1997))

More old ideas

- Back propogation (many earlier references, but Yann LeCun in 1987)
- Automatic differentiation (by 1976, see Griewank, Andreas (2012))
- Long short-term memory (LSTM) (Hochreiter S. and Schmidhuber J. (1997))
- Transfer learning reviews back in 1998 (Thrun S. and Pratt L. (1998))

What's old? - Neural Networks in Law

Belew used neural nets for document retrieval at the first International Conference on Artificial Intelligence and Law, 1987. Document knowledge representation encoded in the weights of the network.

A connectionist approach to conceptual information retrieval

Richard K. Belew
Computer Science & Engineering Dept.
University of California - San Diego
rik@sdcsvax.ucsd.edu

"...(a) word is not a crystal, transparent and unchanged, it is the skin of a living thought."

Chief Justice Holmes, in Towne v. Eisner, 1918.

Abstract

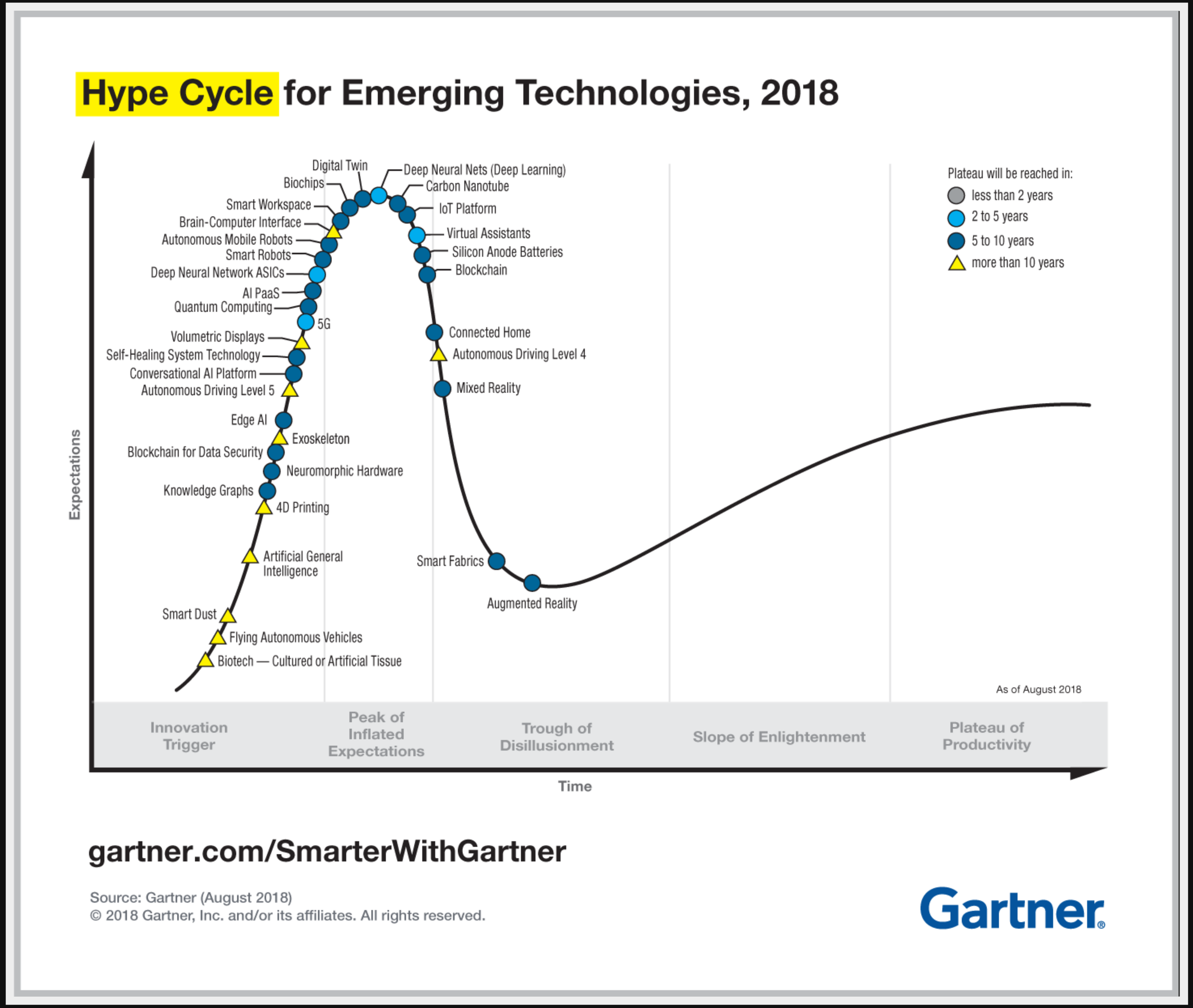
This report proposes that recent advances using low-level connectionist representations offer new possibilities to those interested in free text information retrieval (IR). The AIR system demonstrates that this representation suits the IR domain well, particularly the special problems attending the more sophisti-

From the perspective of AI, the Law is a particularly attractive domain in which to study natural language because it at once embodies all of the centrally important questions of understanding natural language, but works with text that is crafted with more precision than most text.

Connectionism is emerging as a new, significantly different and promising new ***sub-symbolic*** knowledge representation technique in AI. This paper reports on experiments with AIR, a connectionist approach to conceptual information retrieval. It will be shown that this repre-

Belew, R. K. (1987) 'A Connectionist Approach to Conceptual Information Retrieval', ICAIL

'I've heard it all before!'



Gartner 2018, '5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018'

'I've heard it all before!'

'I've heard it all before!'

- The *Gartner Hype Cycle for Emerging Technologies 2018* featured 0 technologies reaching maturity

'I've heard it all before!'

- The *Gartner Hype Cycle for Emerging Technologies 2018* featured 0 technologies reaching maturity
- … and expectations are clustered around technologies that are > 5 years from usefulness

'I've heard it all before!'

- The *Gartner Hype Cycle for Emerging Technologies 2018* featured 0 technologies reaching maturity
- … and expectations are clustered around technologies that are > 5 years from usefulness
- So what are we to make of the promises being made by legal tech firms.

What's new

What's new

- Moore's law no longer relevant - hardware specialised to machine learning workloads (GPU, TPU)

What's new

- Moore's law no longer relevant - hardware specialised to machine learning workloads (GPU, TPU)
- Packaging of research into well engineered software libraries and scientist-friendly langauges with open global communities (Spacy, Tensorflow, PyTorch, NumFocus)

What's new

- Moore's law no longer relevant - hardware specialised to machine learning workloads (GPU, TPU)
- Packaging of research into well engineered software libraries and scientist-friendly langauges with open global communities (Spacy, Tensorflow, PyTorch, NumFocus)
- Transfer learning for NLP

What's new

- Moore's law no longer relevant - hardware specialised to machine learning workloads (GPU, TPU)
- Packaging of research into well engineered software libraries and scientist-friendly langaugues with open global communities (Spacy, Tensorflow, PyTorch, NumFocus)
- Transfer learning for NLP
- Cloud computing

Dedicated ML hardware

Dedicated ML hardware

- Dedicated ML training and serving hardware on-demand (TPU, GraphCore)

Dedicated ML hardware

- Dedicated ML training and serving hardware on-demand (TPU, GraphCore)
- Billed by the minute by public IAAS providers

Dedicated ML hardware

- Dedicated ML training and serving hardware on-demand (TPU, GraphCore)
- Billed by the minute by public IAAS providers
- *but* still expensive to train state-of-the-art models

Standard NLP courses

(demonstrations at https://demo.allennlp.org):

ThoughtRiver™ intelligent contract pre-screening

Standard NLP courses

- tokenisation and segmentation

(demonstrations at https://demo.allennlp.org):

Standard NLP courses

- tokenisation and segmentation
- part-of-speech classifiers, parse-trees, dependency parsing

(demonstrations at https://demo.allennlp.org):

Standard NLP courses

- tokenisation and segmentation
- part-of-speech classifiers, parse-trees, dependency parsing
- regular expressions

(demonstrations at https://demo.allennlp.org):

Standard NLP courses

- tokenisation and segmentation
- part-of-speech classifiers, parse-trees, dependency parsing
- regular expressions
- named entity recognition

(demonstrations at https://demo.allennlp.org):

ThoughtRiver™ intelligent contract pre-screening

Standard NLP courses

- tokenisation and segmentation
- part-of-speech classifiers, parse-trees, dependency parsing
- regular expressions
- named entity recognition
- semantic role labelling (frames)

(demonstrations at https://demo.allennlp.org):

Standard NLP courses

- tokenisation and segmentation
- part-of-speech classifiers, parse-trees, dependency parsing
- regular expressions
- named entity recognition
- semantic role labelling (frames)
- document classifiers and topic models

(demonstrations at https://demo.allennlp.org):

Standard NLP courses

- tokenisation and segmentation
- part-of-speech classifiers, parse-trees, dependency parsing
- regular expressions
- named entity recognition
- semantic role labelling (frames)
- document classifiers and topic models
- language models

(demonstrations at https://demo.allennlp.org):

Standard NLP courses

- tokenisation and segmentation
- part-of-speech classifiers, parse-trees, dependency parsing
- regular expressions
- named entity recognition
- semantic role labelling (frames)
- document classifiers and topic models
- language models
- embeddings (Word2Vec, Glove, FastText, Doc2Vec, DSSM)

(demonstrations at https://demo.allennlp.org):

Standard NLP courses

- tokenisation and segmentation
- part-of-speech classifiers, parse-trees, dependency parsing
- regular expressions
- named entity recognition
- semantic role labelling (frames)
- document classifiers and topic models
- language models
- embeddings (Word2Vec, Glove, FastText, Doc2Vec, DSSM)
- sequence-to-sequence tasks with recurrent networks

(demonstrations at https://demo.allennlp.org):

Tokenisation and segmentation

Tokenisation and segmentation

- how to decompose a sequence of characters into chunks

Tokenisation and segmentation

- how to decompose a sequence of characters into chunks
- and where to best splice those chunks into segments

Tokenisation and segmentation

- how to decompose a sequence of characters into chunks
- and where to best splice those chunks into segments
- thankfully, English is straightforward in this respect (whitespace is 90% of the task)

Linguistic preprocessing

Linguistic preprocessing

- traditional 'stacked' NLP has separate models for linguistic primatives

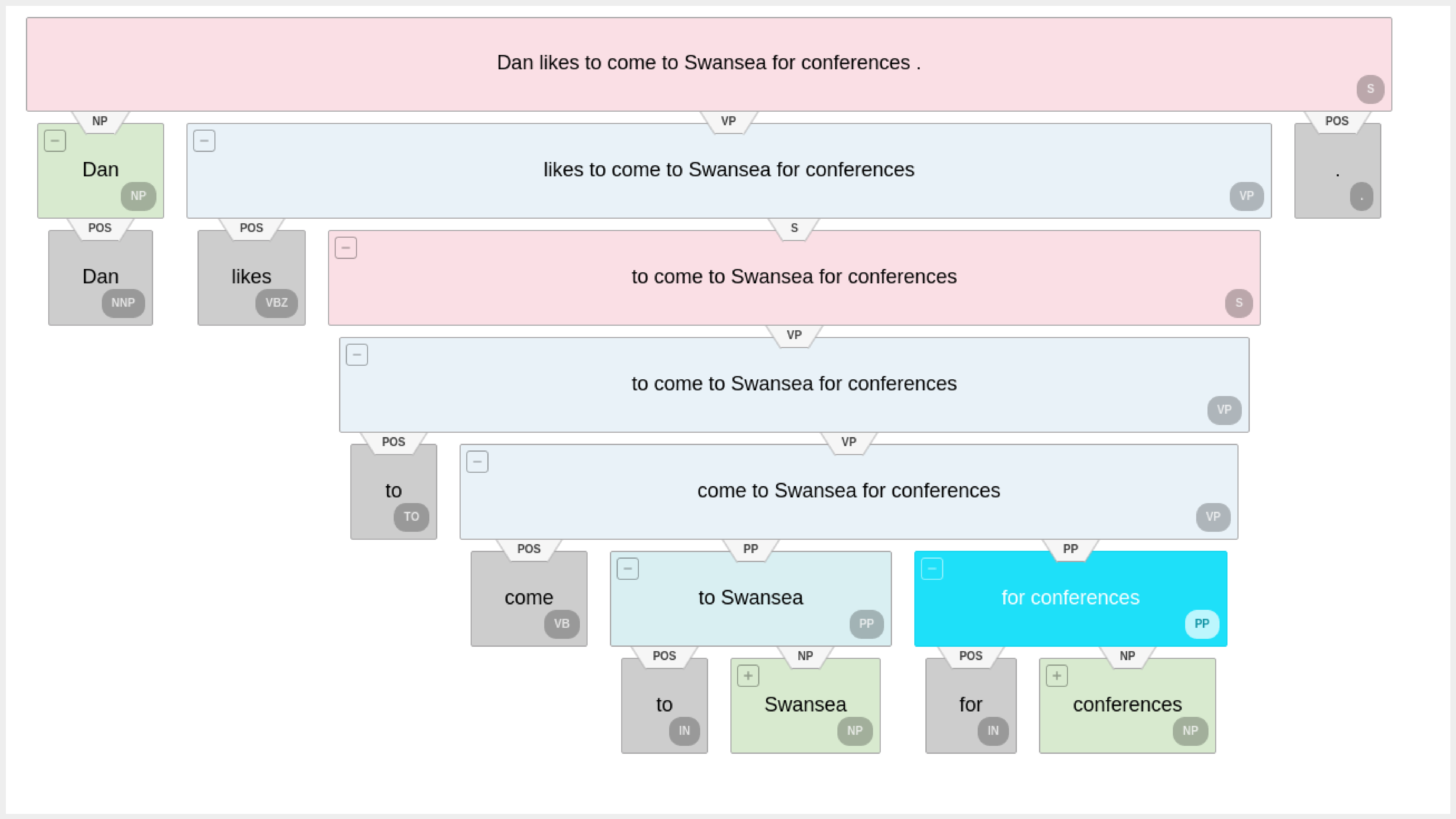
Linguistic preprocessing

- traditional 'stacked' NLP has separate models for linguistic primatives
- part-of-speech classifiers tag tokens with their function (verbs, nouns, determiners)

Linguistic preprocessing

- traditional 'stacked' NLP has separate models for linguistic primatives
- part-of-speech classifiers tag tokens with their function (verbs, nouns, determiners)
- parse-trees (consituency parsing), dependency parsing

Constituency parsing



Dependency parsing

Language models

-

Language models

- predict the missing word

  -

Language models

- predict the missing word
- the internals encode important language features...

-

Language models

- predict the missing word
- the internals encode important language features...
- which can be extracted and used on other tasks

-

Regular expressions

Regular expressions

- origins in finite state automata

Regular expressions

- origins in finite state automata
- 'hard' pattern recognition

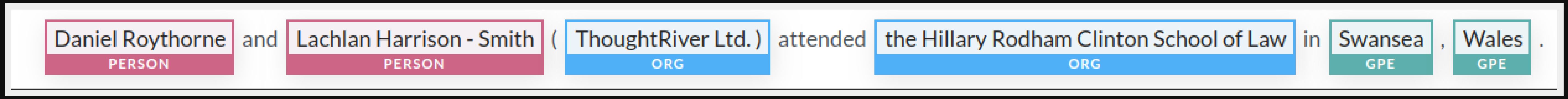
Regular expressions

- origins in finite state automata
- 'hard' pattern recognition
- extremely effective in bounded contexts

Regular expressions

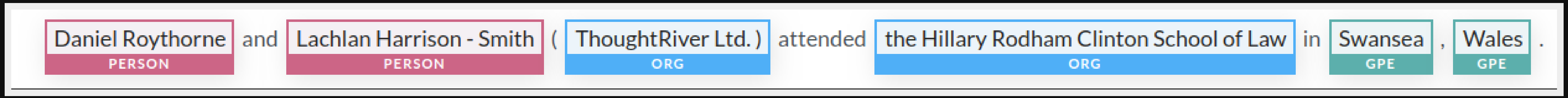
- origins in finite state automata
- 'hard' pattern recognition
- extremely effective in bounded contexts
- but nightmareish to maintain

Named entity recognition (NER)



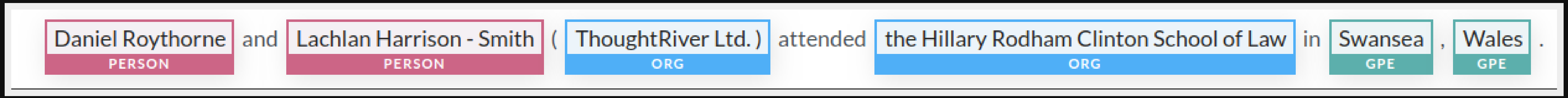
Named entity recognition (NER)

- subsequences of tokens represent meaningful 'things'



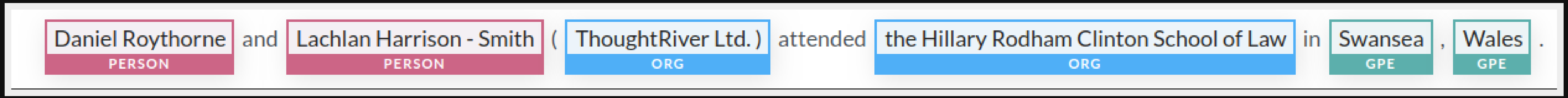
Named entity recognition (NER)

- subsequences of tokens represent meaningful 'things'
- attach tags to contiguous tokens



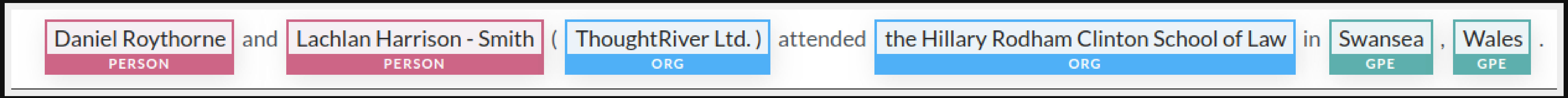
Named entity recognition (NER)

- subsequences of tokens represent meaningful 'things'
- attach tags to contiguous tokens
- e.g. ORG (organisation), PROXY (reference terms), GPE (Countries, Cities, States)



Named entity recognition (NER)

- subsequences of tokens represent meaningful 'things'
- attach tags to contiguous tokens
- e.g. ORG (organisation), PROXY (reference terms), GPE (Countries, Cities, States)
- spaCy models and documentation are great sources



Topic modelling

Topic modelling

- unsupervised method

Topic modelling

- unsupervised method
- infer from linear, generative models (e.g. Latent Dirichlet Allocation)

Topic modelling

- unsupervised method
- infer from linear, generative models (e.g. Latent Dirichlet Allocation)
- remains useful for contract classification

Embeddings

Embeddings

- e.g. word2vec (Mikolov, 2013)

Embeddings

- e.g. word2vec (Mikolov, 2013)
- the representations learned by a language model are useful

Embeddings

- e.g. word2vec (Mikolov, 2013)
- the representations learned by a language model are useful
- allow us to turn a token into a vector with useful properties

Sequence-to-sequence tasks

Sequence-to-sequence tasks

- machine translation ((Google Translate) [https://translate.google.com/])

Sequence-to-sequence tasks

- machine translation ((Google Translate) [https://translate.google.com/])
- text summarisation ((frase.io review) [https://blog.frase.io/20-applications-of-automatic-summarization-in-the-enterprise/])

Sequence-to-sequence tasks

- machine translation ((Google Translate)[https://translate.google.com/])
- text summarisation ((frase.io review)[https://blog.frase.io/20-applications-of-automatic-summarization-in-the-enterprise/])
- diary management ((Amy)[https://x.ai/how-it-works/])

Newer ideas

Newer ideas

- transfer learning in NLP

Newer ideas

- transfer learning in NLP
- neural attention

Newer ideas

- transfer learning in NLP
- neural attention
- neural Turing machines

Newer ideas

- transfer learning in NLP
- neural attention
- neural Turing machines
- weak supervision

Transfer learning

Transfer learning

- ULMFit, BERT and ELMO

Transfer learning

- ULMFit, BERT and ELMO
- 'NLP's ImageNet moment'

Neural attention

Neural attention

- 'Attention is all you need'

Neural attention

- 'Attention is all you need'
- allow recurrent models to encompass larger contexts

Neural Turing machines

Neural Turing machines

- neural networks together with 'addressable' memory (e.g. (DNC) [https://www.nature.com/articles/nature20101])

Neural Turing machines

- neural networks together with 'addressable' memory (e.g. (DNC) [https://www.nature.com/articles/nature20101])
- haven't scaled to industrial problems

Weak supervision

Weak supervision

- methods to use 'soft' labels to bootstrap our training data

Weak supervision

- methods to use 'soft' labels to bootstrap our training data
- similarities with data fusion techniques from signal processing

Examples from contract pre-screening

Examples from contract pre-screening

- Who are the parties and how are they referred to?

Examples from contract pre-screening

- Who are the parties and how are they referred to?
- Is the contract under the jurisdiction of a European Union country?

Examples from contract pre-screening

- Who are the parties and how are they referred to?
- Is the contract under the jurisdiction of a European Union country?
- Does an NDA prescribe a permitted use for confidential information?

Examples from contract pre-screening

- Who are the parties and how are they referred to?
- Is the contract under the jurisdiction of a European Union country?
- Does an NDA prescribe a permitted use for confidential information?
- Does a purchase agreement specify a delivery date in the past?

ThoughtRiver™ intelligent contract pre-screening

Examples from contract pre-screening

- Who are the parties and how are they referred to?
- Is the contract under the jurisdiction of a European Union country?
- Does an NDA prescribe a permitted use for confidential information?
- Does a purchase agreement specify a delivery date in the past?
- Does anything look out of place?

Examples from contract pre-screening

- Who are the parties and how are they referred to?
- Is the contract under the jurisdiction of a European Union country?
- Does an NDA prescribe a permitted use for confidential information?
- Does a purchase agreement specify a delivery date in the past?
- Does anything look out of place?
- When do provisions of an agreement terminate?

Examples from contract pre-screening

- Who are the parties and how are they referred to?
- Is the contract under the jurisdiction of a European Union country?
- Does an NDA prescribe a permitted use for confidential information?
- Does a purchase agreement specify a delivery date in the past?
- Does anything look out of place?
- When do provisions of an agreement terminate?
- But first, the glamourous bits ...

Preprocessing

Garbage in, garbage out is still a truism.

Preprocessing

Garbage in, garbage out is still a truism.

- Microsoft Word is ubiquitous

Preprocessing

Garbage in, garbage out is still a truism.

- Microsoft Word is ubiquitous
- Injesting silos of images, .pdfs and deprecated file formats

Preprocessing

Garbage in, garbage out is still a truism.

- Microsoft Word is ubiquitous
- Injesting silos of images, .pdfs and deprecated file formats
- .docx parsing and source document representations (rich content retention)

Preprocessing

Garbage in, garbage out is still a truism.

- Microsoft Word is ubiquitous
- Injesting silos of images, .pdfs and deprecated file formats
- .docx parsing and source document representations (rich content retention)
- document type classification

Preprocessing

Garbage in, garbage out is still a truism.

- Microsoft Word is ubiquitous
- Injesting silos of images, .pdfs and deprecated file formats
- .docx parsing and source document representations (rich content retention)
- document type classification
- language detection

Preprocessing

Garbage in, garbage out is still a truism.

- Microsoft Word is ubiquitous
- Injesting silos of images, .pdfs and deprecated file formats
- .docx parsing and source document representations (rich content retention)
- document type classification
- language detection
- normalisation

Party detection

Party detection

- What legal entities are parties to the agreement?

Party detection

- What legal entities are parties to the agreement?
- How to pick out references to those parties?

Party detection

- What legal entities are parties to the agreement?
- How to pick out references to those parties?
- What about references to groups, and reciprocal terms (e.g. 'the parties')

This [type of agreement] (the “Agreement”) is made on .... (“Effective Date”). BETWEEN: XAVIER INC whose operational office is at Greenville Park, Los Angeles, CA, 923040, USA (“XAVIER”); and YOLO LTD with its registered office at Stream Business Park, Santa Monica, CA, 90404, USA (“YOLO”). Separately a “party” and together the “parties”.

Jurisdiction

6.1 This Agreement shall be governed for all purposes by the laws of Sudan and the parties irrevocably submit to the exclusive jurisdiction of the courts of North Korea.

Behavioural prescriptions and normative conditions

Behavioural prescriptions and normative conditions

- statements of rights, prohibitions, obligations

Behavioural prescriptions and normative conditions

- statements of rights, prohibitions, obligations
- and the states of the world where they apply

Behavioural prescriptions and normative conditions

- statements of rights, prohibitions, obligations
- and the states of the world where they apply
- e.g. confidentiality clauses in NDAs

The use of the Confidential Information by the Recipient hereunder shall be limited solely to and for the purposes of Project Microsoft (the “Permitted Use”).

Engineering challenges

Engineering challenges

- data-driven systems are *far* more complex than regular software

Engineering challenges

- data-driven systems are *far* more complex than regular software
- Simple software: data-B -> ALGORITHM -> data-A

Engineering challenges

- data-driven systems are *far* more complex than regular software
- Simple software: data-B -> ALGORITHM -> data-A
- ML model: (data-C -> META-ALGORITHM -> ALGORITHM-A) + (Simple software)

Engineering challenges

- data-driven systems are *far* more complex than regular software
- Simple software: data-B -> ALGORITHM -> data-A
- ML model: (data-C -> META-ALGORITHM -> ALGORITHM-A) + (Simple software)
- ML product: (data-A -> META-META-ALGORITHM -> META-ALGORITHM-A) + (ML model)

ML deployment issues

- Machine Learning: The High Interest Credit Card of Technical Debt
- Software Engineering for Machine Learning: A Case Study
- What's your ML test score? A rubric for ML production systems

Challenges

Challenges

- Labelled data bottleneck

Challenges

- Labelled data bottleneck
- Legal language is not natural language

Challenges

- Labelled data bottleneck
- Legal language is not natural language
- Complex domain ontology

Challenges

- Labelled data bottleneck
- Legal language is not natural language
- Complex domain ontology
- Extensive explicit and implicit context

Challenges

- Labelled data bottleneck
- Legal language is not natural language
- Complex domain ontology
- Extensive explicit and implicit context
- Ethics and regulation

Challenges

- Labelled data bottleneck
- Legal language is not natural language
- Complex domain ontology
- Extensive explicit and implicit context
- Ethics and regulation
- System complexity is high where behaviour is non-stationary

Challenges

- Labelled data bottleneck
- Legal language is not natural language
- Complex domain ontology
- Extensive explicit and implicit context
- Ethics and regulation
- System complexity is high where behaviour is non-stationary
- Business process integration and engendering 'data science empathy'

Training data curation

Training data curation

- expert annotation is expensive

Training data curation

- expert annotation is expensive
- managing data campaigns is nuanced (annotator quality, tooling, bias, training corpus selection)

Training data curation

- expert annotation is expensive
- managing data campaigns is nuanced (annotator quality, tooling, bias, training corpus selection)
- Mechanical Turk is not an option

Training data curation

- expert annotation is expensive
- managing data campaigns is nuanced (annotator quality, tooling, bias, training corpus selection)
- Mechanical Turk is not an option
- minimise volume of training data required

Methods to minimise training volumes

Plenty of options, but no shinkwrapped solutions.

Methods to minimise training volumes

Plenty of options, but no shinkwrapped solutions.

- transfer learning

Methods to minimise training volumes

Plenty of options, but no shinkwrapped solutions.

- transfer learning
- active learning

Methods to minimise training volumes

Plenty of options, but no shinkwrapped solutions.

- transfer learning
- active learning
- semi-supervised learning

Methods to minimise training volumes

Plenty of options, but no shinkwrapped solutions.

- transfer learning
- active learning
- semi-supervised learning
- weak supervision

Legal language is not natural language

Legal language is not natural language

- much lower entropy -> smaller models

Legal language is not natural language

- much lower entropy -> smaller models
- transferrability of results from academic research is not a given

ThoughtRiver™
intelligent contract pre-screening

Complex domain ontology

Complex domain ontology

- harness the structure of legal concepts

Complex domain ontology

- harness the structure of legal concepts
- constrain models to give logically consistent answers (e.g. hierarchical classifiers)

Extensive explicit and implicit context

Extensive explicit and implicit context

- coalesce relevant information from distant parts of documents

ThoughtRiver intelligent contract pre-screening

Extensive explicit and implicit context

- coalesce relevant information from distant parts of documents
- integration of external knowledge bases (e.g. company registers, OpenCorporates)

Ethics and regulation

Ethics and regulation

- bias, fairnes

Ethics and regulation

- bias, fairnes
- explainability

Ethics and regulation

- bias, fairnes
- explainability
- privacy preservation