

# Learning How to Say It: Language Generation and Deep Learning

Alexander M Rush

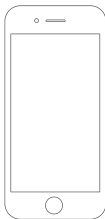
# Talk Outline

- ① Background: Core Model and Implementation
- ② Work 1: Rethinking Model Training (*Beam Search Optimization*)
- ③ **Work 2:** Rethinking Generation (*Learning Neural Templates*)
- ④ Future Directions

# Talking About Data

 $x$ 

Frederick Parker-Rhodes	
<b>Born</b>	21 November 1914 Newington, <a href="#">Yorkshire</a>
<b>Died</b>	2 March 1987 (aged 72)
<b>Residence</b>	UK
<b>Nationality</b>	British
<b>Known for</b>	Contributions to <a href="#">computational linguistics</a> , <a href="#">combinatorial physics</a> , <a href="#">bit-string physics</a> , <a href="#">plant pathology</a> , and <a href="#">mycology</a>
<b>Scientific career</b>	
<b>Fields</b>	<a href="#">Mycology</a> , <a href="#">Plant Pathology</a> , <a href="#">Mathematics</a> , <a href="#">Linguistics</a> , <a href="#">Computer Science</a>
<b>Author abbrev. (botany)</b>	Park.-Rhodes

 $\theta$ 

# Talking About Data

 $x$ 

Frederick Parker-Rhodes	
<b>Born</b>	21 November 1914 Newington, <a href="#">Yorkshire</a>
<b>Died</b>	2 March 1987 (aged 72)
<b>Residence</b>	UK
<b>Nationality</b>	British
<b>Known for</b>	Contributions to <a href="#">computational linguistics</a> , <a href="#">combinatorial physics</a> , <a href="#">bit-string physics</a> , <a href="#">plant pathology</a> , and <a href="#">mycology</a>
<b>Scientific career</b>	
<b>Fields</b>	<a href="#">Mycology</a> , <a href="#">Plant Pathology</a> , <a href="#">Mathematics</a> , <a href="#">Linguistics</a> , <a href="#">Computer Science</a>
<b>Author abbrev. (botany)</b>	Park.-Rhodes

 $\theta$  $y_{1:T}$ 

Frederick Parker-Rhodes (21 November 1914 2 March 1987) was an English linguist, plant pathologist, computer scientist, mathematician, mystic, and mycologist.

# Talking About Data

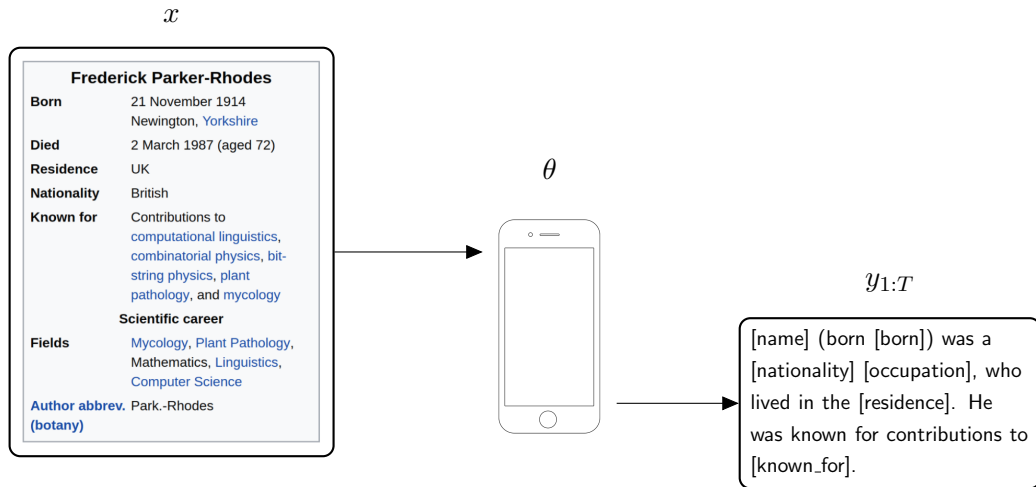
 $x$ 

Frederick Parker-Rhodes	
<b>Born</b>	21 November 1914 Newington, <a href="#">Yorkshire</a>
<b>Died</b>	2 March 1987 (aged 72)
<b>Residence</b>	UK
<b>Nationality</b>	British
<b>Known for</b>	Contributions to <a href="#">computational linguistics</a> , <a href="#">combinatorial physics</a> , <a href="#">bit-string physics</a> , <a href="#">plant pathology</a> , and <a href="#">mycology</a>
<b>Scientific career</b>	
<b>Fields</b>	<a href="#">Mycology</a> , <a href="#">Plant Pathology</a> , <a href="#">Mathematics</a> , <a href="#">Linguistics</a> , <a href="#">Computer Science</a>
<b>Author abbrev. (botany)</b>	Park.-Rhodes

 $\theta$  $y_{1:T}^*$ 

Frederick Parker-Rhodes (21 November 1914 – 2 March 1987) was an English mycology and **plant pathology**, **mathematics** at the University of UK.

# Alternative Approach: Templated Generation



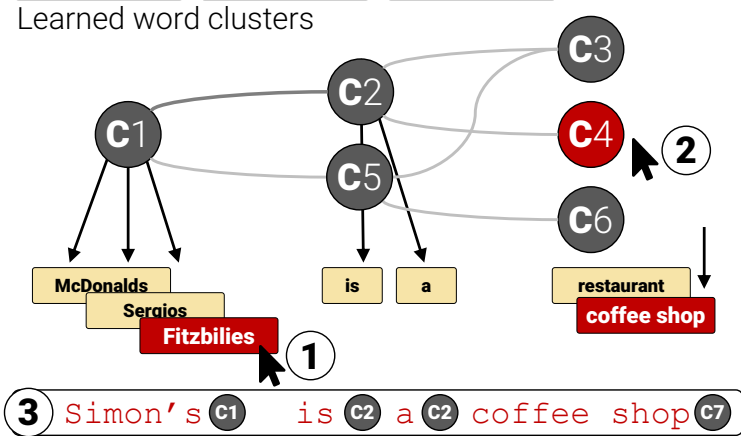
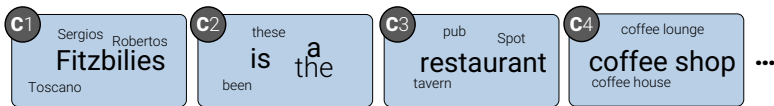
# Arguments for Rule-Based Generation

Guarantees about the quality, in particular,

- ① Interpretable in its factual content.
- ② Controllable in terms of style and form.

Can we achieve this with a deep learning based system?

# Learning Neural Templates for Generation





# Approach: Deep Latent-Variable Models

Goal: Expose specific choices as *discrete* latent variables  $z$ .

$$p(y, z \mid x; \theta)$$

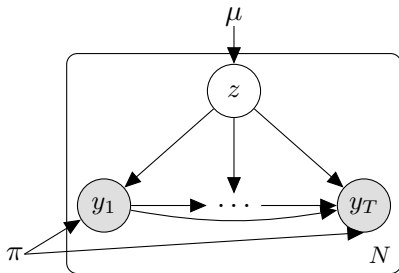
- $x, y, \theta$  as before, *what to talk about / how to say it*
- $z$  is a collection of latent variables

## Example 1: Conditional Sentence Clustering

Generative process:

- 1 Draw cluster  $z \in \{1, \dots, K\}$  from a Categorical.
- 2 Draw words  $y_{1:T}$  from RNN with parameters  $\pi_z$ .

$$p(y, z|x; \theta) = \mu_z \times \text{RNN}(y_{1:T}; \pi_z)$$



## Example 2: Summary with Copy

Let  $z$  be a binary latent variable.

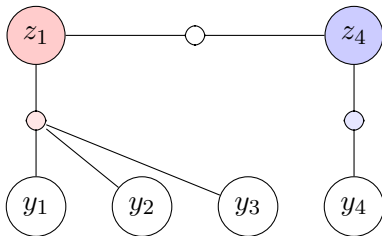
- If  $z = 0$ , let the model generate a new word.
- If  $z = 1$ , let the model copy a word from the source.

### Pointer-generator model + coverage summary

francis saili has signed a two-year deal to join munster later this year .  
the 24-year-old was part of the new zealand under-20 side that won the junior world championship in italy in 2011 .  
saili 's signature is something of a coup for munster and head coach anthony foley .

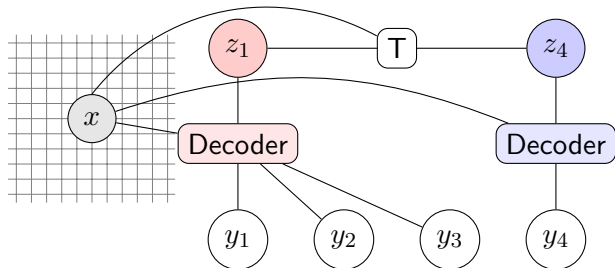
## Base Model: Hidden Semi-Markov Model

- HMM: discrete latent states with single emissions (e.g. words).
- HSMM: discrete latent states produce multiple emissions (e.g. phrases).
- Parameterized with *transition*, *emission*, and *length* distributions.



## Our Proposal: Neural Hidden Semi-Markov Model

- Employ HSMM as a conditional latent variable language model,  $p(y_1, \dots, y_T, z \mid x)$ .
- Transition Distribution: NN between states.
- Emission Distribution: Encoder-Decoder, one per state.



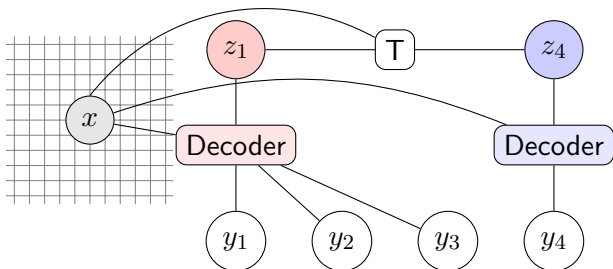
## Technical Methodology: Fitting Parameters

Fit model by minimizing negative log-marginal likelihood on training data.

$$\min_{\theta} -\log \sum_z p(y, z \mid x; \theta)$$

Details: Use dynamic programming to efficiently compute HSMM forward algorithm for sum, backprop with autograd, all inference is exact.

## From Neural HSMM to Templates



Compute argmax latent variables to find common *templates*.

$$z_{1:T}^* = \arg \max_{z_{1:T}} p(y_{1:T}, z_{1:T} \mid x; \theta)$$

*[The Wrestlers]<sub>185</sub> [is a]<sub>29</sub> [coffee shop]<sub>164</sub> [that serves]<sub>188</sub> [English]<sub>139</sub> [food]<sub>18</sub> [in the]<sub>32</sub>  
[moderate]<sub>125</sub> [price range]<sub>180</sub> [.]<sub>90</sub>*

# Example Templates: Wikipedia

1. 

aftab ahmed	(	born	1951	)	is an american	actor
anderson da silva	;	born on	1970		was an american	actress
david jones		born 1	1974		is an english	cricketer
...	...	...	...	...	...	...
2. 

aftab ahmed	was a	world war i	member of the	austrian	house of representatives
anderson da silva	is a former	liberal	party member of the	pennsylvania	legislature
david jones	is a	baseball	recipient of the	montana	senate
...	...	...	...	...	...
3. 

adjutant	aftab ahmed	was a	world war i	member of the	knesset
lieutenant	anderson da silva	is a former	liberal	party member of the	scottish parliament
captain	david jones	is a	baseball	recipient of the	fc lokomotiv liski
...	...	...	...	...	...
4. 

william	" billy " watson	1913	-	1917	was an american	football player
john william	smith	(   c. 1900	in	surrey, england	was an australian	rules footballer
james "	jim " edward	1913	-	british columbia	is an american	defenceman
...	...	...	...	...	...	...
	who plays for	collingwood	in the	victorial football league	vfl	
	who currently plays for	st kilda	of the	national football league	(   afl   )	
	who played with	carlton	and the	australian football league	(   nfl   )	
...	...	...	...	...	...	...
5. 

aftab ahmed	is a	member of the	knesset
anderson da silva	is a former	party member of the	scottish parliament
david jones	is a female	recipient of the	fc lokomotiv liski
...	...	...	...



# Neural Template Generation Approach

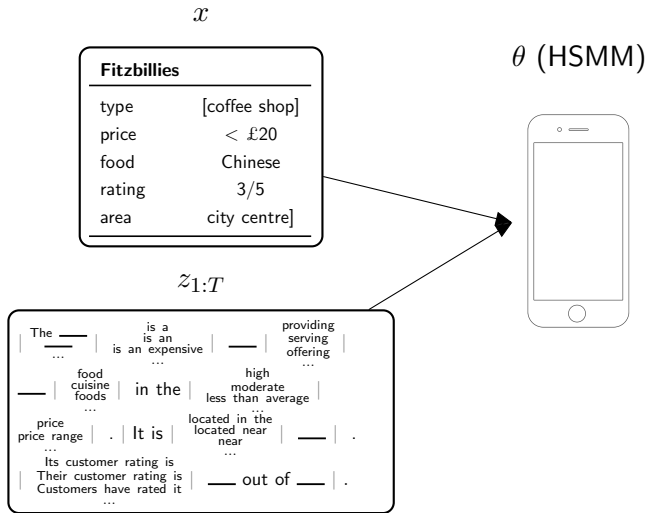
$x$

Fitzbillies	
type	[coffee shop]
price	< £20
food	Chinese
rating	3/5
area	city centre]

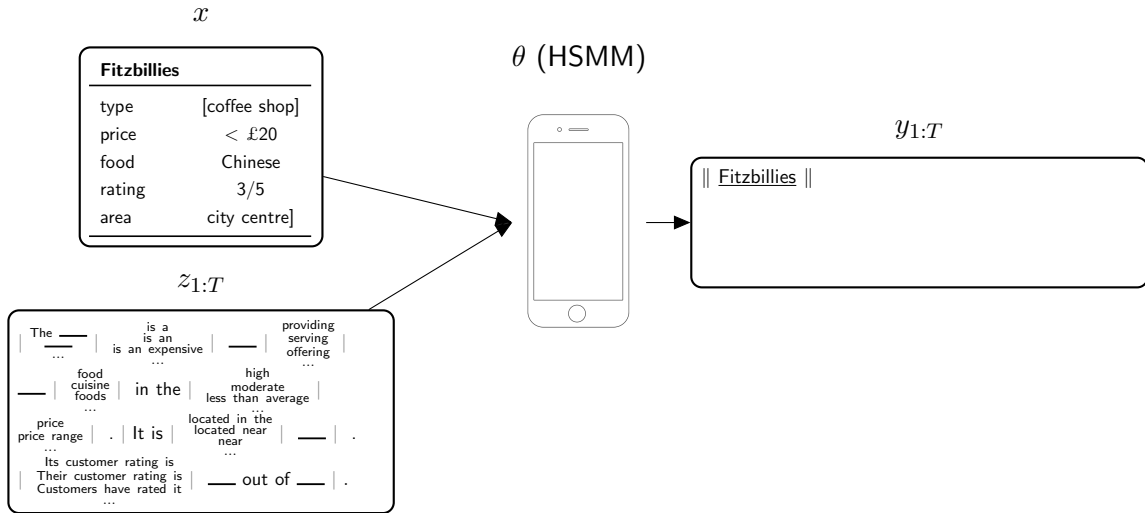
$\theta$  (HSMM)



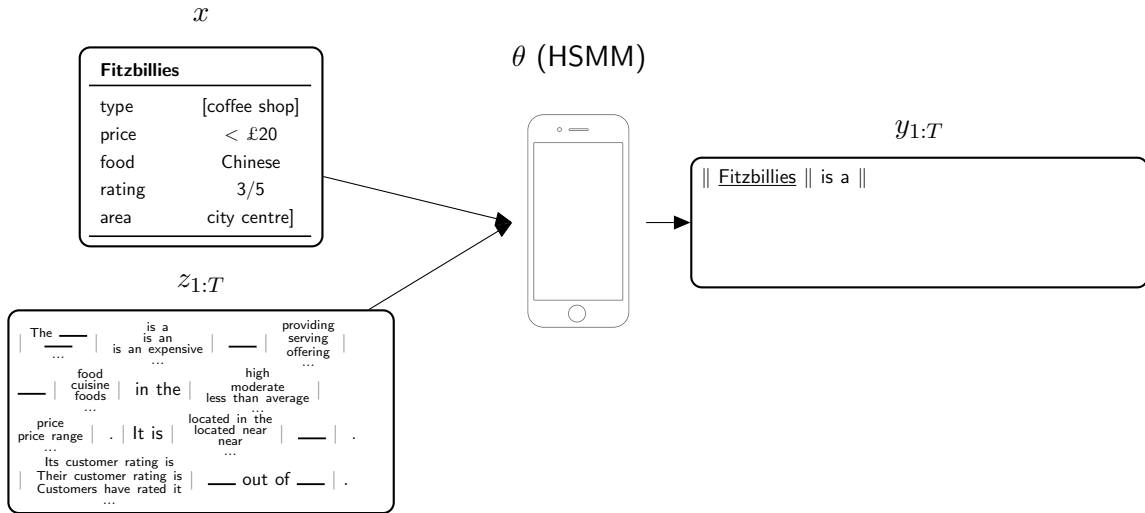
# Neural Template Generation Approach



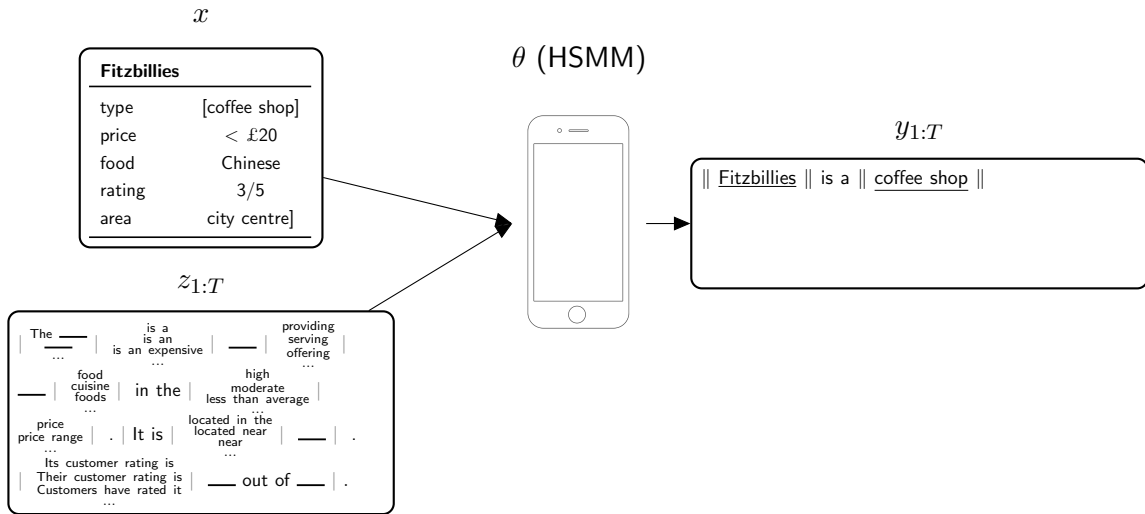
# Neural Template Generation Approach



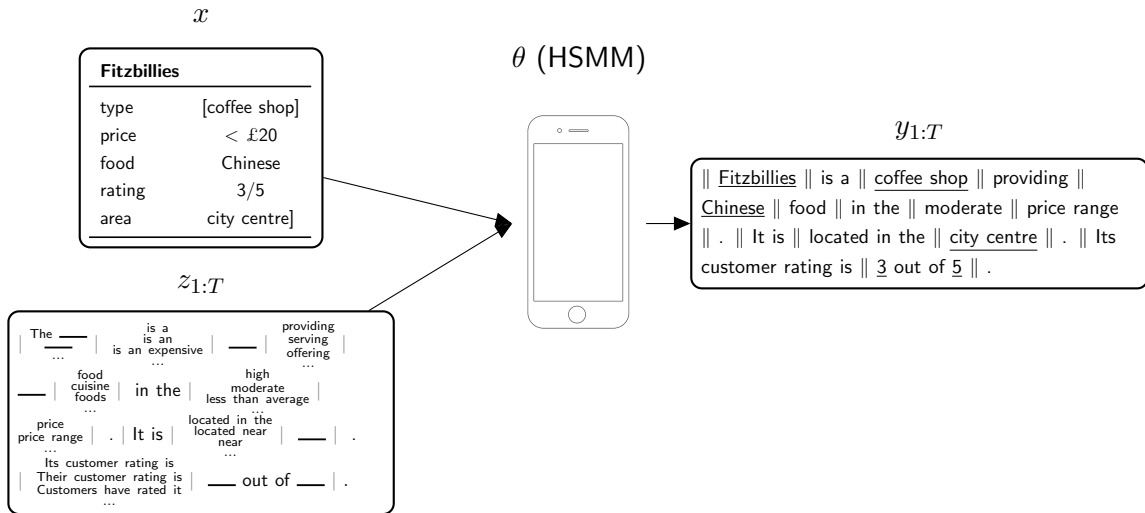
# Neural Template Generation Approach



# Neural Template Generation Approach



# Neural Template Generation Approach



# Issue 1: Interpretability

---

kenny warren

---

**name:** kenny warren, **birth date:** 1 april 1946,

**birth name:** kenneth warren deutscher, **birth place:** brooklyn, new york,

**occupation:** ventriloquist, comedian, author,

**notable work:** book - the revival of ventriloquism in america

---

1. kenny warren deutscher ( april 1, 1946 ) is an american ventriloquist.
  2. kenny warren deutscher ( april 1, 1946 , brooklyn,) is an american ventriloquist.
  3. kenny warren deutscher ( april 1, 1946 ) is an american ventriloquist, best known for his the revival of ventriloquism.
  4. "kenny" warren is an american ventriloquist.
  5. kenneth warren "kenny" warren (born april 1, 1946 ) is an american ventriloquist, and author.
-

## Issue 2: Controllability

---

### **The Golden Palace**

---

name[The Golden Palace], type[coffee shop], food[Chinese],  
priceRange[cheap] custRating[5 out of 5], area[city centre],

---

1. The Golden Palace is a coffee shop located in the city centre.
  2. In the city centre is a cheap Chinese coffee shop called The Golden Palace.
  3. The Golden Palace that serves Chinese food in the cheap price range. It is located in the city centre. Its customer rating is 5 out of 5.
  4. The Golden Palace is a Chinese coffee shop.
  5. The Golden Palace is a Chinese coffee shop with a customer rating of 5 out of 5.
-



# Results

	BLEU	NIST
Test		
Substitution	43.78	6.88
Neural Template	56.72	7.63
Full Neural Model	65.93	8.59

	BLEU	NIST	ROUGE-4
Conditional KN-LM	19.8	5.19	10.7
NNLM (field)	33.4	7.52	23.9
NNLM (field & word)	34.7	7.98	25.8
Neural Template	33.8	7.51	28.2

# Talk Outline

- ① Background: Core Model and Implementation
- ② Work 1: Rethinking Model Training (*Beam Search Optimization*)
- ③ Work 2: Rethinking Generation (*Learning Neural Templates*)
- ④ **Future Challenges in Text Generation**

# Three Challenge in Text Generation

- ① Long-Form Generation with High-Level Reasoning
- ② Compact and Efficient Generation
- ③ Latent-Variable Modeling for NLP

# Long-Form Generation with Explicit Reasoning

(3)

TEAM	WIN	LOSS	PTS	FG_PCT	RB	AS ...
Hawks	11	12	103	49	47	27
Heat	7	15	95	43	34	20

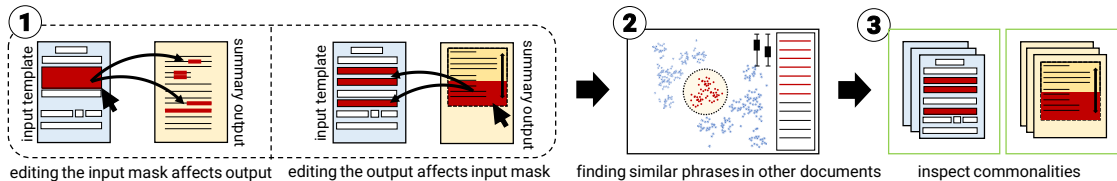
(2)

PLAYER	AS	RB	PT	FG	FGA	CITY ...
Tyler Johnson	5	2	27	8	16	Miami
Dwight Howard	11	17	23	9	11	Atlanta
Paul Millsap	2	9	21	8	12	Atlanta
Goran Dragic	4	2	21	8	17	Miami
Wayne Ellington	2	3	19	7	15	Miami
Dennis Schroder	7	4	17	8	15	Atlanta
Rodney McGruder	5	5	11	3	8	Miami
Hasan Whiteside	2	12	8	4	12	Miami
...						

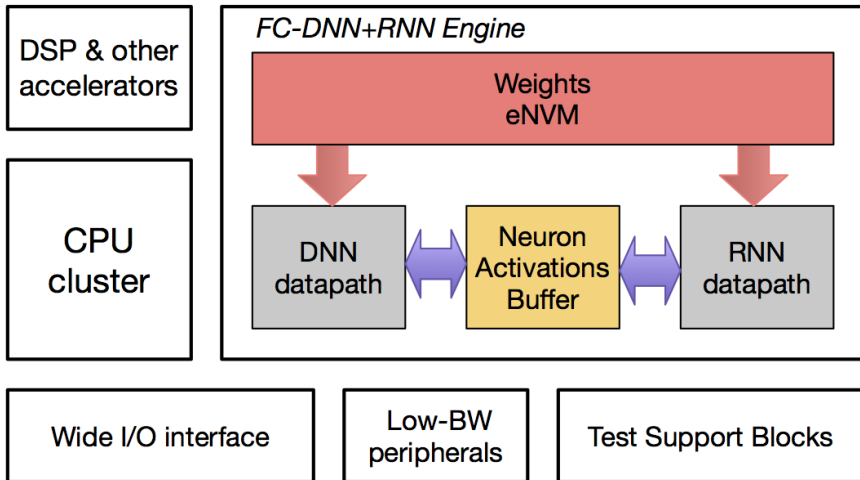
(1)

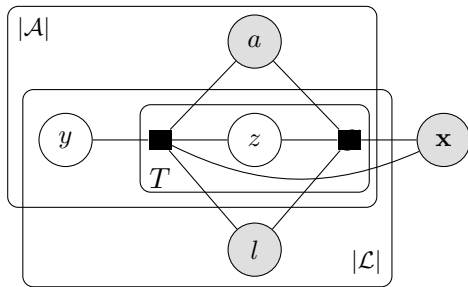
[The Atlanta Hawks defeated the Miami Heat, 103 - 95, at Philips Arena on Wednesday.] [Defense was key for the Hawks, as they held the Heat to 42 percent shooting and forced them to commit 16 turnovers. Atlanta also dominated in the paint, winning the rebounding battle, 47 - 34, and outscoring them in the paint 58 - 26. The Hawks shot 49 percent from the field and assisted on 27 of their 43 made baskets.] [Miami (7 - 15) are as beat-up as anyone right now. Hassan Whiteside really struggled in this game, as he amassed eight points, 12 rebounds and one blocks on 4 - of - 12 shooting ...]

# Summary



## ***Universal Translator SoC***





Thanks





Yuntian Deng, Anssi Kanervisto, and Alexander M. Rush. 2016. What You Get Is What You See: A Visual Markup Decompiler. In *Arxiv*.

Yuntian Deng, Yoon Kim, Justin Chiu, Demi Guo, and Alexander Rush. 2018. Latent alignment and variational attention. In *Advances in Neural Information Processing Systems*, pages 9735–9747.

Yoon Kim, Carl Denton, Luong Hoang, and Alexander M. Rush. 2017. Structured attention networks. [abs/1702.00887](https://arxiv.org/abs/1702.00887).

Yoon Kim, Yacine Jernite, David Sontag, and Alexander M. Rush. 2016. Character-Aware Neural Language Models. In *AAAI*.

Yoon Kim and Alexander M. Rush. 2016. Sequence-Level Knowledge Distillation. In *EMNLP*.

Yoon Kim, Sam Wiseman, Andrew C. Miller, David Sontag, and Alexander M. Rush. 2018. Semi-amortized variational autoencoders.

Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M. Rush. 2017.

Opennmt: Open-source toolkit for neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, System Demonstrations*, pages 67–72.

Brandon Reagen, Udit Gupta, Robert Adolf, Michael M Mitzenmacher, Alexander M Rush, Gu-Yeon Wei, and David Brooks. 2017. Weightless: Lossy weight encoding for deep neural network compression. *arXiv preprint arXiv:1711.04686*.

Alexander Rush. 2018. The annotated transformer. In *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, pages 52–60.

Alexander M Rush, Sumit Chopra, and Jason Weston. 2015. A Neural Attention Model for Abstractive Sentence Summarization. In *EMNLP*, September, pages 379–389.

Allen Schmaltz, Yoon Kim, Alexander M. Rush, and Stuart M. Shieber. 2016. Sentence-Level Grammatical Error Identification as Sequence-to-Sequence Correction. In *arxiv*.

Jean Senellart, Dakun Zhang, WANG Bo, Guillaume Klein, Jean-Pierre Ramatchandirin,

Josep Crego, and Alexander Rush. 2018. Opennmt system description for wnmt 2018: 800 words/sec on a single-core cpu. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pages 122–128.

Hendrik Strobelt, Sebastian Gehrmann, Michael Behrisch, Adam Perer, Hanspeter Pfister, and Alexander M Rush. 2019. Seq2seq-v is: A visual debugging tool for sequence-to-sequence models. *IEEE transactions on visualization and computer graphics*, 25(1):353–363.

Hendrik Strobelt, Sebastian Gehrmann, Bernd Huber, Hanspeter Pfister, and Alexander M. Rush. 2016. Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks. In *Arxiv*.

Sam Wiseman and Alexander M. Rush. 2016. Sequence-to-Sequence Learning as Beam-Search Optimization. In *EMNLP*.

Sam Wiseman, Alexander M. Rush, and Stuart M. Shieber. 2017a. Challenges in Data-to-Document Generation. In *EMNLP*.

Sam Wiseman, Stuart M. Shieber, and Alexander M. Rush. 2017b. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 2253–2263.

Sam Wiseman, Stuart M Shieber, and Alexander M Rush. 2018. Learning neural templates for text generation. *arXiv preprint arXiv:1808.10122*.