CSCE 638 Natural Language Processing Foundation and Techniques

Lecture 21: Non-Autoregressive Generation

Kuan-Hao Huang

Spring 2026

Quiz 3

- Average: 78.74
- Std: 11.65
- Q1: 72
- Median: 81
- Q3: 86

Check Gradescope for details. For questions, send emails to csce638-ta-26s@lists.tamu.edu with "[CSCE 638] Subject …" or check with TA in office hours

Quiz 4

- April 13 (Monday)
- Coverage: mainly Lecture 16 to 20
  - Naturally include some concepts in Lecture 1 to 15
- In-class, 20 minutes, closed book, no cheat sheet
- Written quiz, 5 questions
  - Please bring a pen
- Tips
  - Get familiar with formula
  - Understand the intuition behind the formula and the design
  - Know the pros and cons of different approaches

Invited Talk

- **Date:** 4/15 online @ Zoom
  - https://tamu.zoom.us/my/khhuang?pwd=oAdWOKVOCGPApqDbJnVtktdW2AE6nb.1
  - **Talk:** Less Is More: Why Compression May Be the Missing Incentive for LLM Generalization
  - **Speaker:** Ben Zhou, Assistant Professor, Arizona State University

Invited Talk

Abstract:
Modern LLMs have such enormous capacity that they can memorize supervision signals without forming reusable abstractions. Humans, by contrast, are forced by biological and social constraints to compress experience into transferable concepts—arguably a key driver of general intelligence. We argue that compression should be an explicit training objective to encourage reusable, controllable generalization in large models. In this talk, I will draw connections to information-theoretic principles such as MDL and the information bottleneck, then present four of our recent works that operationalize compression as training-time signals via reinforcement learning, showing how constraining the information channel consistently forces models to reason rather than memorize.

Invited Talk

Speaker Bio:

Ben Zhou is an Assistant Professor in the School of Computing and Augmented Intelligence at Arizona State University. Ben's research uses data and symbolic cognitive processes to improve model reasoning, controllability, and trustworthiness from learning/inference schemes and architectural perspectives. Ben obtained his Ph.D. degree from the University of Pennsylvania. He is a recipient of the ENIAC fellowship from the University of Pennsylvania and a finalist for the CRA Outstanding Undergraduate Researcher Award. Additional information is available at http://xuanyu.me/.

Final Presentation

- Each team has 7 minutes for presentation
  - You have to stop once you reach 7 minutes
- The presentation should include
  - The topic you choose
  - Novelty/challenges
  - Your approach/design
  - Experiments/evaluation
  - Results, findings, insights/demo
- Clarity is important
  - Teach your classmate about your topic
- Time control is also important

Final Presentation

- Presentation dates
  - 4/20, 4/22, 4/27
- Online
  - https://tamu.zoom.us/my/khhuang?pwd=oAdWOKVOCGPApqDbJnVtktdW2AE6nb.1
- Presentation order
  - https://docs.google.com/spreadsheets/d/1qUZPFI4wciToJsXye8-WN4L7xVG38IWdS2GCCzmu84A/edit?usp=sharing
- Everyone has to join Zoom with your name displayed

Recap: Autoregressive Language Models

$$
\begin{aligned}
P(w_1, w_2, w_3, \ldots, w_l) &= P(w_1)P(w_2, w_3, \ldots, w_l|w_1) \\
&= P(w_1)P(w_2|w_1)(w_3, \ldots, w_l|w_1, w_2) \\
&= P(w_1)P(w_2|w_1)P(w_3|w_1, w_2)(w_4, \ldots, w_l|w_1, w_2, w_3) \\
&= \prod_{i=1}^{l} P(w_i|w_1, w_2, \ldots, w_{i-1})
\end{aligned}
$$

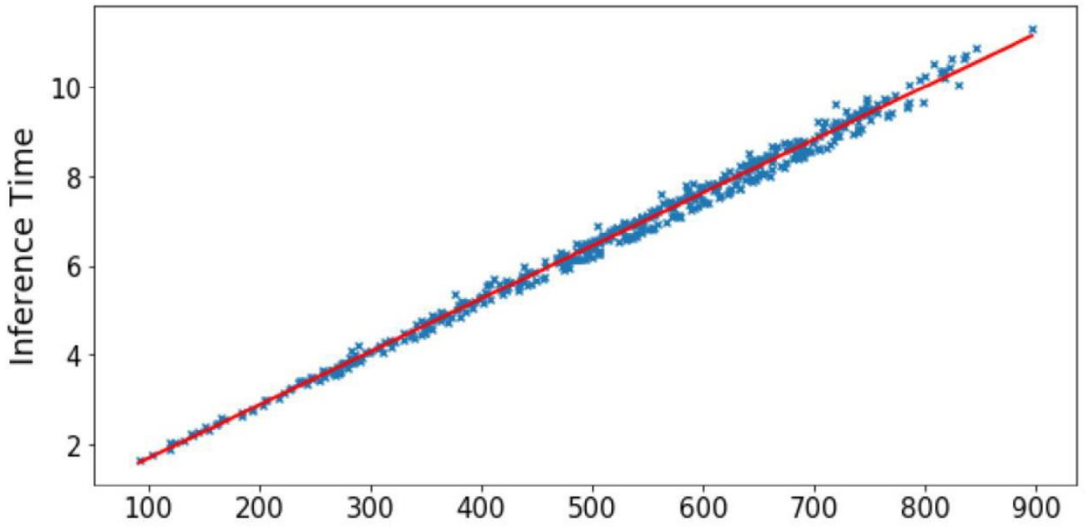
$$
\begin{aligned}
P(\textit{She likes to go hiking}) = P(\textit{She}) \cdot P(\text{likes}|\text{She}) \cdot P(\text{to}|\text{She likes}) \\
\cdot P(\text{go}|\text{She likes to}) \cdot P(\text{hiking}|\text{She likes to go})
\end{aligned}
$$

Recap: Autoregressive Language Models

Binge ... on | - | and | of | is
Binge drinking ... is | and | had | in | was
Binge drinking may ... be | also | have | not | increase
Binge drinking may not ... be | have | cause | always | help
Binge drinking may not necessarily ... be | lead | cause | results | have
Binge drinking may not necessarily kill ... you | the | a | people | your
Binge drinking may not necessarily kill or ... even | injure | kill | cause | prevent
Binge drinking may not necessarily kill or even ... kill | prevent | cause | reduce | injure
Binge drinking may not necessarily kill or even damage ... your | the | a | you | someone
Binge drinking may not necessarily kill or even damage brain ... cells | functions | tissue | neurons
Binge drinking may not necessarily kill or even damage brain cells, ... some | it | the | is | long

Non-Autoregressive Generation

- Can we generate text in ways other than word by word?
  - Other order
  - Parallel decoding
- Why?
  - Faster decoding



Non-Autoregressive Generation

- Can we generate text in ways other than word by word?
  - Other order
  - Parallel decoding
- Why?
  - Faster decoding
  - More similar to human writing (revision)

There is a bird on the tree.

There is a red bird on the green tree.

There is a red bird with a big beak on the green tree.

Document Editing

INTRODUCTION

Several initiatives have been launched in many countries with the aim of modernizing ~~the~~ public services. In this sense different reports and documents have indicated the need for investing in technologies ~~for~~ to offer~~ing~~ better services to citizens and organizations (Department of Public Expenditure and Reform, 2011; United Nations, 2012) and thus reduce~~ing~~ the burden for them. This trend is named "electronic g~~G~~overnment (e-g~~G~~overnment)," and it can be defined as the "the use of information technology to enable and improve the efficiency with which government services are provided to citizens, employees, businesses, and agencies" (Carter & Belanger, 2005).

Although several countries have improved their services through the use of more sophisticated ~~web~~ Web pages, according to the Department of Public Expenditure and Reform (2012), countries still need to introduce more technologies and automated processes with the aim of reducing the burden of processes currently needed for citizens and organizations. In several public processes the citizens and organization must ~~have to~~ present physical documents ~~which~~ that are delivered manually from one section to another ~~section~~, sometimes producing ~~sometimes~~ delays in the delivery due to human causes such as illness~~es~~, oversight, or overwork of the public worker. By means of automated processes, this documentation is immediately available to the next section or administration in the business process once the documentation has been analyzed and completed by the corresponding section or administration. In this sense public bodies must~~have to~~ ensure that the sharing of data between different public organizations provides a reduction ~~of~~ in the number of times citizens or businesses must~~have to~~ ask for data.

https://www.econtentpro.com/copyediting/samples

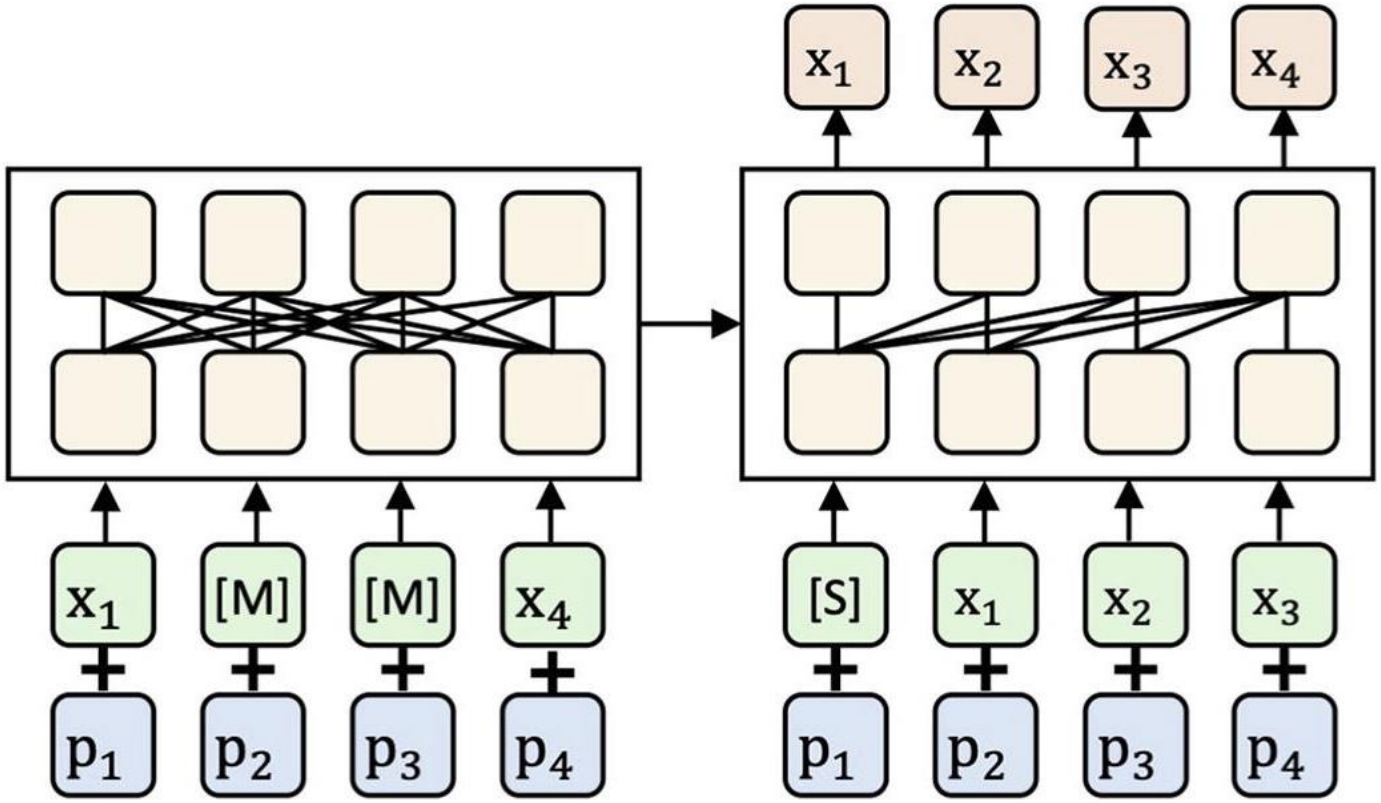
Challenges

- How to add words in the middle?
- Parallel decoding → fluency issue

Semi-Autoregressive Neural Machine Translation

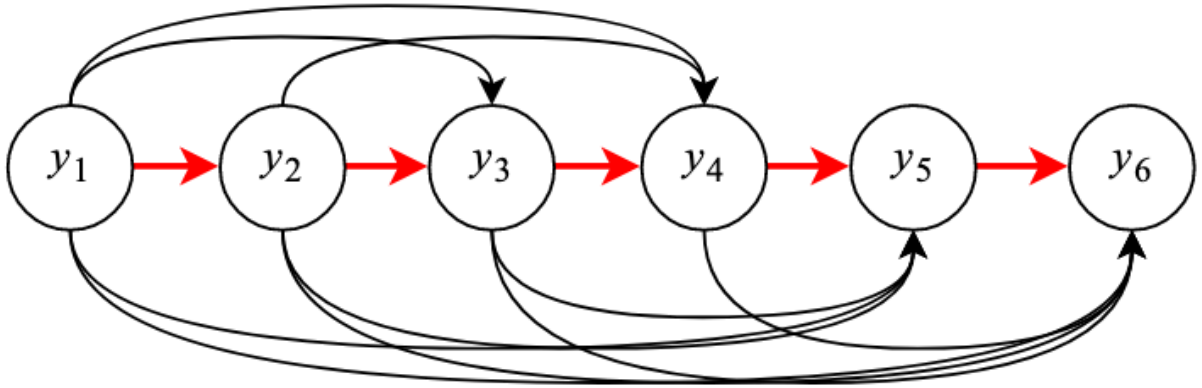
Chunqi Wang* **Ji Zhang** **Haiqing Chen**
Alibaba Group
{shiyuan.wcq, zj122146, haiqing.chenhq}@alibaba-inc.com

Machine Translation with Seq2Seq

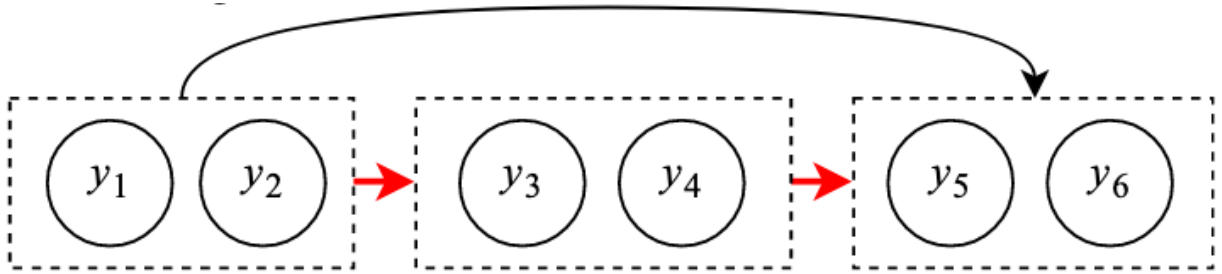


Semi-Autoregressive Decoding

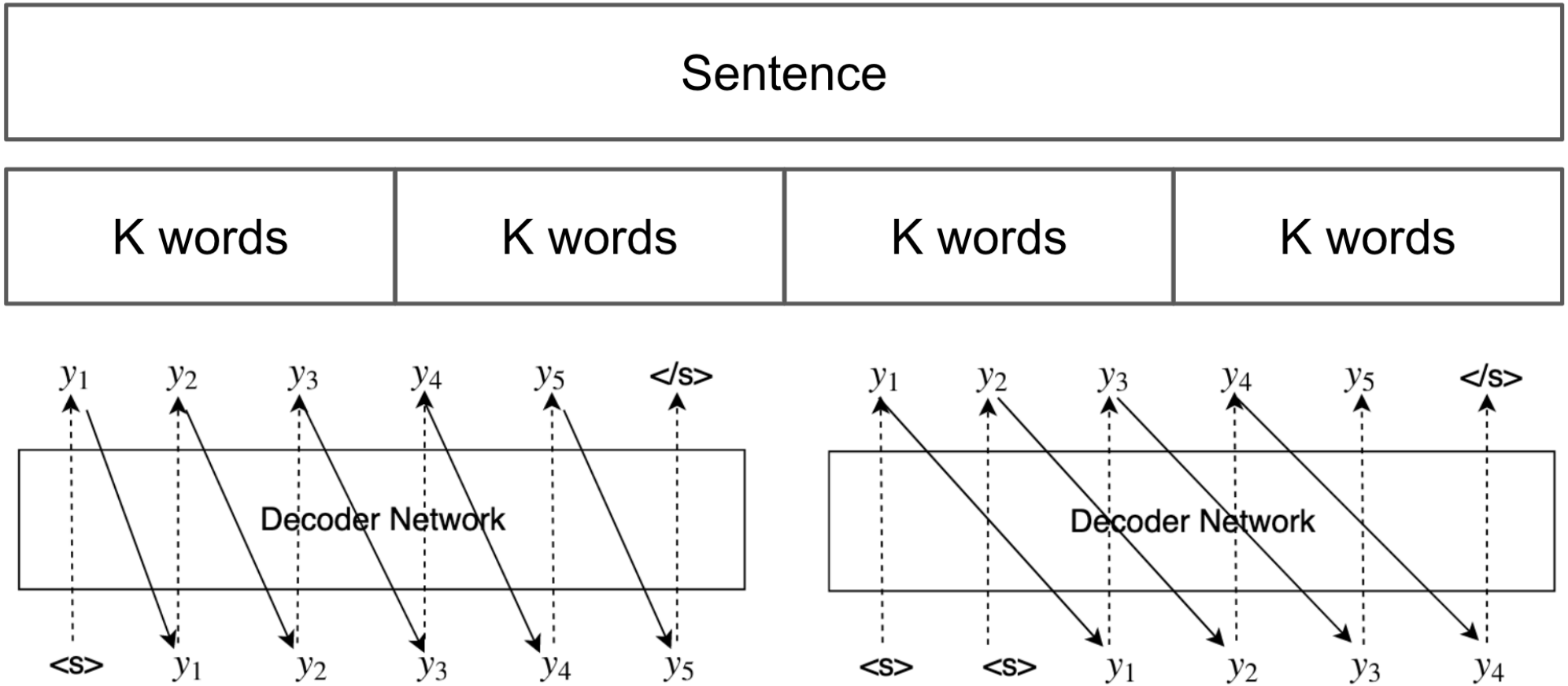
- Autoregressive decoding



- Semi-autoregressive decoding



Semi-Autoregressive Transformer



Mask Modification

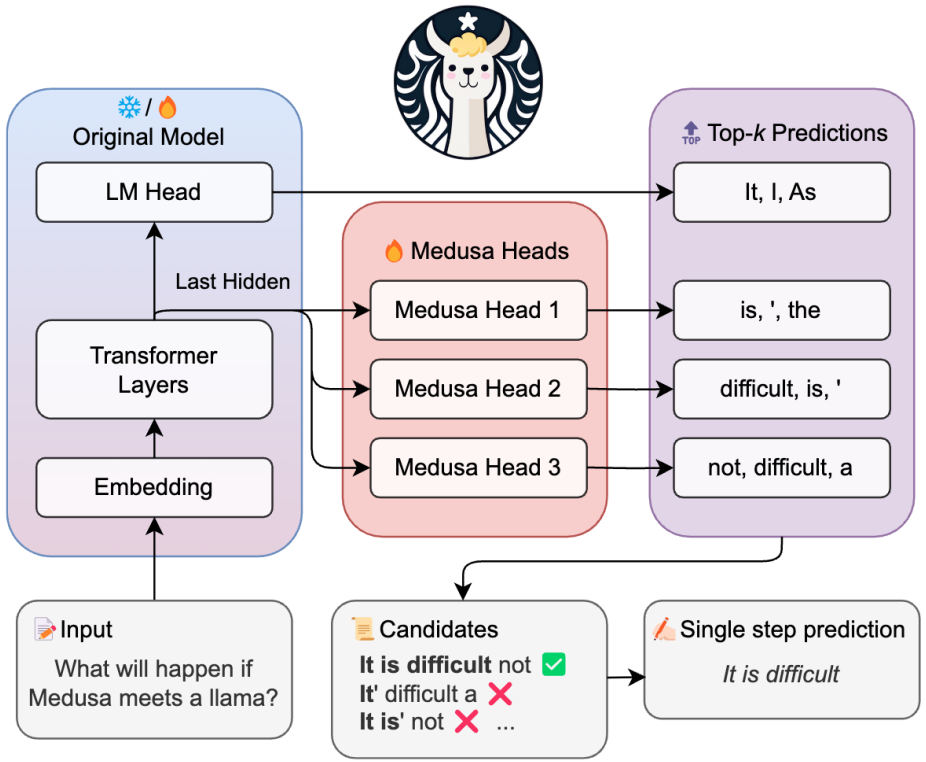
$$\begin{bmatrix} 1 & \mathbf{0} & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & \mathbf{0} & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & \mathbf{0} \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & \mathbf{1} & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & \mathbf{1} \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

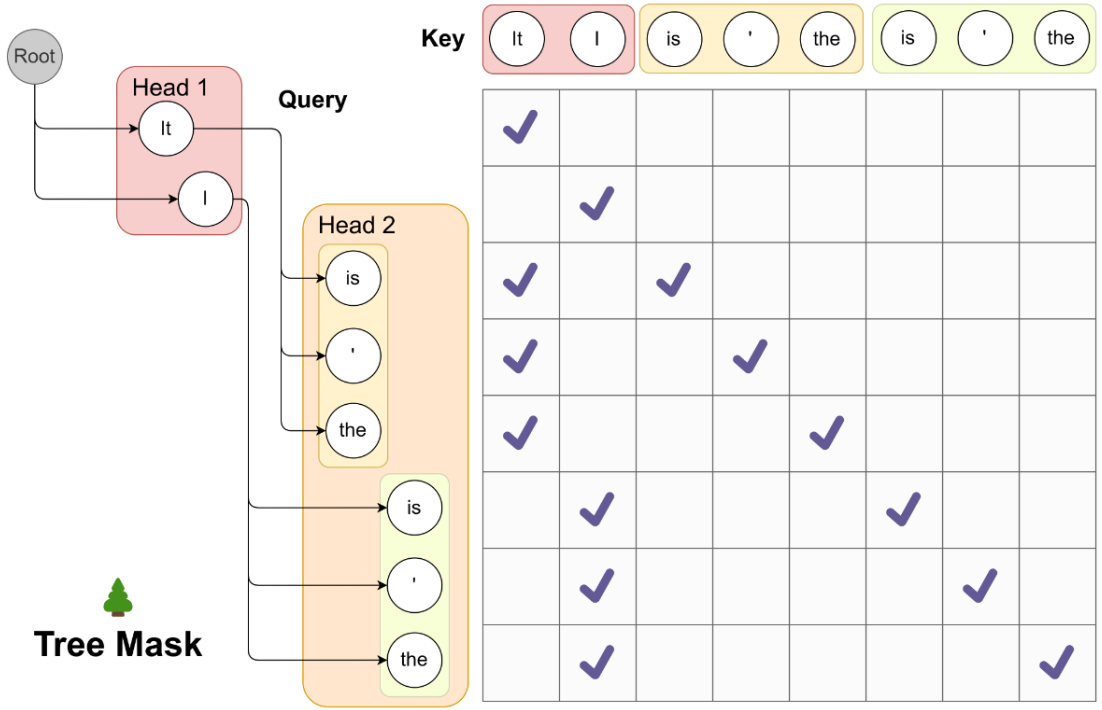
Machine Translation Performance

Model	Beam Size	BLEU	Degeneration	Latency	Speedup
Transformer	4	27.11	0%	346ms	1.00×
	1	26.01	4%	283ms	1.22×
Transformer, N=2	4	24.30	10%	163ms	2.12×
	1	23.37	14%	113ms	3.06×
This Work					
SAT, K=2	4	26.90	1%	229ms	1.51×
	1	26.09	4%	167ms	2.07×
SAT, K=4	4	25.71	5%	149ms	2.32×
	1	24.67	9%	91ms	3.80×
SAT, K=6	4	24.83	8%	116ms	2.98×
	1	23.93	12%	62ms	5.58×

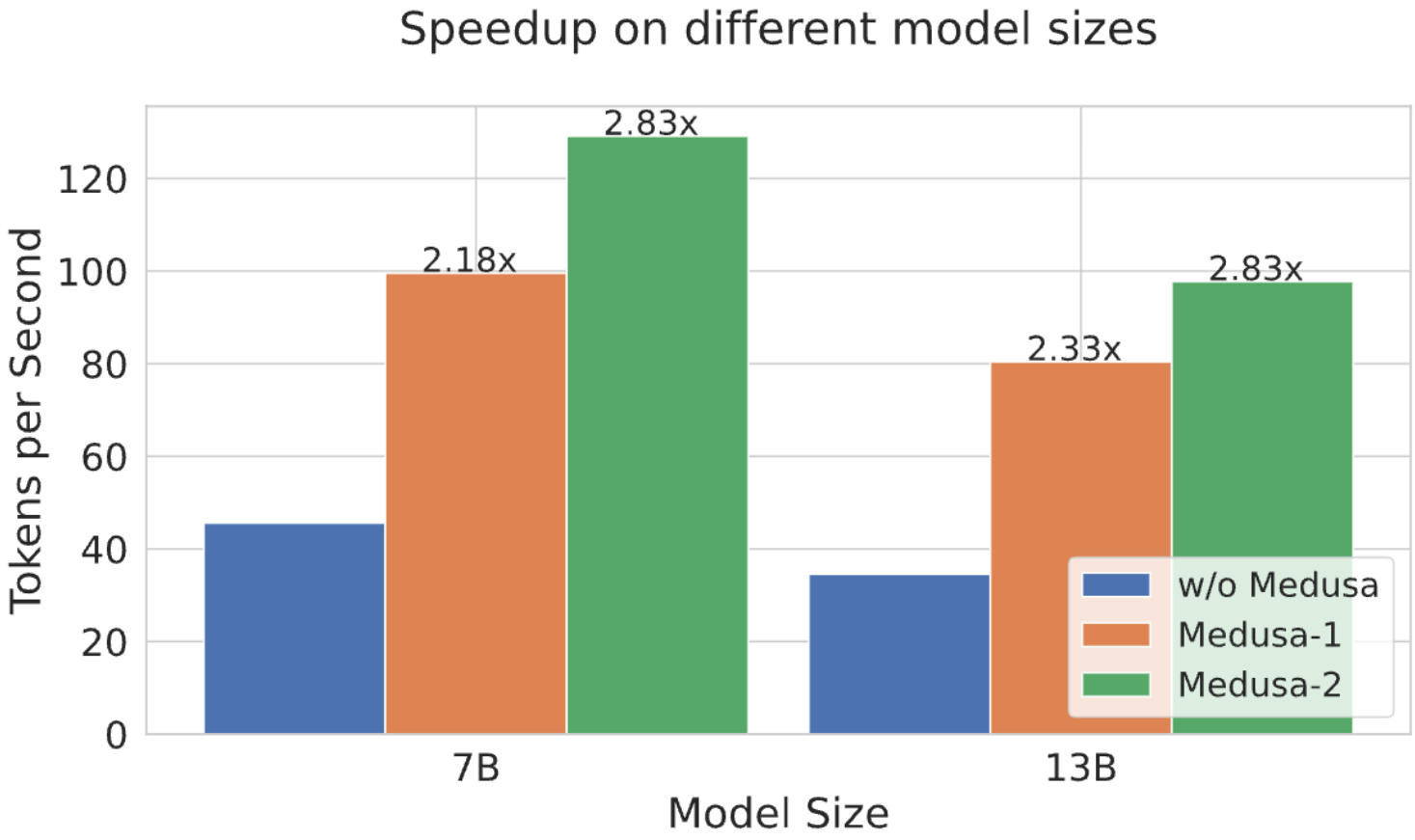
MEDUSA: Simple LLM Inference Acceleration Framework with Multiple Decoding Heads

Tianle Cai [*1,2] Yuhong Li [*3] Zhengyang Geng [4] Hongwu Peng [5] Jason D. Lee [1] Deming Chen [3] Tri Dao [1,2]





Decoding Speed

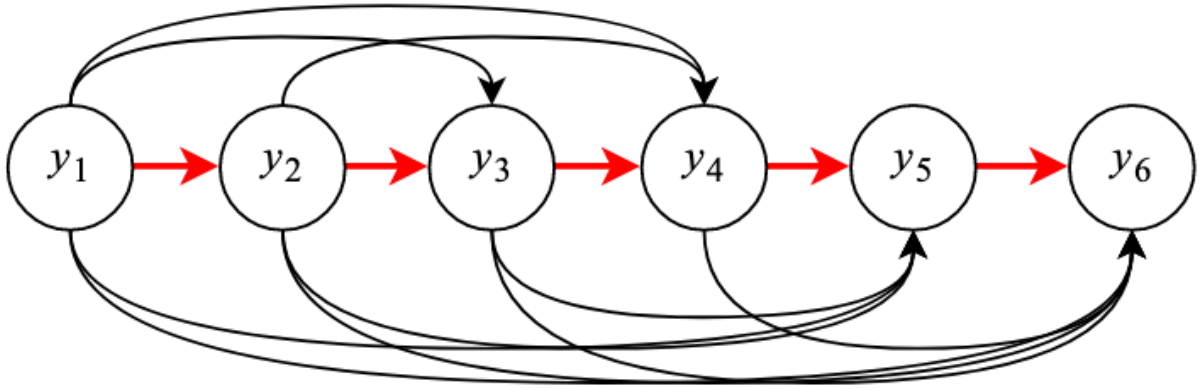


NON-AUTOREGRESSIVE NEURAL MACHINE TRANSLATION

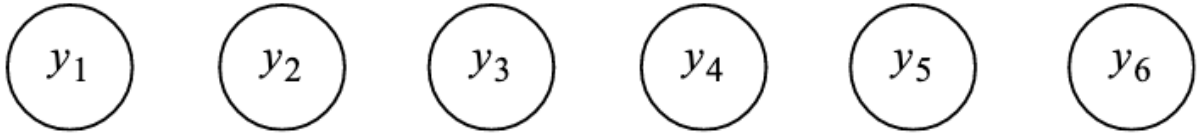
Jiatao Gu[†,*], James Bradbury[‡], Caiming Xiong[‡], Victor O.K. Li[†] & Richard Socher[‡]
[‡]Salesforce Research
{james.bradbury,cxiong,rsocher}@salesforce.com
[†]The University of Hong Kong
{jiataogu, vli}@eee.hku.hk

Fully Non-Autoregressive Decoding

- Autoregressive decoding

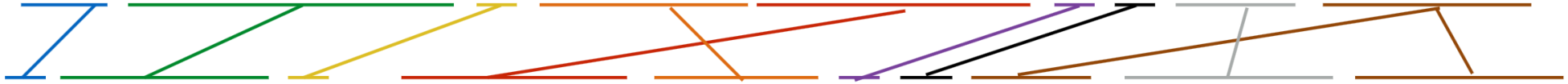


- Fully non-autoregressive decoding



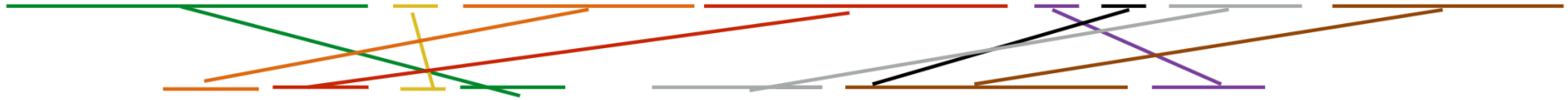
Motivation (Word Alignment)

The development of artificial intelligence is a really big deal.



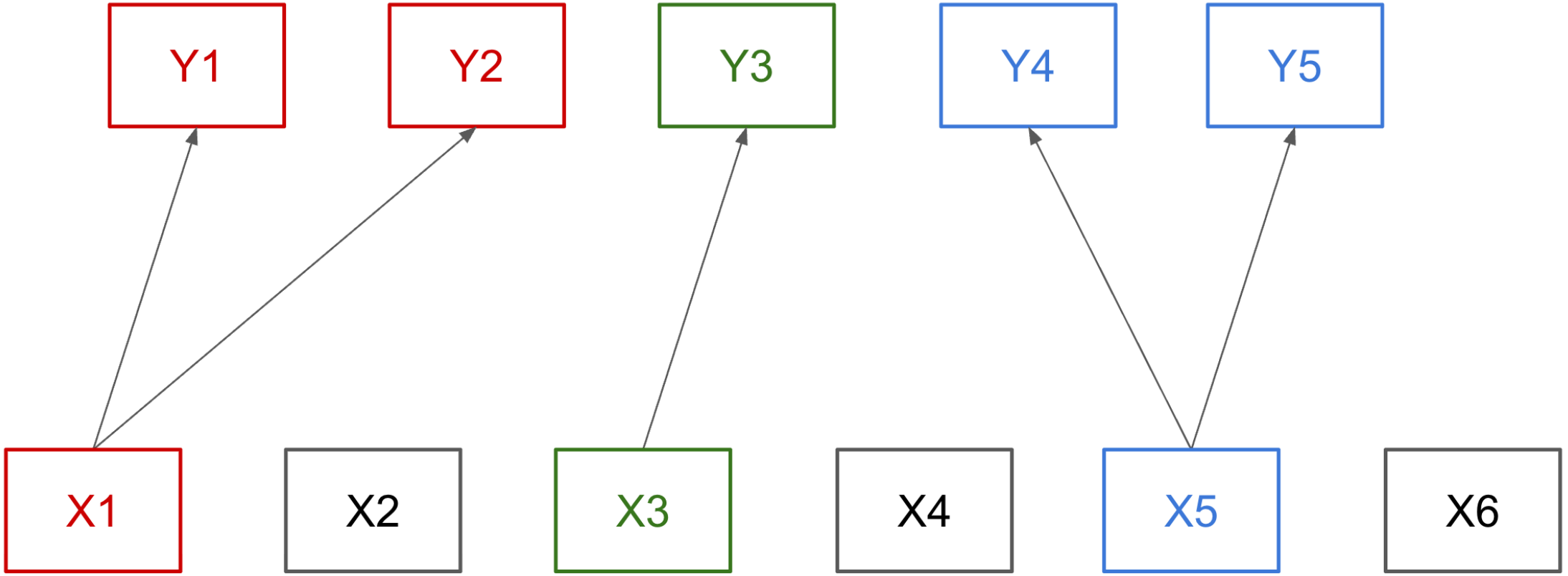
El desarrollo de la inteligencia artificial es un asunto realmente importante.

The development of artificial intelligence is a really big deal.

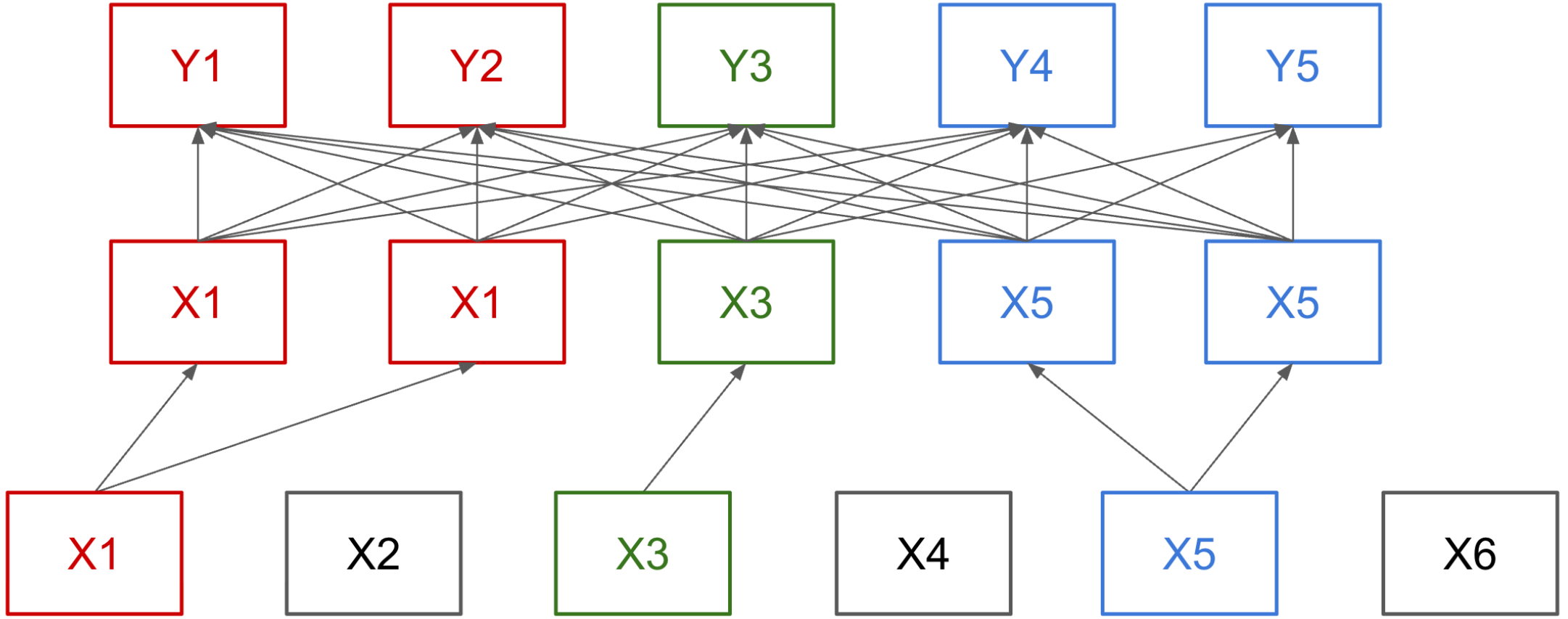


人工知能の発展は本当にすごいことです。

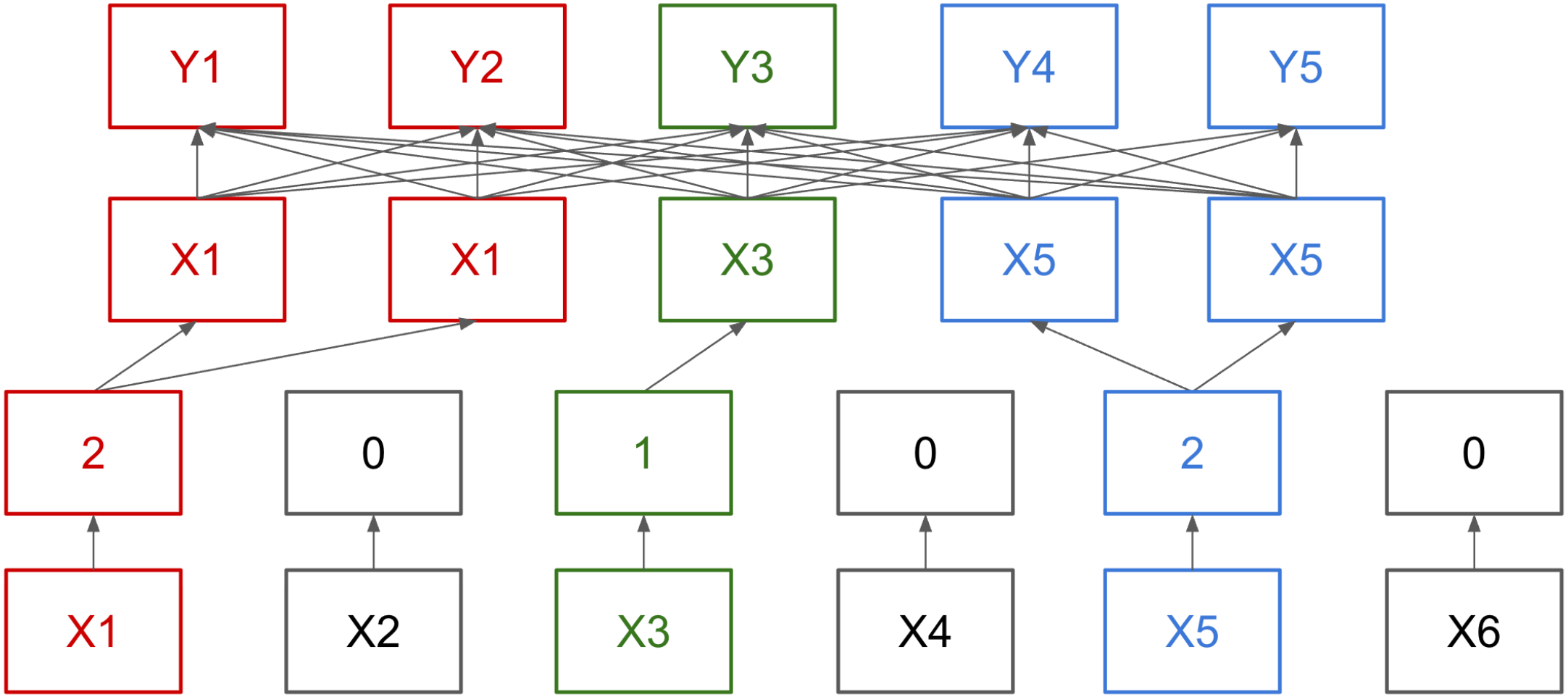
Motivation (Word Alignment)



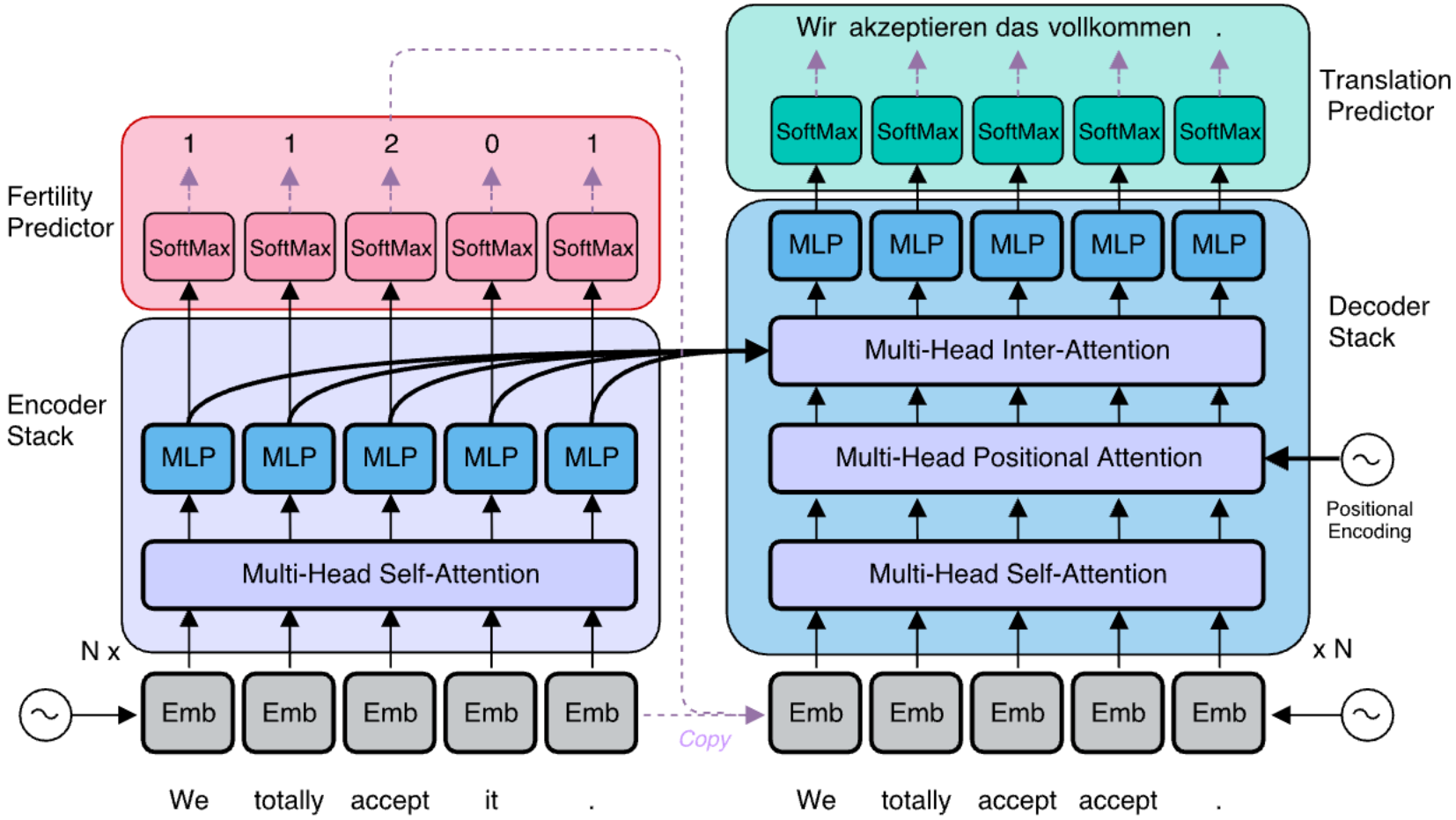
Motivation (Word Alignment)



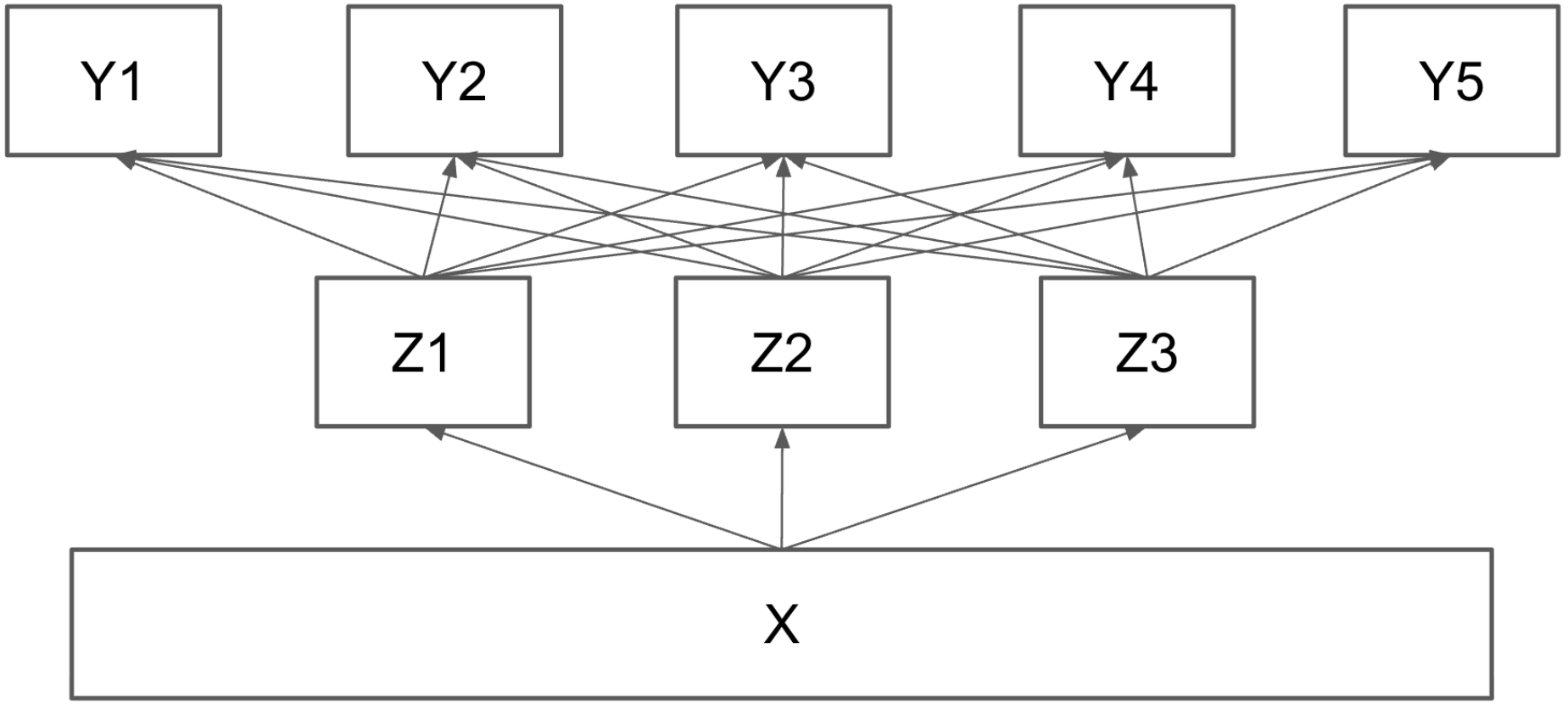
Fertility Predictor



Fertility Predictor



Decode with Latent Variables (A More General Version)



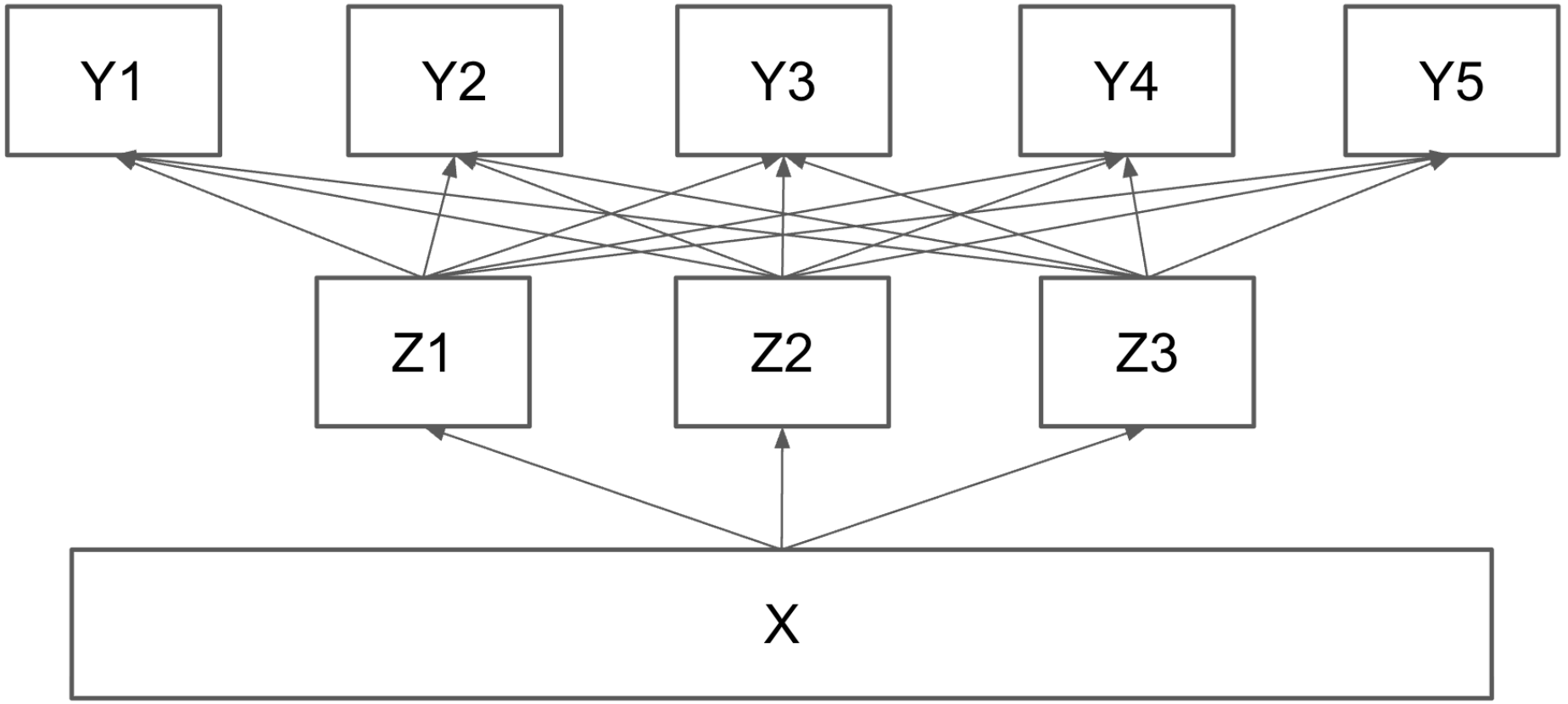
Results

Models	WMT14		WMT16		IWSLT16		
	En→De	De→En	En→Ro	Ro→En	En→De	Latency	Speedup
NAT	17.35	20.62	26.22	27.83	25.20	39 ms	15.6×
NAT (+FT)	17.69	21.47	27.29	29.06	26.52	39 ms	15.6×
NAT (+FT + NPD $s = 10$)	18.66	22.41	29.02	30.76	27.44	79 ms	7.68×
NAT (+FT + NPD $s = 100$)	19.17	23.20	29.79	**31.44**	28.16	257 ms	2.36×
Autoregressive ($b = 1$)	22.71	26.39	31.35	31.03	28.89	408 ms	1.49×
Autoregressive ($b = 4$)	23.45	27.02	31.91	31.76	29.70	607 ms	1.00×

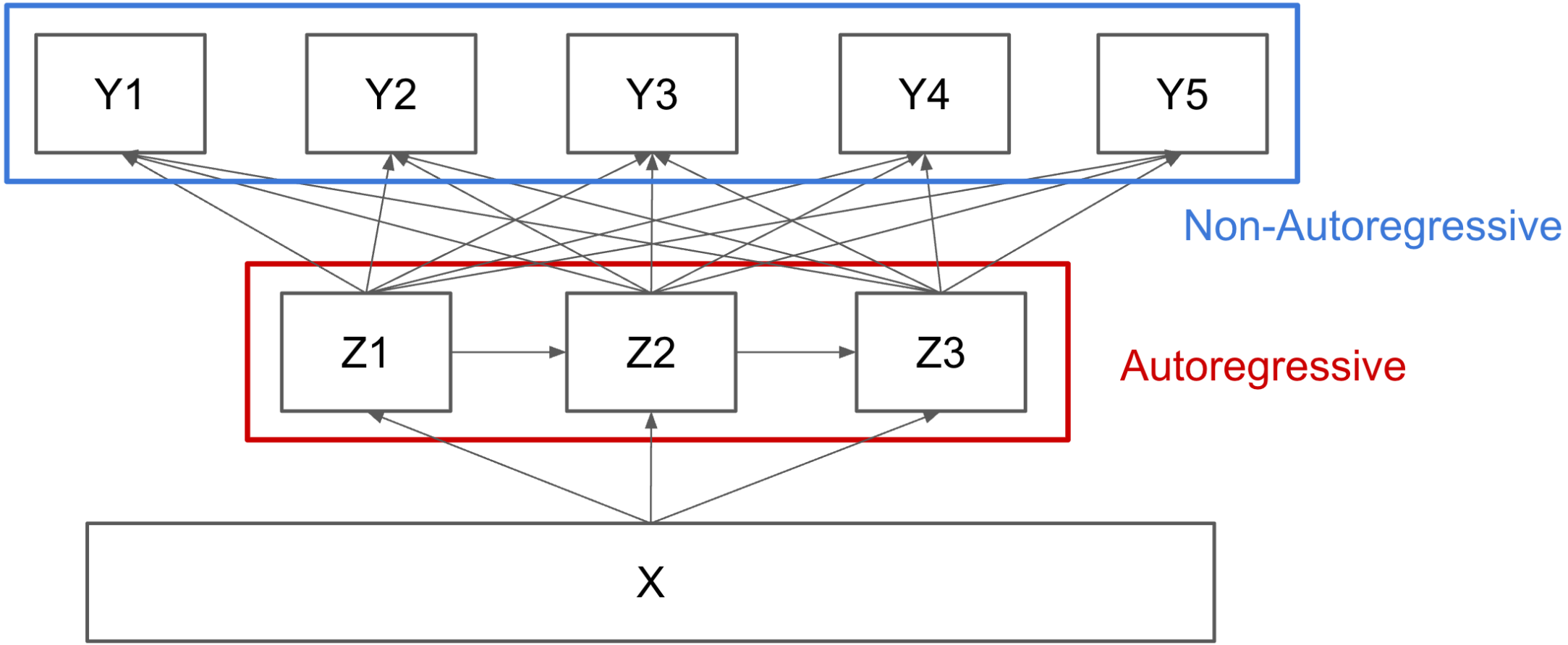
Fast Decoding in Sequence Models Using Discrete Latent Variables

Łukasz Kaiser [1] **Aurko Roy** [1] **Ashish Vaswani** [1] **Niki Parmar** [1] **Samy Bengio** [1] **Jakob Uszkoreit** [1]
Noam Shazeer [1]

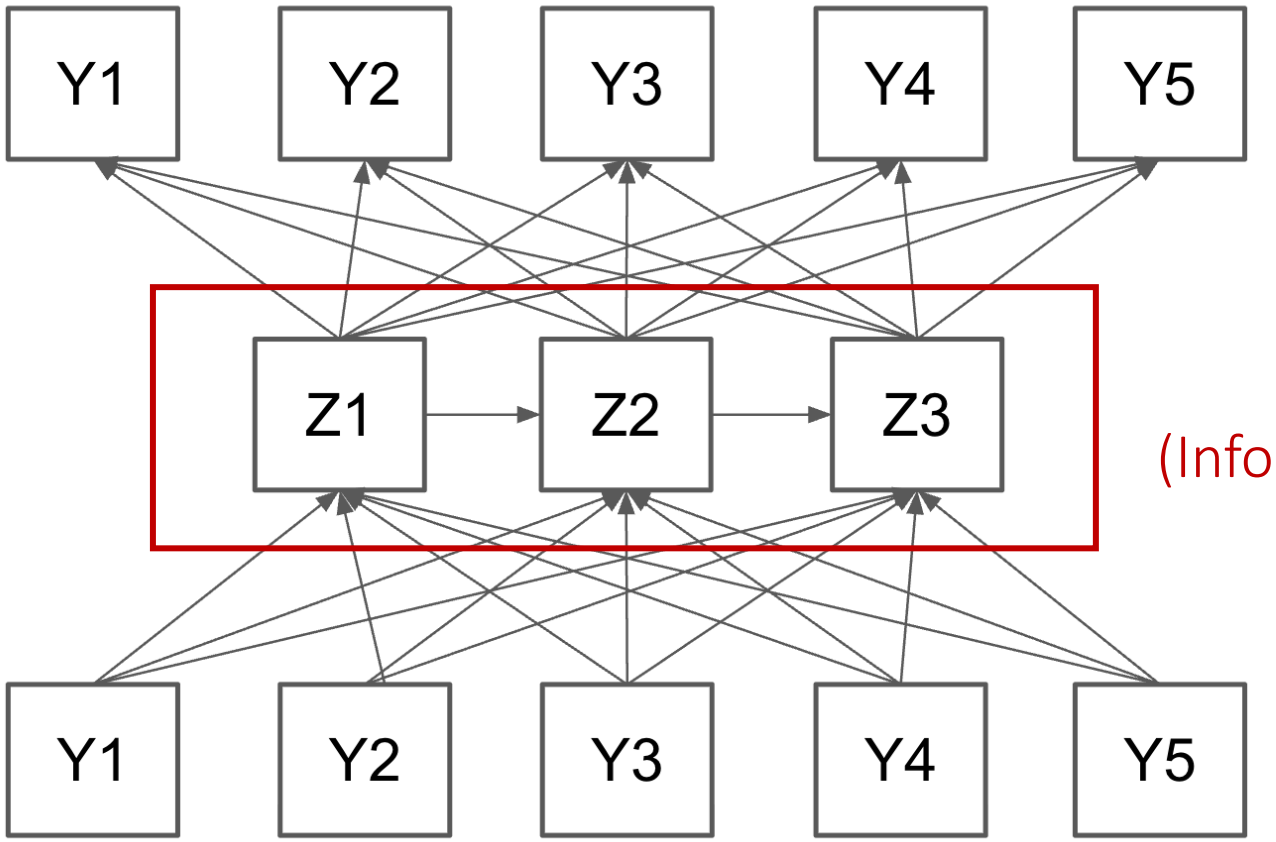
Decode with Latent Variables



Decode with Autoregressive Latent Variables



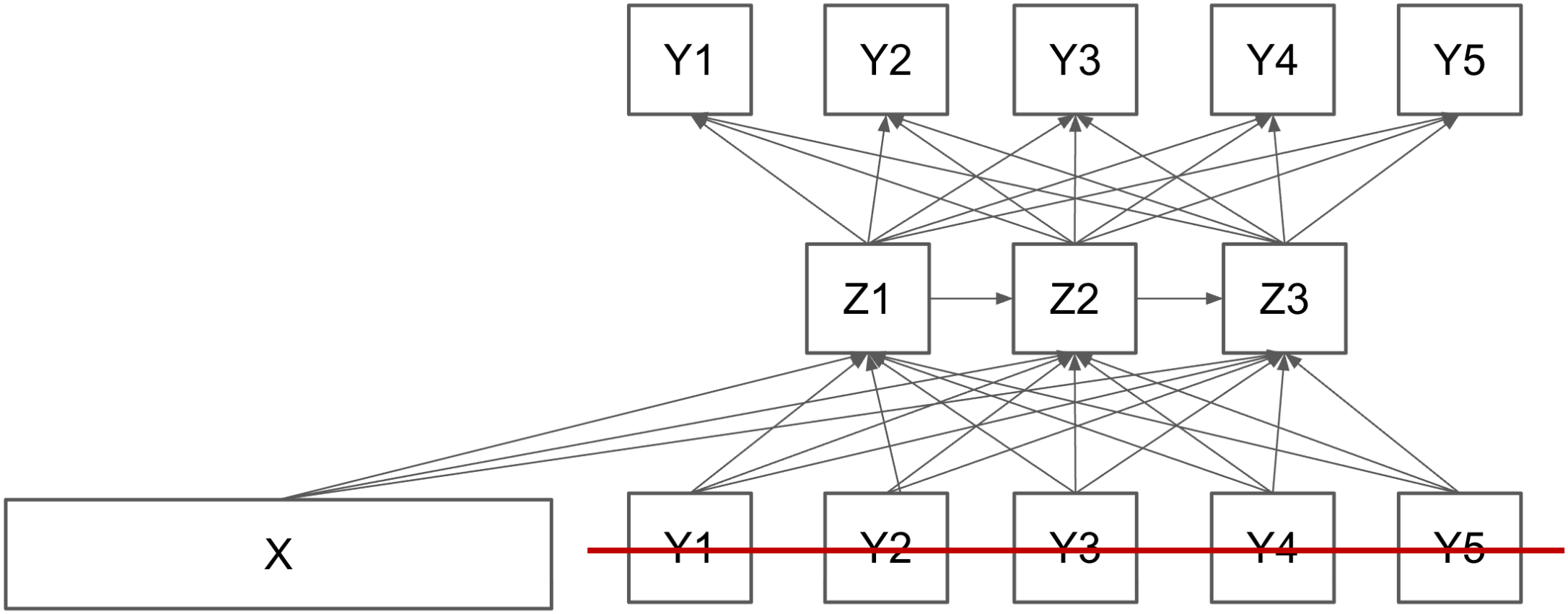
How to Decide Latent Variables?



Discrete values
(Information Bottleneck)

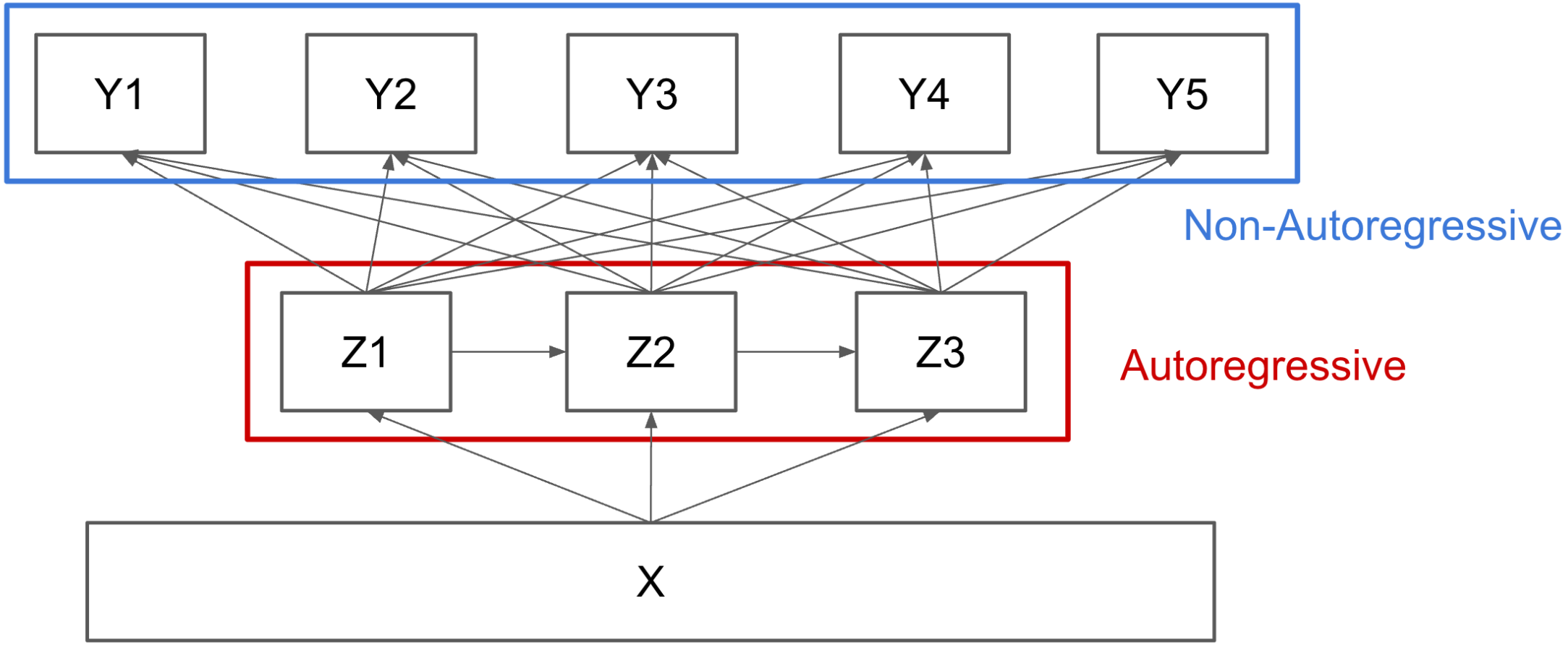
Reconstruction from latent variables

Learning Mapping Between Inputs and Latent Variables



Predict Output Length

A separate classifier to predict the output length



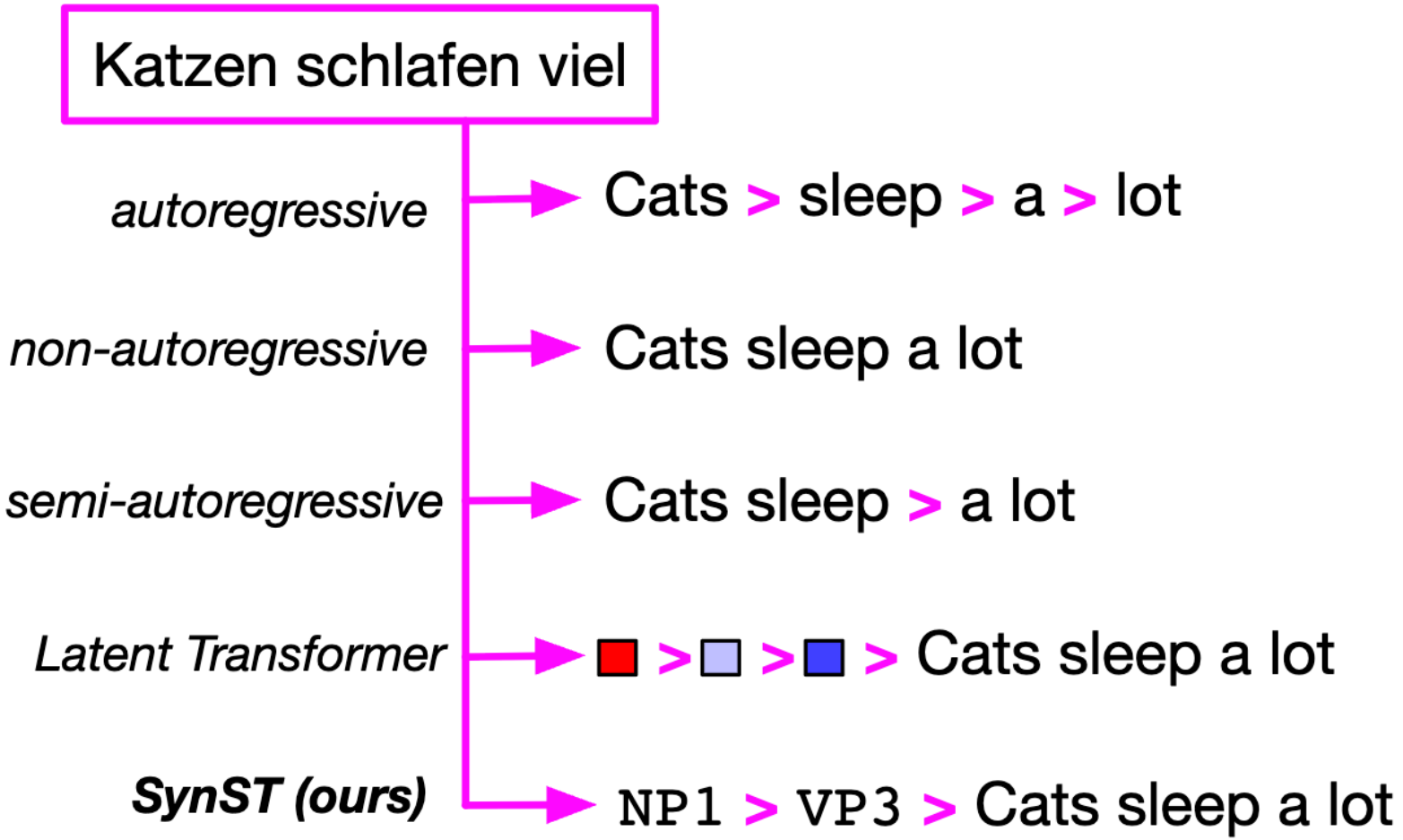
Results

Model	BLEU	**Latency**	
		$b = 1$	$b = 64$
Baseline (no beam-search)	22.7	408 ms	-
NAT	17.7	39 ms	-
NAT+NPD=10	18.7	79 ms	-
NAT+NPD=100	19.2	257 ms	-
LT, Improved Semhash	19.8	105 ms	8 ms
LT, VQ-VAE	2.78	148 ms	7 ms
LT, s-DVQ	19.7	177 ms	7 ms
LT, p-DVQ	19.8	182 ms	8 ms

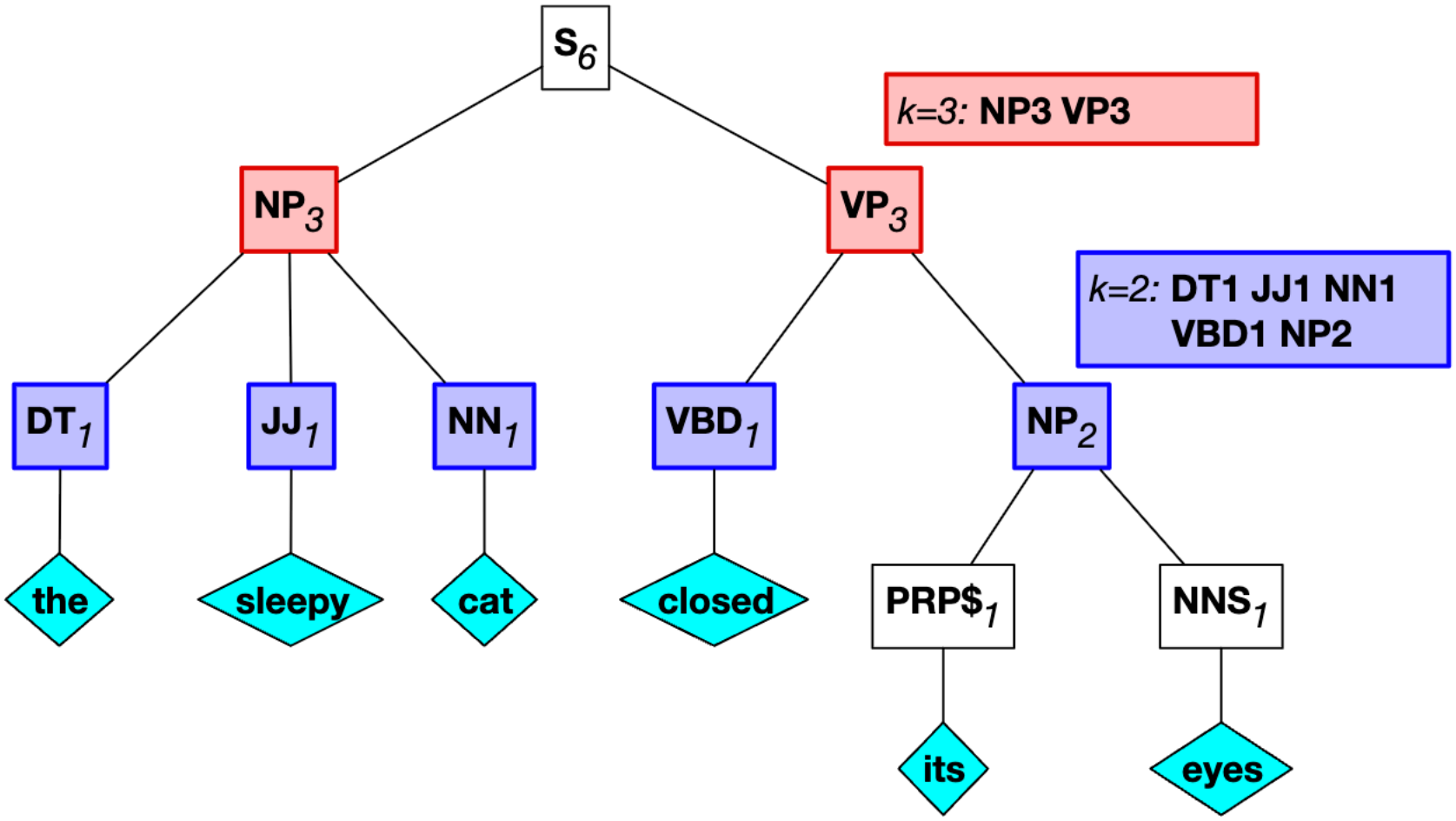
Syntactically Supervised Transformers for Faster Neural Machine Translation

Nader Akoury, Kalpesh Krishna, Mohit Iyyer
College of Information and Computer Sciences
University of Massachusetts Amherst
{nsa,kalpesh,miyyer}@cs.umass.edu

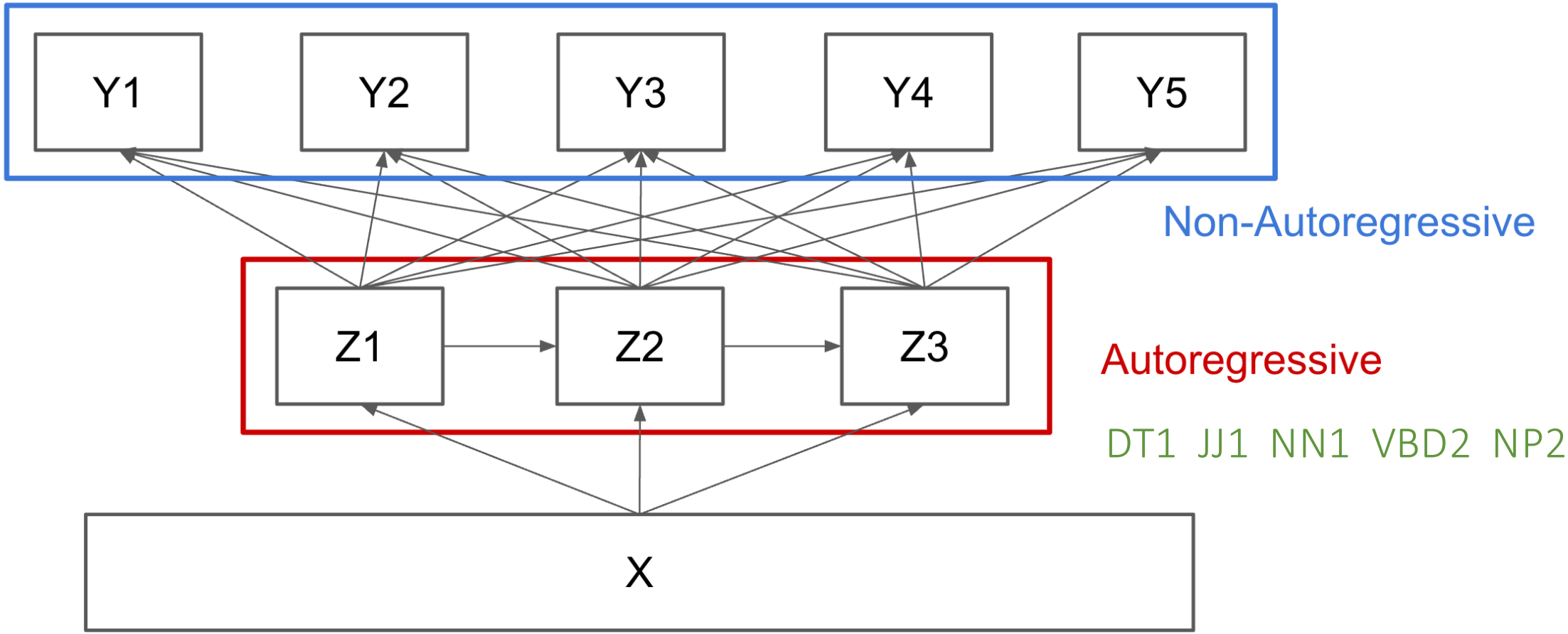
Consider Syntactic Information for Latent Variables



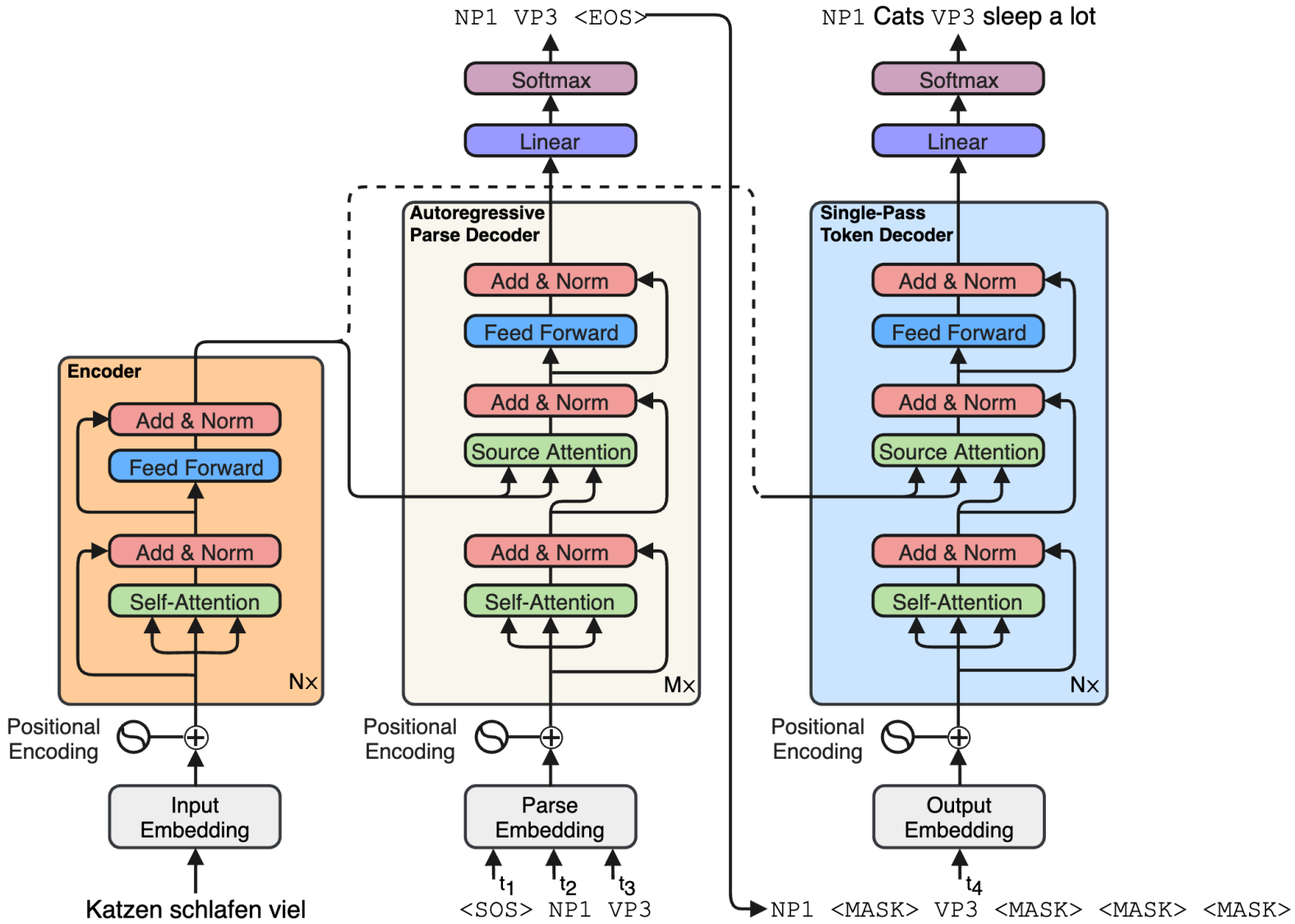
Syntactic Information: Constituency Parse Trees



Consider Syntactic Information for Latent Variables



Syntactically Supervised Transformer (SynST)



Results

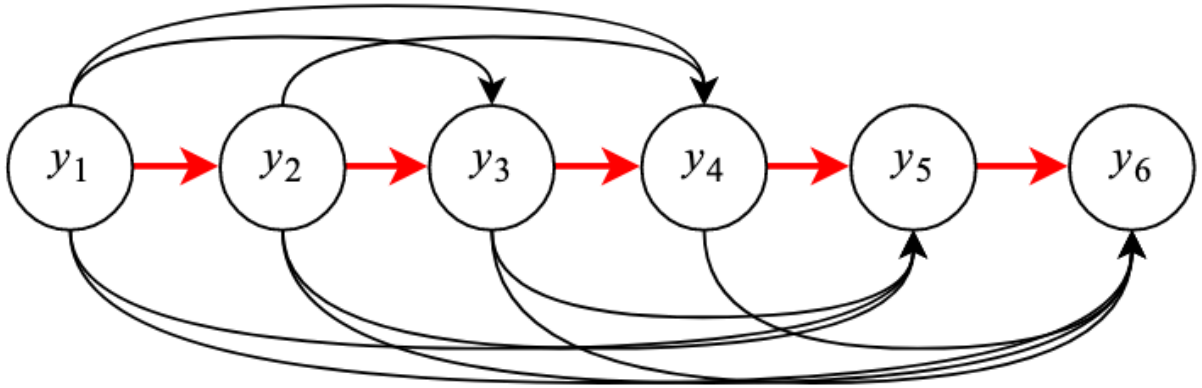
Model	WMT En-De		WMT De-En		IWSLT En-De		WMT En-Fr	
	BLEU	Speedup	BLEU	Speedup	BLEU	Speedup	BLEU	Speedup
Baseline ($b = 1$)	25.82	1.15×	29.83	1.14×	28.66	1.16×	39.41	1.18×
Baseline ($b = 4$)	26.87	1.00×	30.73	1.00×	30.00	1.00×	40.22	1.00×
SAT ($k = 2$)	22.81	2.05×	26.78	2.04×	25.48	2.03×	36.62	2.14×
SAT ($k = 4$)	16.44	3.61×	21.27	3.58×	20.25	3.45×	28.07	3.34×
SAT ($k = 6$)	12.55	4.86×	15.23	4.27×	14.02	4.39×	24.63	4.77×
LT*	19.8	3.89×	-	-	-	-	-	-
SynST($k = 6$)	20.74	4.86×	25.50	5.06×	23.82	3.78×	33.47	5.32×

Mask-Predict: Parallel Decoding of Conditional Masked Language Models

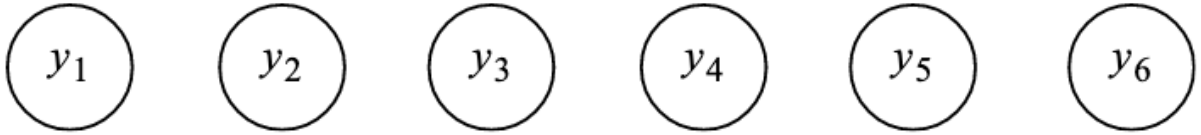
Marjan Ghazvininejad* **Omer Levy*** **Yinhan Liu*** **Luke Zettlemoyer**
Facebook AI Research
Seattle, WA

Fully Non-Autoregressive Decoding

- Autoregressive decoding



- Fully non-autoregressive decoding



How about iterative refinement?

Mask-Predict: Iterative Refinement

- **Input:** Der Abzug der franzsischen Kampftruppen wurde am 20. November abgeschlossen
- **Step 1:** Predict the output length based on the input (12)

Mask-Predict: Iterative Refinement

- **Step 2:** Iterative non-autoregressive refinement

[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]
The	departure	departure	the	French	combat	completed	completed	on	20	November
0.9	0.2	0.3	0.1	0.4	0.2	0.2	0.2	0.1	0.6	0.7
The	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	20	November
	departure	of	French	combat	troops	troops	completed	on		
	0.2	0.9	0.8	0.7	0.8	0.1	0.2	0.9		
The	[Mask]	of	French	combat	troops	[Mask]	[Mask]	on	20	November
	withdrawal					was	completed			
	0.9					0.8	0.9			
The	withdrawal	of	French	combat	troops	was	completed	on	20	November

Results

Model	Dimensions (Model/Hidden)	Iterations	WMT'14 EN-DE	WMT'14 DE-EN	WMT'16 EN-RO	WMT'16 RO-EN
NAT w/ Fertility (Gu et al., 2018)	512/512	1	19.17	23.20	29.79	31.44
CTC Loss (Libovický and Helcl, 2018)	512/4096	1	17.68	19.80	19.93	24.71
Iterative Refinement (Lee et al., 2018)	512/512	1	13.91	16.77	24.45	25.73
	512/512	10	21.61	25.48	29.32	30.19
(Dynamic #Iterations)	512/512	?	21.54	25.43	29.66	30.30
Small CMLM with Mask-Predict	512/512	1	15.06	19.26	20.12	20.36
	512/512	4	**24.17**	**28.55**	**30.00**	30.43
	512/512	10	**25.51**	**29.47**	**31.65**	**32.27**
Base CMLM with Mask-Predict	512/2048	1	18.05	21.83	27.32	28.20
	512/2048	4	**25.94**	**29.90**	**32.53**	**33.23**
	512/2048	10	**27.03**	**30.53**	**33.08**	**33.31**
Base Transformer (Vaswani et al., 2017)	512/2048	N	27.30	— —	— —	— —
Base Transformer (Our Implementation)	512/2048	N	27.74	31.09	34.28	33.99
Base Transformer (+Distillation)	512/2048	N	27.86	31.07	— —	— —
Large Transformer (Vaswani et al., 2017)	1024/4096	N	28.40	— —	— —	— —
Large Transformer (Our Implementation)	1024/4096	N	28.60	31.71	— —	— —

Diffusion-LM Improves Controllable Text Generation

Xiang Lisa Li
Stanford University
xlisali@stanford.edu

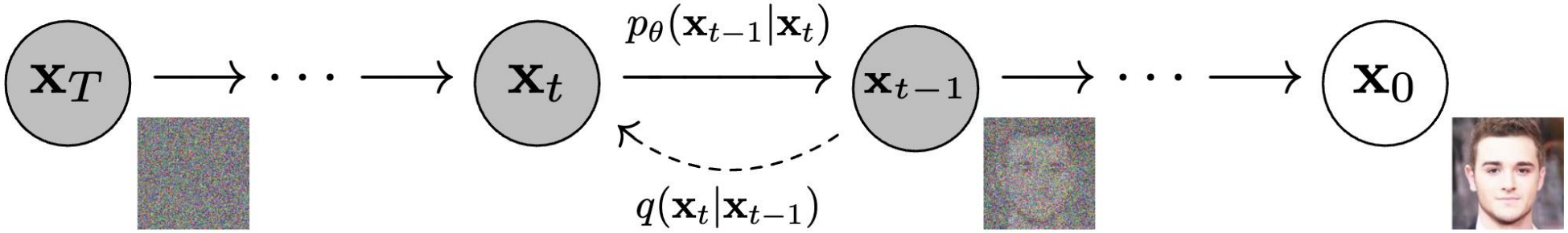
John Thickstun
Stanford University
jthickst@stanford.edu

Ishaan Gulrajani
Stanford Univeristy
igul@stanford.edu

Percy Liang
Stanford Univeristy
pliang@cs.stanford.edu

Tatsunori B. Hashimoto
Stanford Univeristy
thashim@stanford.edu

Image Diffusion



$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$$

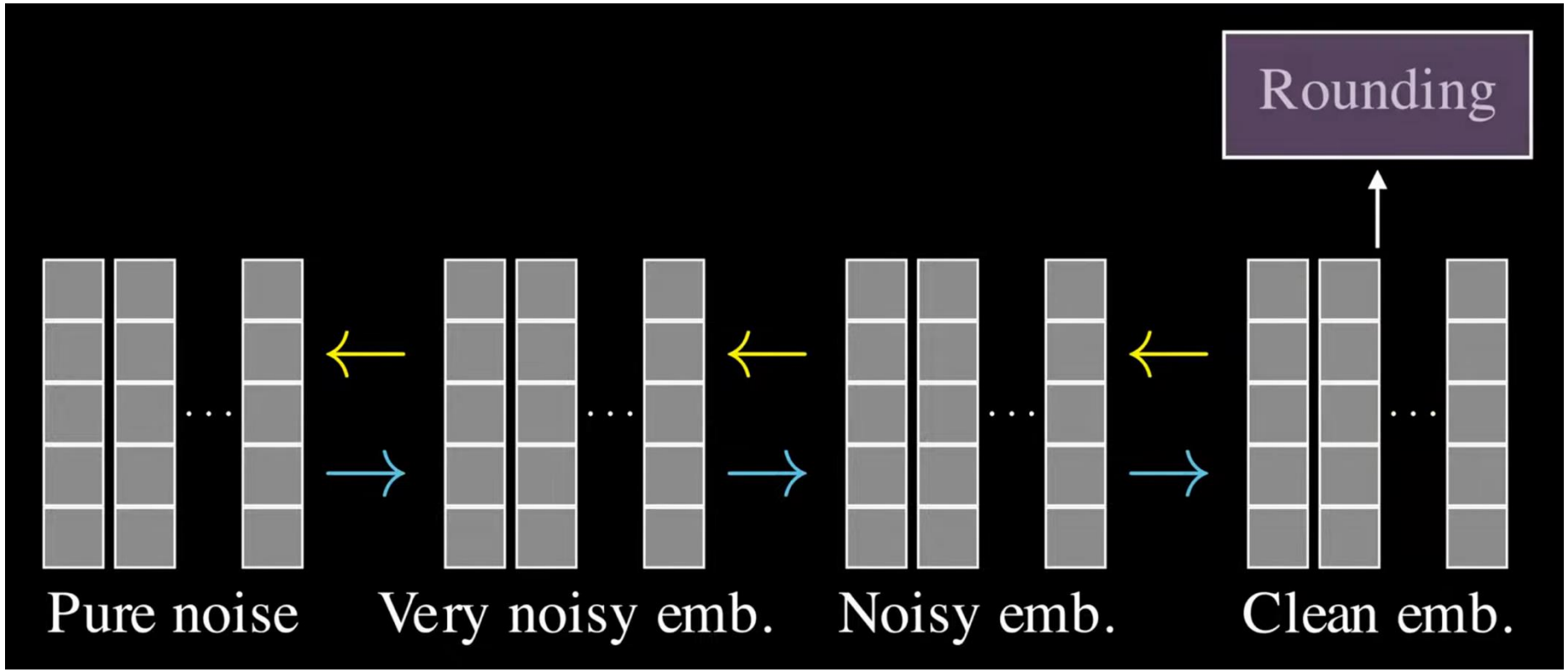
Gradually add noise to image

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$$

Learn to denoise

Image Diffusion Examples

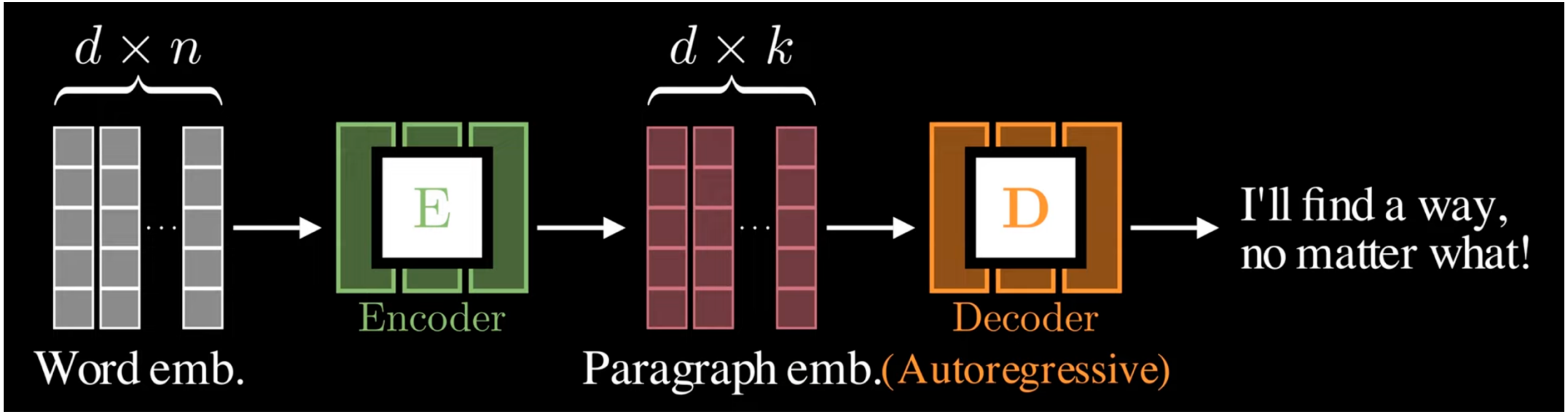
How About Text?

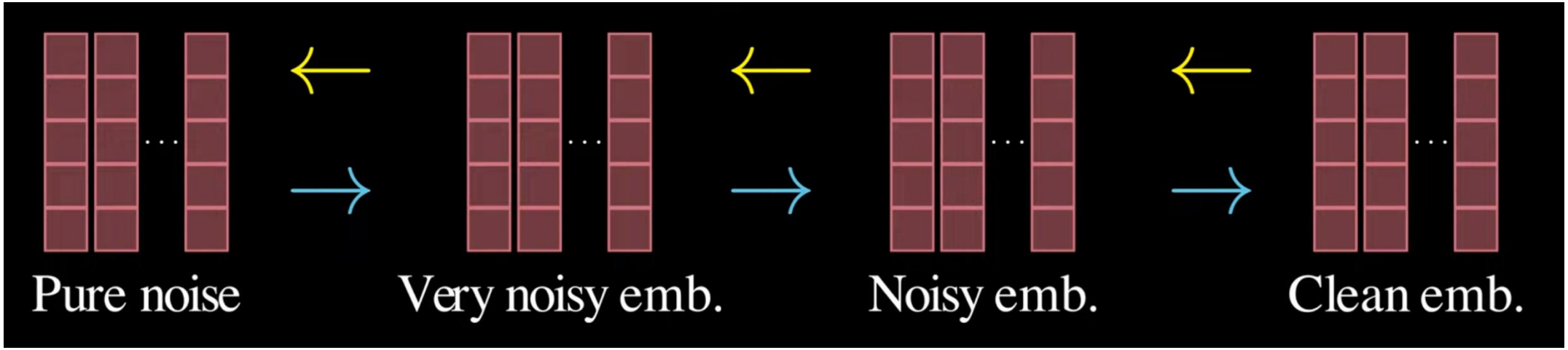


Word-Level Diffusion

https://www.youtube.com/watch?v=8BTOoc0yDVA

Sentence-Level Diffusion





https://www.youtube.com/watch?v=8BTOoc0yDVA

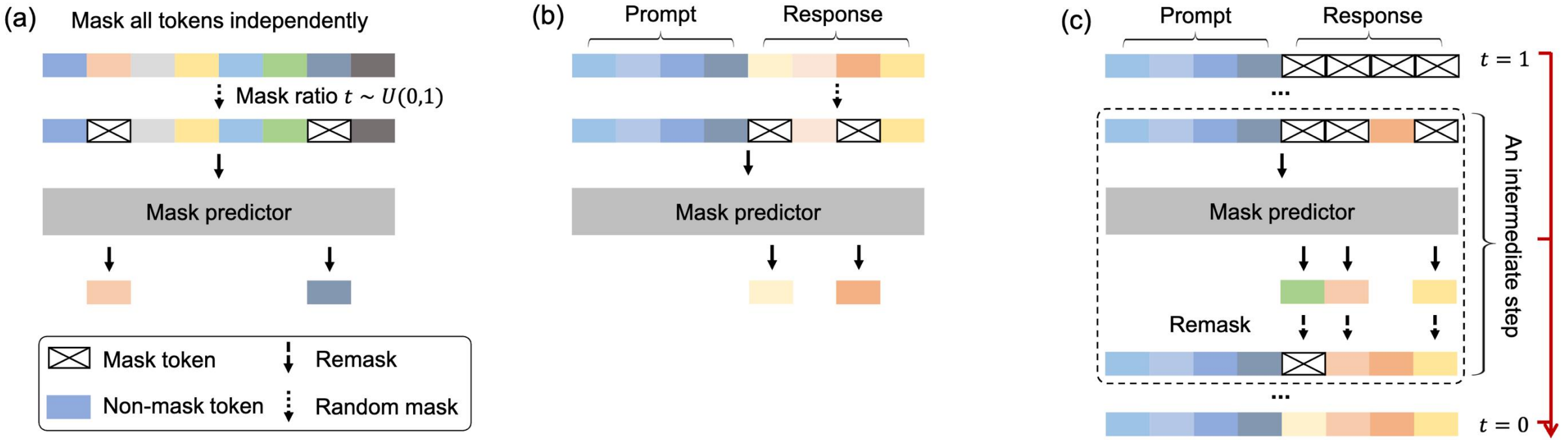
Recap: Mask-Predict

[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]
The 0.9	departure 0.2	departure 0.3	the 0.1	French 0.4	combat 0.2	completed 0.2	completed 0.2	on 0.1	20 0.6	November 0.7
The	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	[Mask]	20	November
	departure 0.2	of 0.9	French 0.8	combat 0.7	troops 0.8	troops 0.1	completed 0.2	on 0.9		
The	[Mask]	of	French	combat	troops	[Mask]	[Mask]	on	20	November
	withdrawal 0.9					was 0.8	completed 0.9			
The	withdrawal	of	French	combat	troops	was	completed	on	20	November

Large Language Diffusion Models

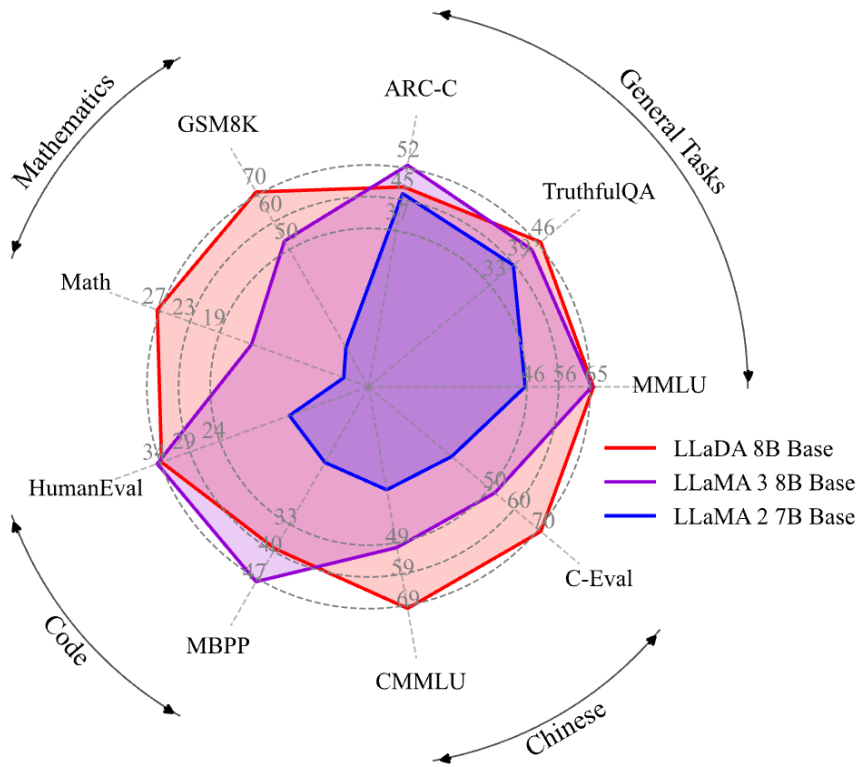
Shen Nie[1,2,3*†] **Fengqi Zhu**[1,2,3*†] **Zebin You**[1,2,3†] **Xiaolu Zhang**[4‡] **Jingyang Ou**[1,2,3]
Jun Hu[4‡] **Jun Zhou**[4] **Yankai Lin**[1,2,3‡] **Ji-Rong Wen**[1,2,3] **Chongxuan Li**[1,2,3‡§]
[1] Gaoling School of Artificial Intelligence, Renmin University of China
[2] Beijing Key Laboratory of Research on Large Models and Intelligent Governance
[3] Engineering Research Center of Next-Generation Intelligent Search and Recommendation, MOE
[4] Ant Group
{nieshen,fengqizhu,chongxuanli}@ruc.edu.cn

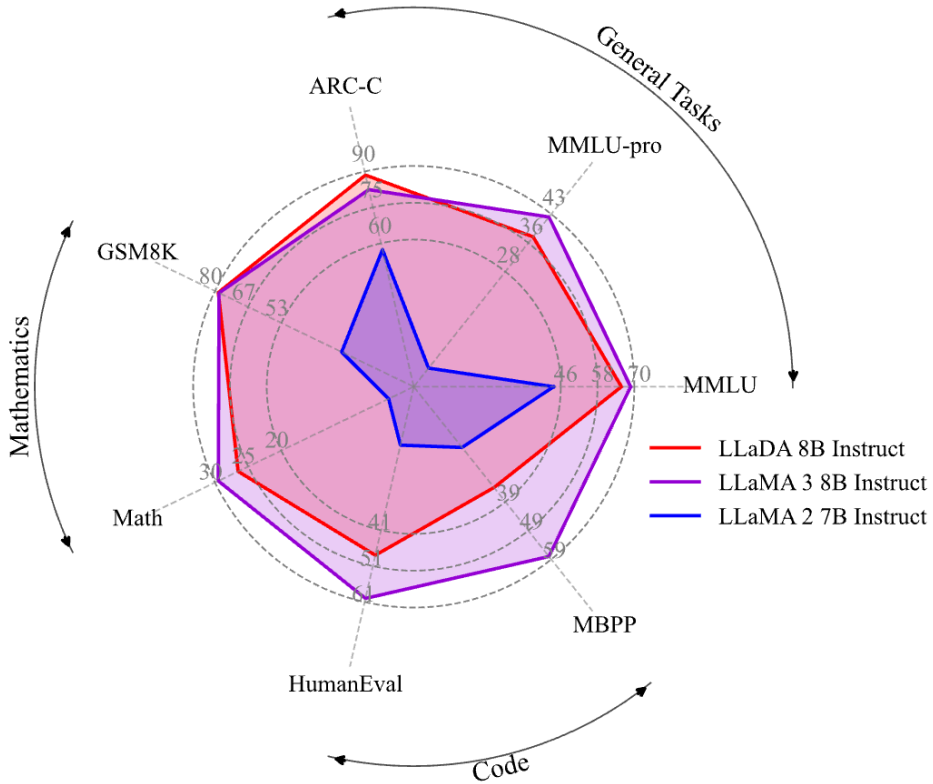
Masked Diffusion Language Models (LLaDA)



Large Language Diffusion Models, 2025

Masked Diffusion Language Models (LLaDA)





Large Language Diffusion Models, 2025

Insertion Transformer: Flexible Sequence Generation via Insertion Operations

Mitchell Stern [1 2] **William Chan** [1] **Jamie Kiros** [1] **Jakob Uszkoreit** [1]

Insertion Operation

- Generate text by inserting words

This is a book

<slot 0>	This	<slot 1>	is	<slot 2>	a	<slot 3>	book	<slot 4>

Insert “boring” at <slot 3>

This is a boring book

<slot 0>	This	<slot 1>	is	<slot 2>	a	<slot 3>	boring	<slot 4>	book	<slot 5>

Insert “very” at <slot 3>

This is a very boring book

Example

t	Canvas	Insertion
0	[]	(ate, 0)
1	[ate]	(together, 1)
2	[ate, together]	(friends, 0)
3	[friends, ate, together]	(three, 0)
4	[three, friends, ate, together]	(lunch, 3)
5	[three, friends, ate, lunch, together]	(⟨EOS⟩, 5)

Example (Parallel Version)

t	Canvas	Insertions
0	[]	(ate, 0)
1	[ate]	(friends, 0), (together, 1)
2	[friends, ate, together]	(three, 0), (lunch, 2)
3	[three, friends, ate, lunch, together]	($\langle$EOS$\rangle$, 5)

Advantages

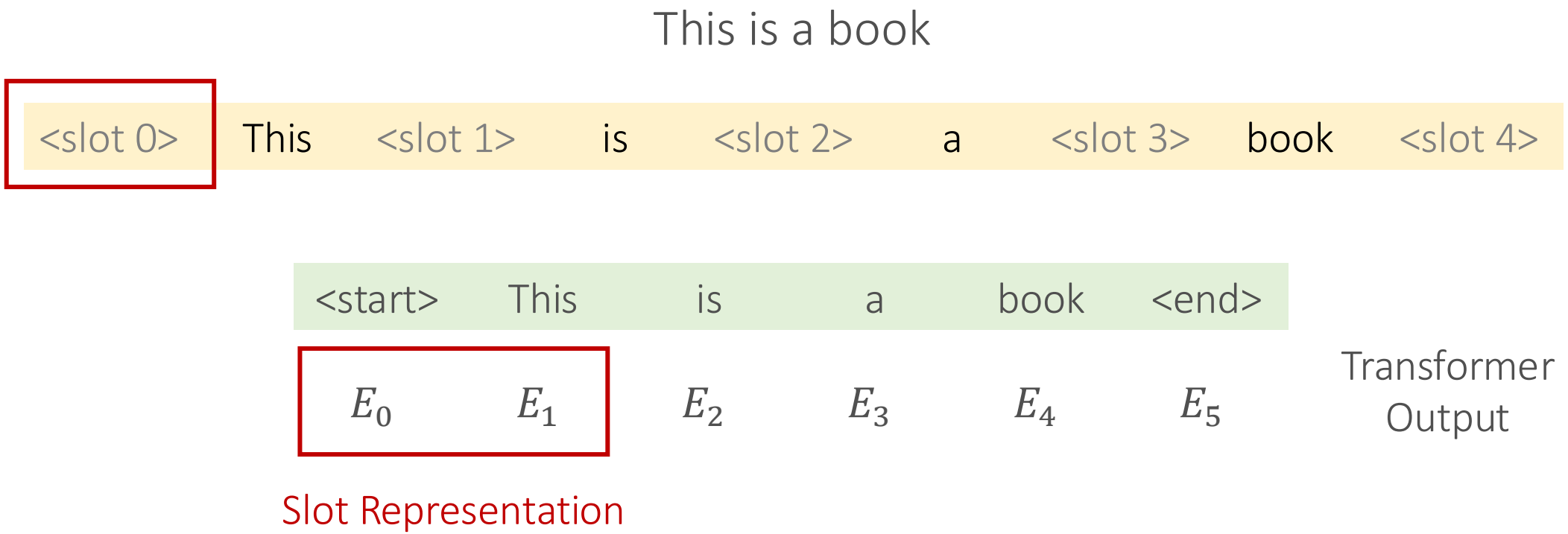
- More similar to human writing
- Don’t need to predict the output length in advance
  - Dynamically decide when to stop

There is a bird on the tree.

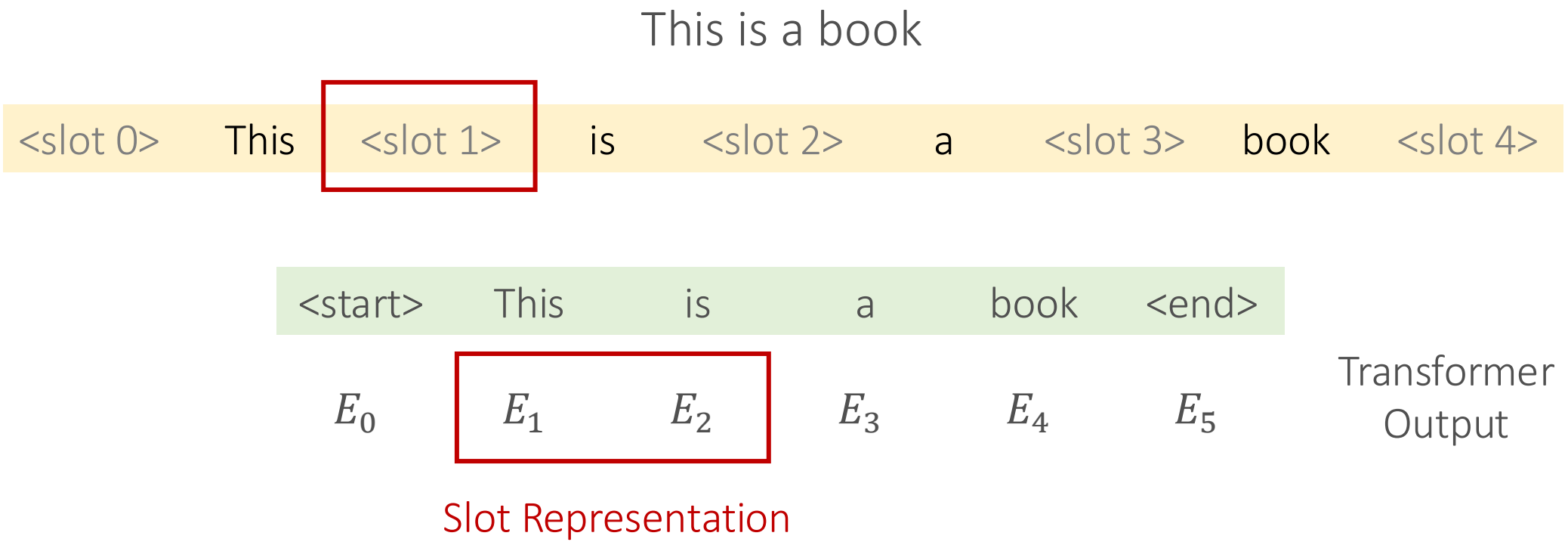
There is a red bird on the green tree.

There is a red bird with a big beak on the green tree.

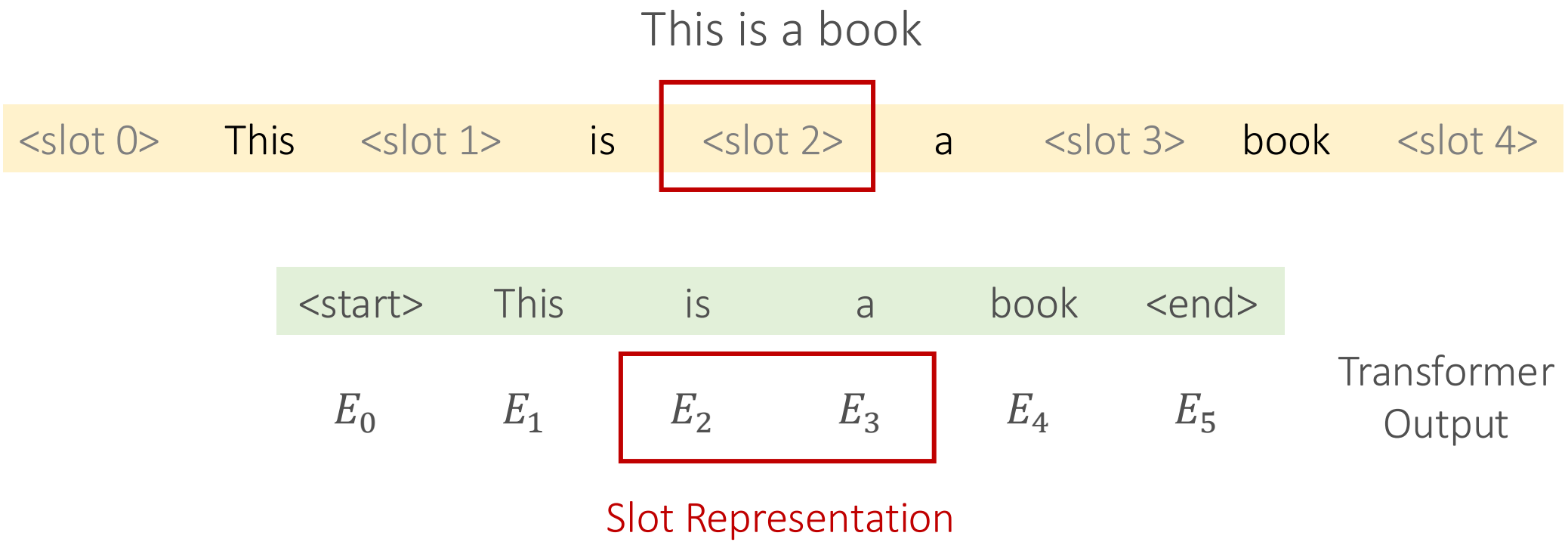
Insertion Transformer



Insertion Transformer



Insertion Transformer



Predict words based on slot representations

$$P = W^{\top}\text{concat}(E_i, E_{i+1})$$

Training Loss

A B C D E F G H I J K L M N O

<S0> A <S1> C <S2> D <S3> I <S4> M <S5>

$$Loss^{(0)} = -\log p(EOS)$$

$$Loss^{(1)} = -\log p(B)$$

$$Loss^{(2)} = -\log p(EOS)$$

$$Loss^{(3)} = \text{Avg}(-\log p(E), -\log p(F), -\log p(G), -\log p(H))$$

$$Loss^{(4)} = \text{Avg}(-\log p(J), -\log p(K), -\log p(L))$$

$$Loss^{(5)} = \text{Avg}(-\log p(N), -\log p(O))$$

Training Loss (Balanced Binary Tree)

A B C D E F G H I J K L M N O

<S0> A <S1> C <S2> D <S3> I <S4> M <S5>

$$Loss^{(0)} = -\log p(EOS)$$

$$Loss^{(1)} = -\log p(B)$$

$$Loss^{(2)} = -\log p(EOS)$$

$$Loss^{(3)} = \text{Avg}(-\log p(E), -\log p(F), -\log p(G), -\log p(H))$$

$$Loss^{(4)} = \text{Avg}(-\log p(J), -\log p(K), -\log p(L))$$

$$Loss^{(5)} = \text{Avg}(-\log p(N), -\log p(O))$$

Training Loss (Balanced Binary Tree)

Results

Loss	Termination	BLEU (+EOS)	BLEU (+EOS) +Distillation	BLEU (+EOS) +Distillation, +Parallel
Left-to-Right	Sequence	20.92 (20.92)	23.29 (23.36)	-
Binary Tree ($\tau = 0.5$)	Slot	20.35 (21.39)	24.49 (25.55)	25.33 (25.70)
Binary Tree ($\tau = 1.0$)	Slot	21.02 (22.37)	24.36 (25.43)	25.43 (25.76)
Binary Tree ($\tau = 2.0$)	Slot	20.52 (21.95)	24.59 (25.80)	25.33 (25.80)
Uniform	Sequence	19.34 (22.64)	22.75 (25.45)	-
Uniform	Slot	18.26 (22.16)	22.39 (25.58)	24.31 (24.91)

Examples

Input: Everyone has the Internet, an iPad and eBooks.

Output: Jeder hat das Internet, ein iPad und eBooks.

Greedy decode (uniform loss):

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

(continued)

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Jeder_ hat_ das_ Internet_ , _ ein_ i Pad _ und_ eB oo ks_ ._

Examples

Input: But on the other side of the state, that is not the impression many people have of their former governor.

Output: Aber auf der anderen Seite des Staates ist das nicht der Eindruck, den viele von ihrem ehemaligen Gouverneur haben.

Parallel decode (binary tree loss):

Aber_ auf_ der_ anderen_ Seite_ des_ Staates_ ist_ das_ nicht_ der_ Eindruck_ , _ den_ viele_ von_ ihrem_ ehemaligen_ Gouverneur _ haben_ ._

Aber_ auf_ der_ anderen_ Seite_ des_ Staates_ ist_ das_ nicht_ der_ Eindruck_ , _ den_ viele_ von_ ihrem_ ehemaligen_ Gouverneur _ haben_ ._

Aber_ auf_ der_ anderen_ Seite_ des_ Staates_ ist_ das_ nicht_ der_ Eindruck_ , _ den_ viele_ von_ ihrem_ ehemaligen_ Gouverneur _ haben_ ._

Aber_ auf_ der_ anderen_ Seite_ des_ Staates_ ist_ das_ nicht_ der_ Eindruck_ , _ den_ viele_ von_ ihrem_ ehemaligen_ Gouverneur _ haben_ ._

Aber_ auf_ der_ anderen_ Seite_ des_ Staates_ ist_ das_ nicht_ der_ Eindruck_ , _ den_ viele_ von_ ihrem_ ehemaligen_ Gouverneur _ haben_ ._