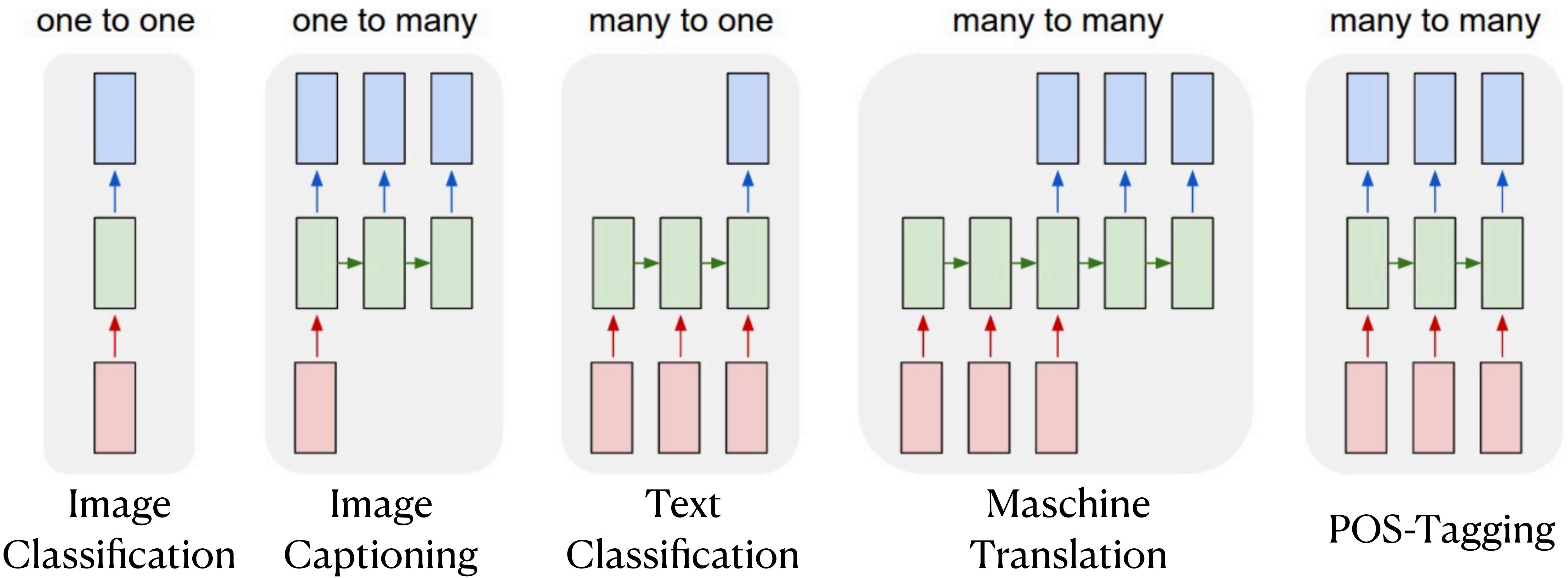


# Нейронные сети в NLP

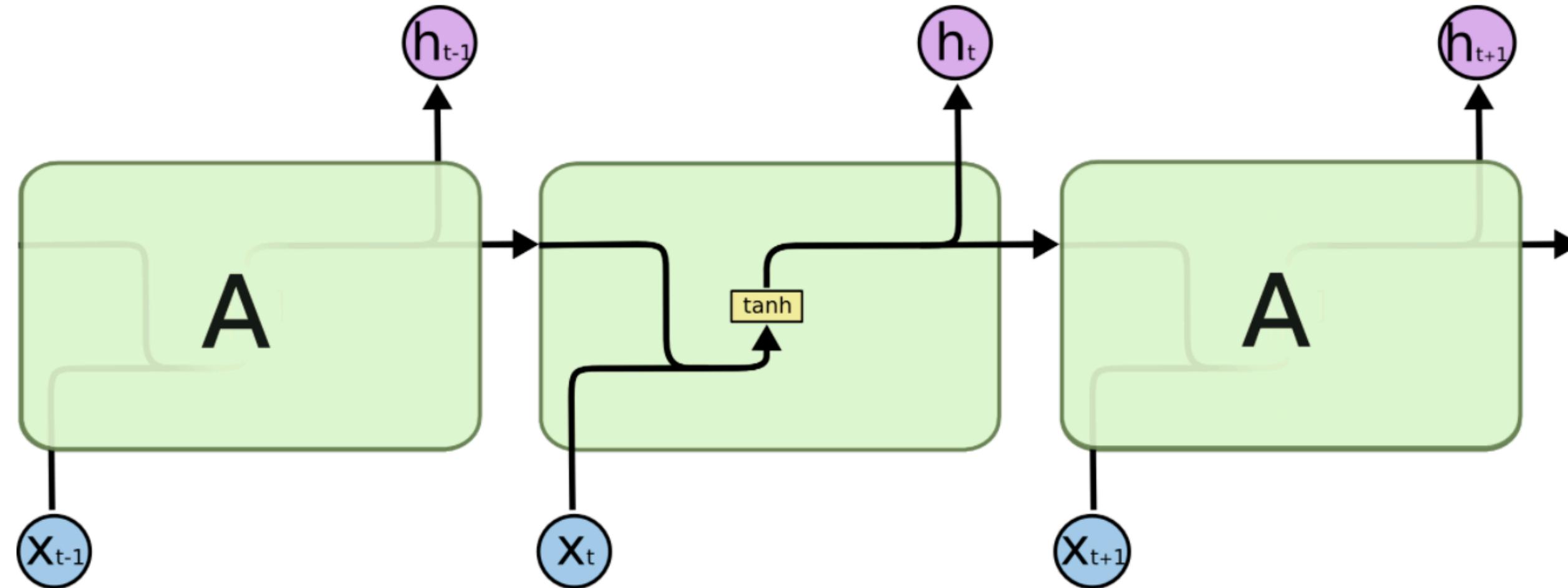
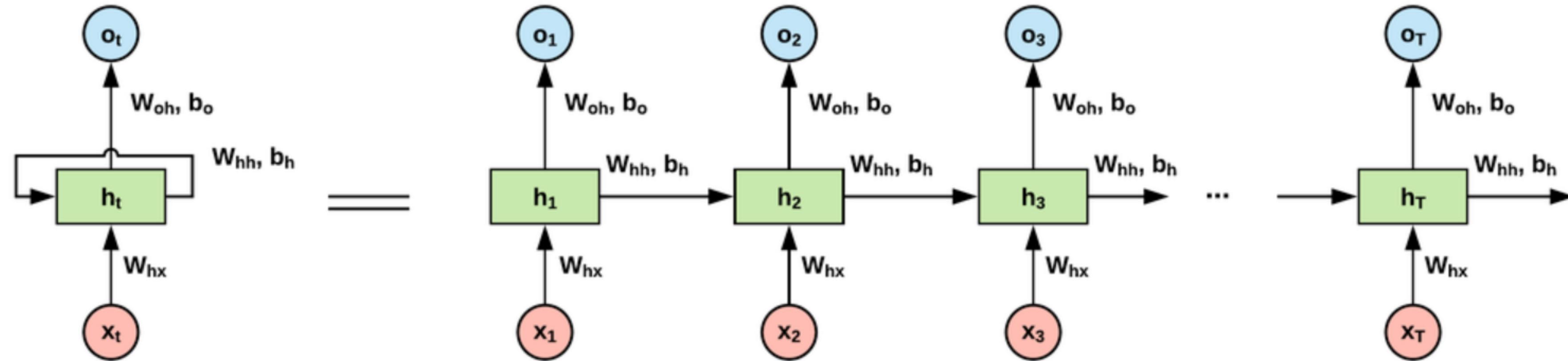
RNN Attention Transformers

Лектор: Алтухов Никита Александрович  
Аналитик данных Сбербанк

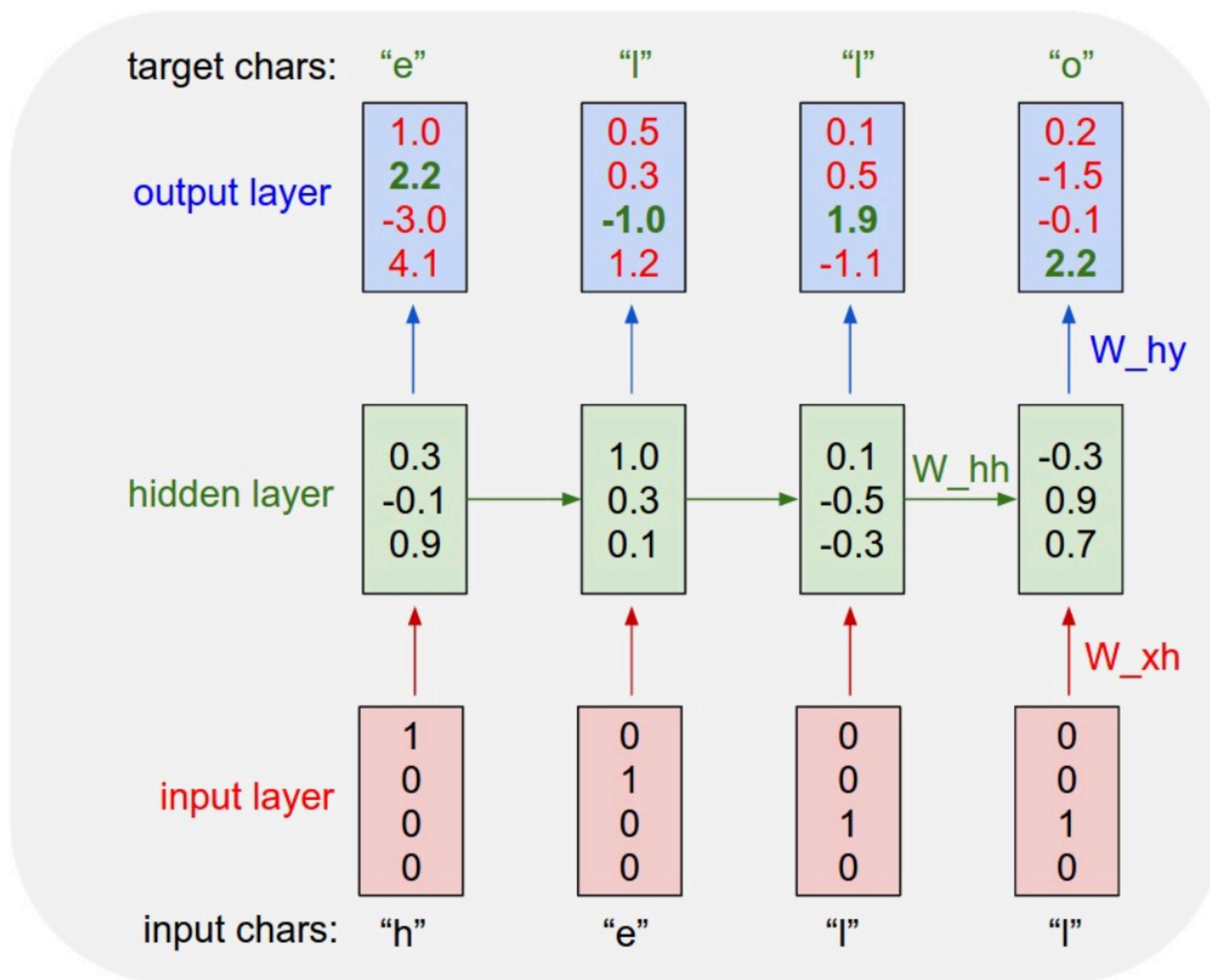
# Sequence модели



# RNN



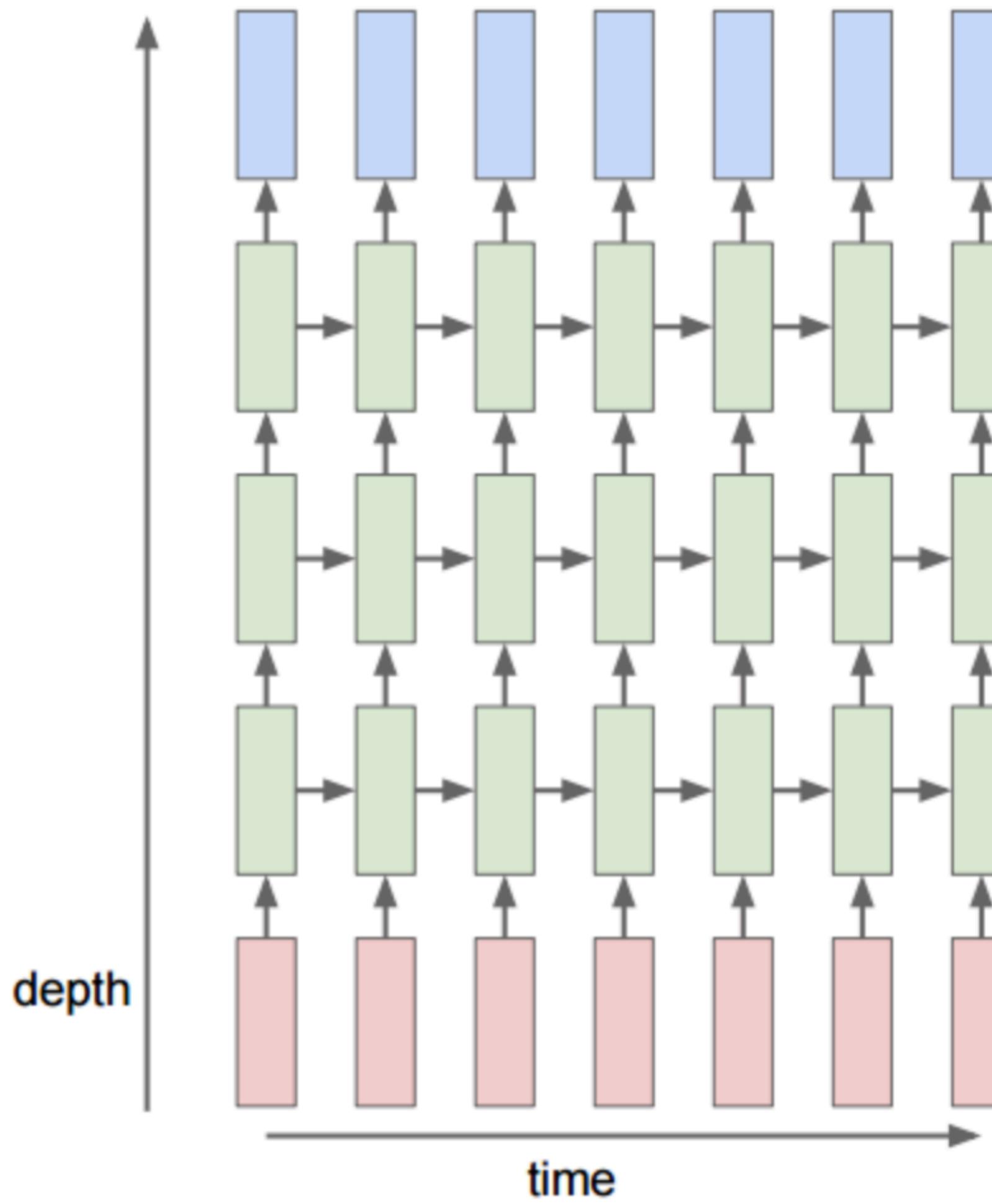
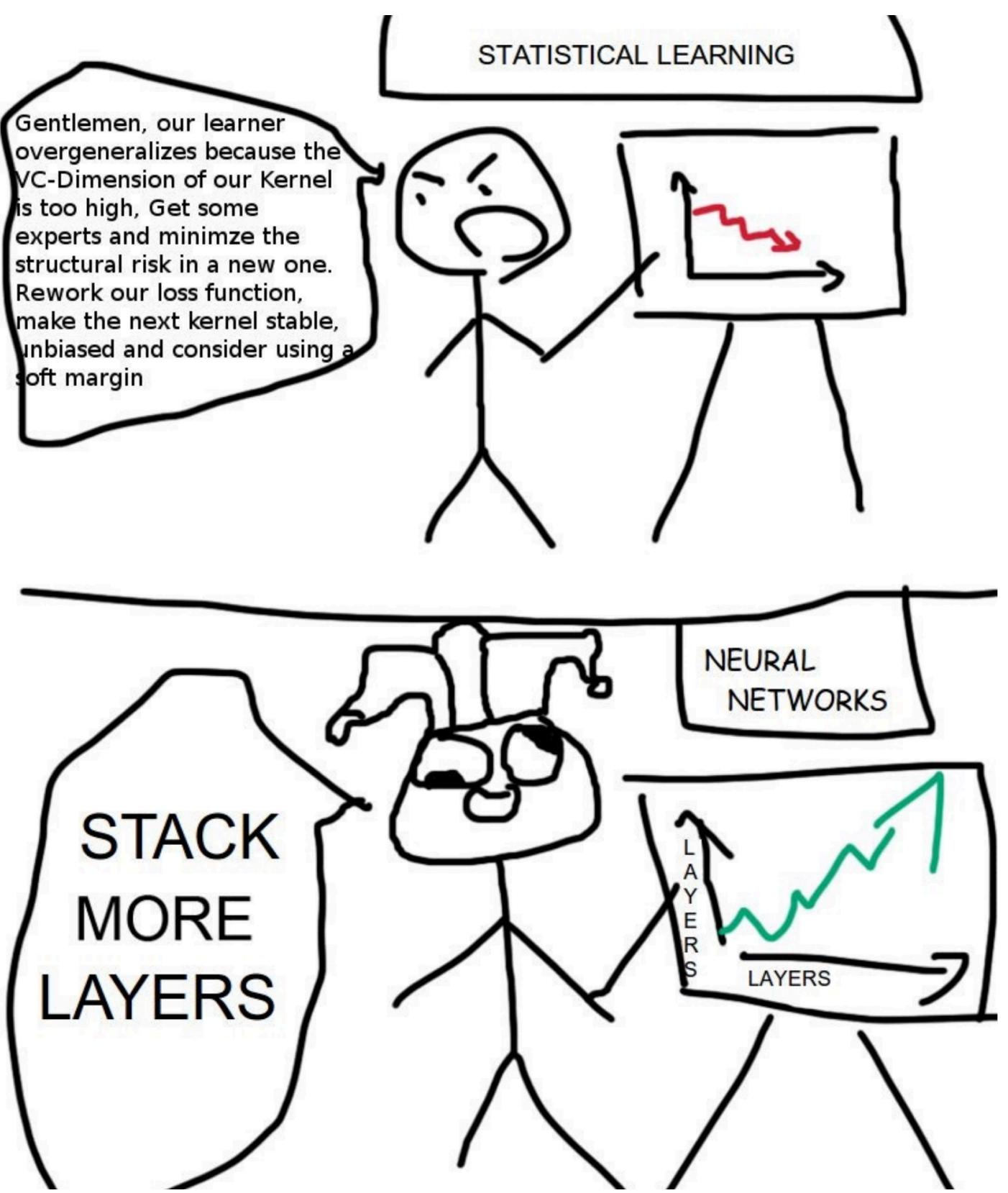
# RNN



- Посимвольная генерация текста
- На обучении выход смещен относительно входа на один символ. Пытаемся предсказать следующий символ
- На предикте выход подается на вход следующего шага
- Добавляем в словарь [BOS] и [EOS]

```
class RNN:
    # ...
    def step(self, x):
        # update the hidden state
        self.h = np.tanh(np.dot(self.W_hh, self.h) + np.dot(self.W_xh, x))
        # compute the output vector
        y = np.dot(self.W_hy, self.h)
        return y
```

# Многослойная RNN



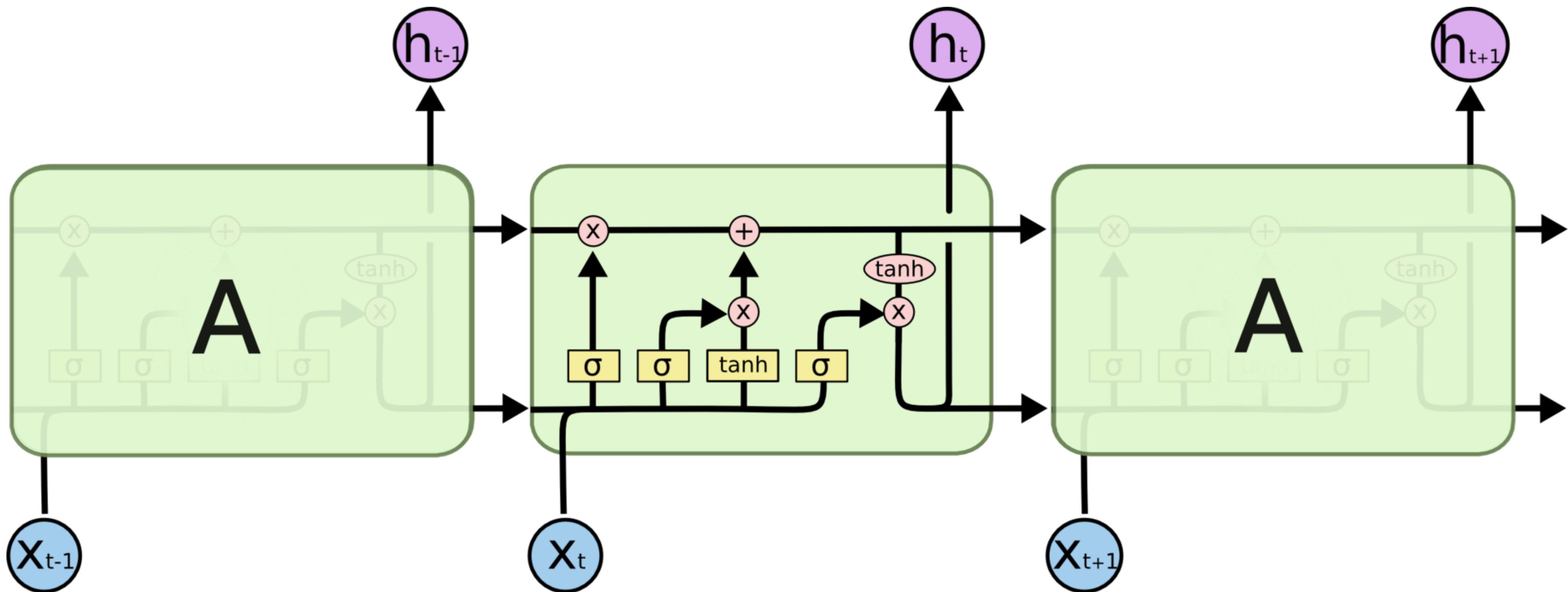
# Проблемы Vanila RNN

- При обучении, когда градиент течет обратно по сети, он умножается много раз (длина последовательности) на матрицу весов. Что делает его либо огромным (если веса  $> 1$ ), либо очень маленьким (веса  $< 1$ ).
- Если используется  $\tanh$ , то градиент превращается в 0.
- Если используется  $\text{relu}$ , то градиент взрывается.

# Проблемы Vanila RNN

- При обучении, когда градиент течет обратно по сети, он умножается много раз (длина последовательности) на матрицу весов. Что делает его либо огромным (если веса  $> 1$ ), либо очень маленьким (веса  $< 1$ ).
- Если используется  $\tanh$ , то градиент превращается в 0.
- Если используется  $\text{relu}$ , то градиент взрывается.

# LSTM



# LSTM

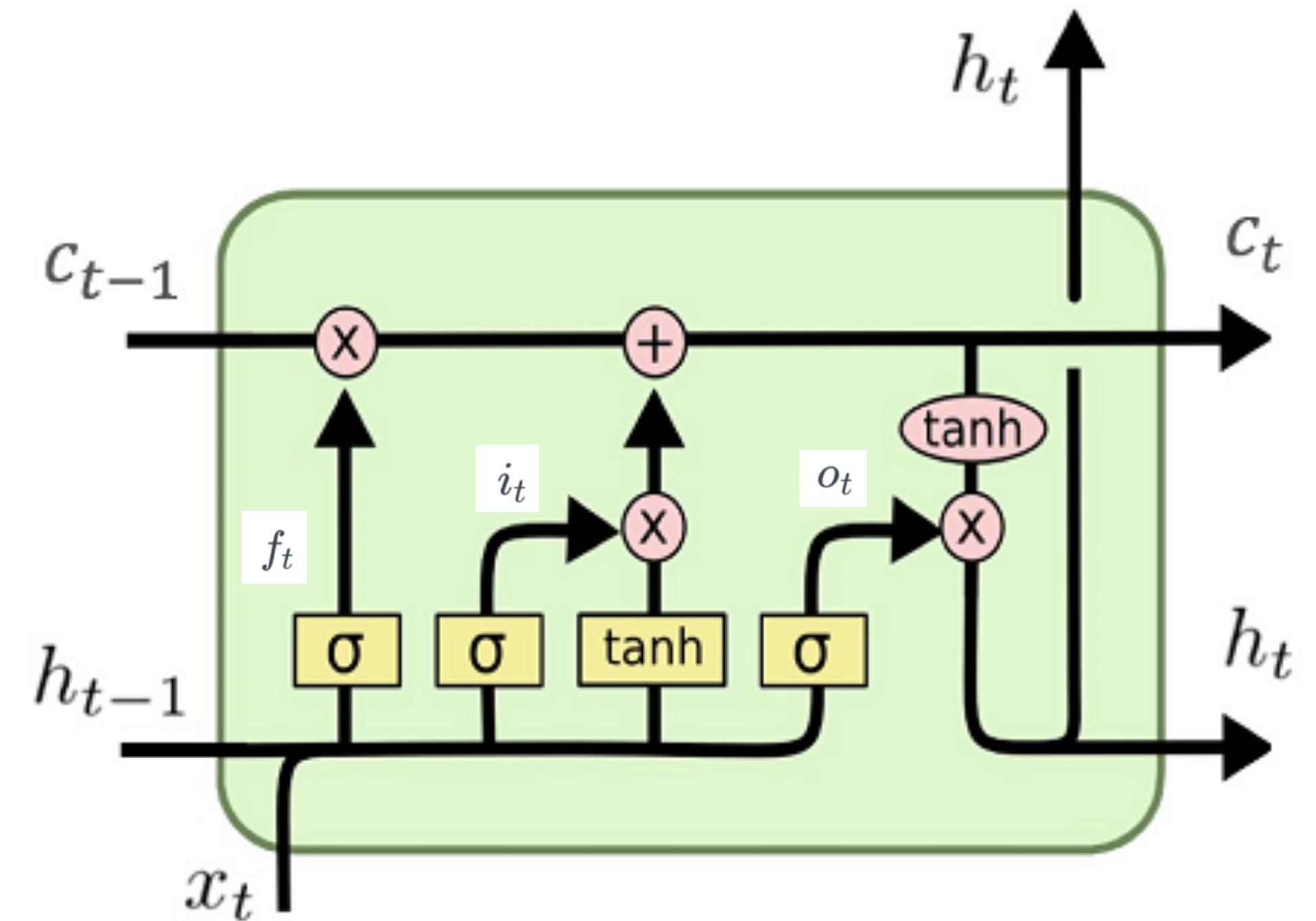
$$f_t = \sigma(W_{hf}h_{t-1} + W_{xf}x + b_f)$$

$$i_t = \sigma(W_{hi}h_{t-1} + W_{xi}x + b_i)$$

$$o_t = \sigma(W_{ho}h_{t-1} + W_{xo}x + b_o)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{gx}x + W_{gh}h_{t-1} + b_g)$$

$$h_t = o_t \odot \tanh(c_t)$$



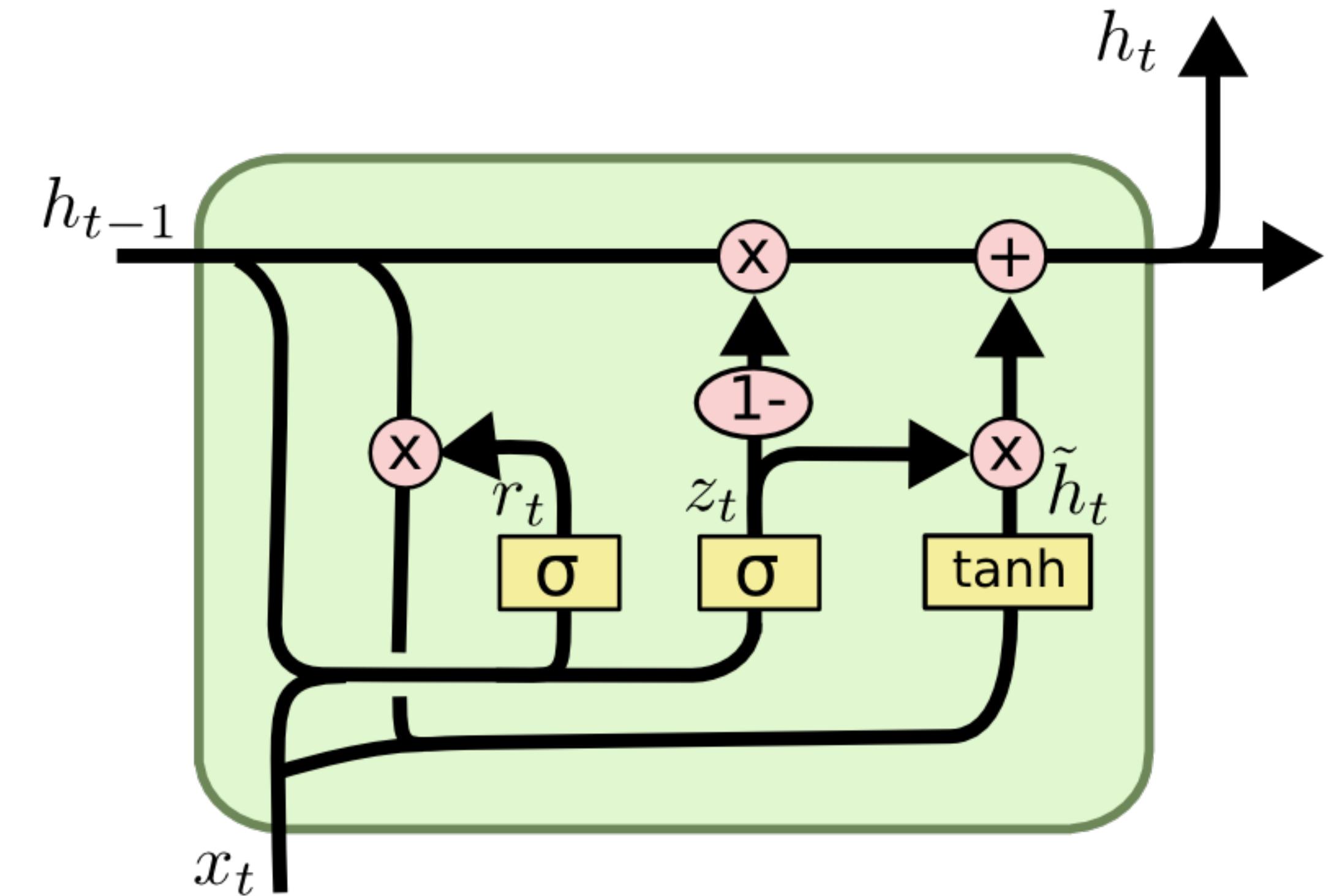
# GRU

$$z_t = \sigma (W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma (W_r \cdot [h_{t-1}, x_t])$$

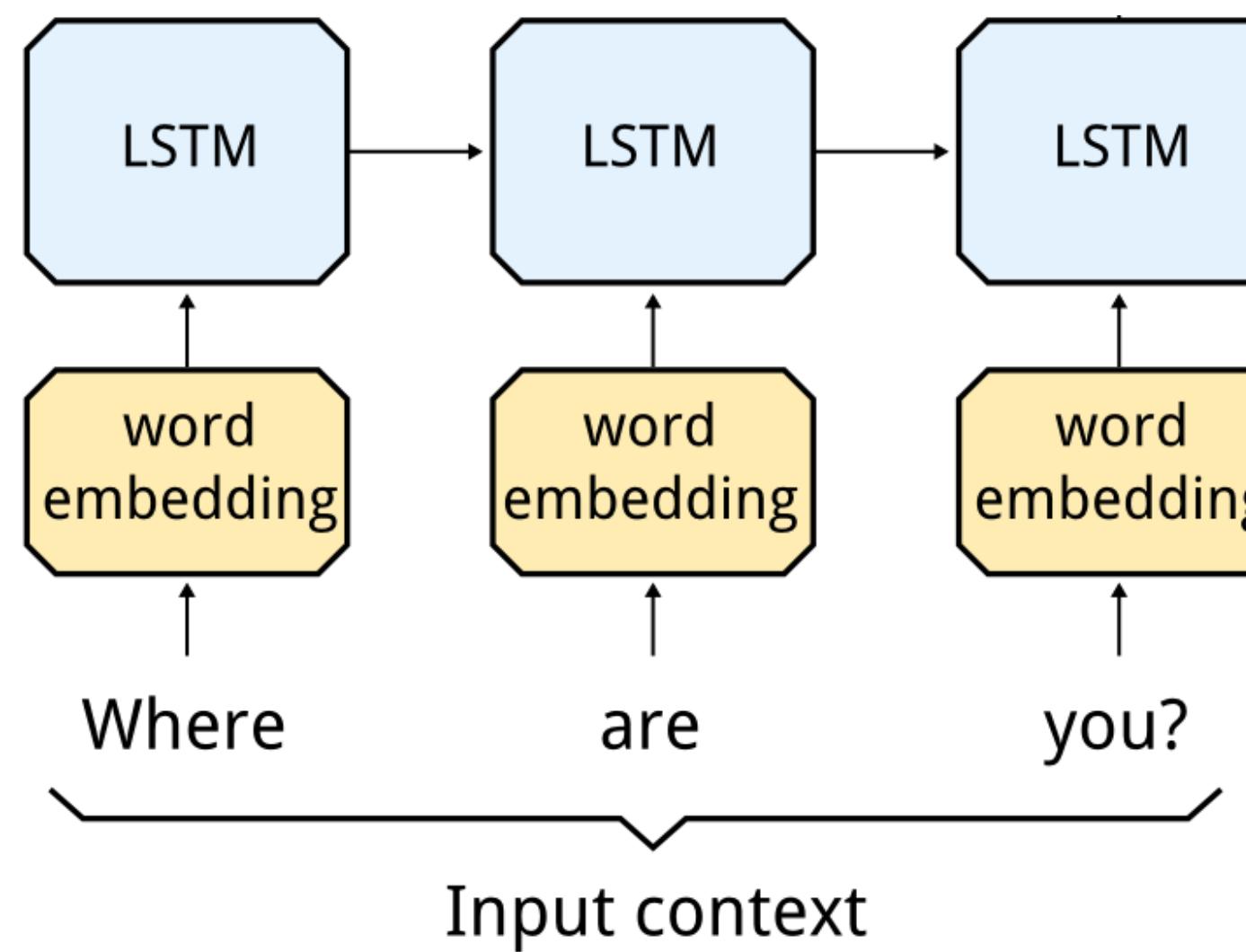
$$\tilde{h}_t = \tanh (W \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

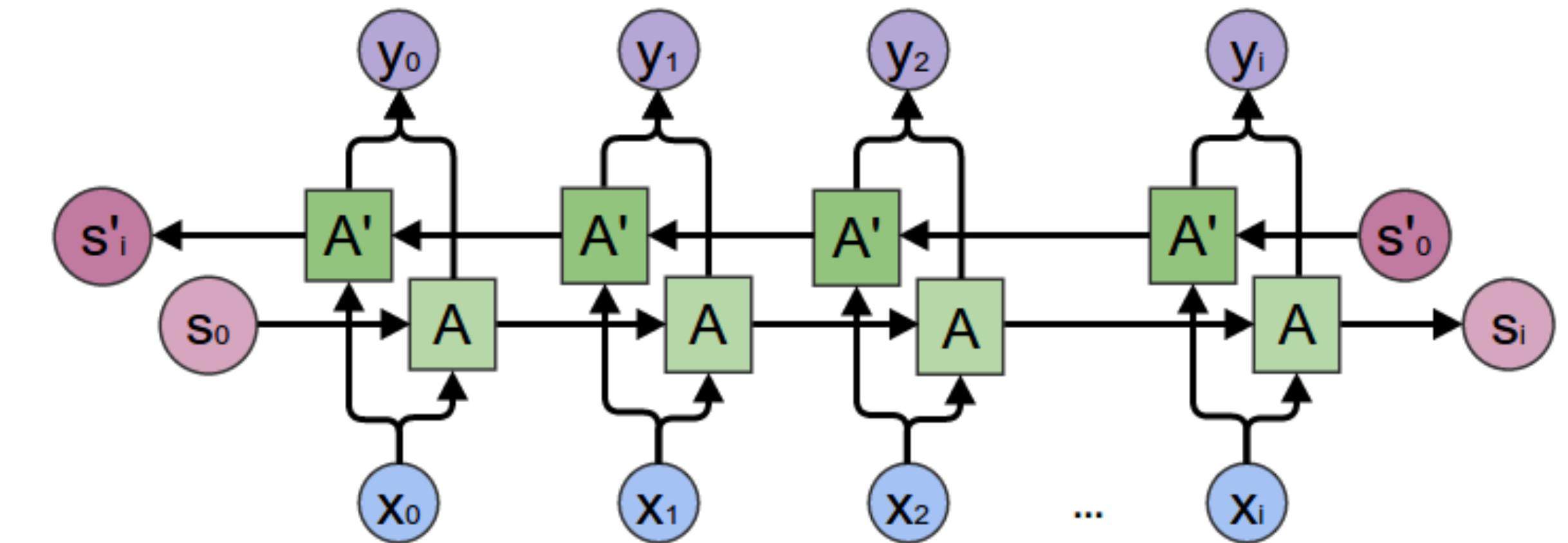


# Как еще усилить?

Embeddings (Word2Vec)



Bidirectional RNN



# Применения RNN

## Wikipedia (100Мб)

Naturalism and decision for the majority of Arab countries' capitalide was grounded by the Irish language by [[John Clair]], [[An Imperial Japanese Revolt]], associated with Guangzham's sovereignty. His generals were the powerful ruler of the Portugal in the [[Protestant Immineners]], which could be said to be directly in Cantonese Communication, which followed a ceremony and set inspired prison, training. The emperor travelled back to [[Antioch, Perth, October 25|21]] to note, the Kingdom of Costa Rica, unsuccessful fashioned the [[Thrales]], [[Cynth's Dajoard]], known in western [[Scotland]], near Italy to the conquest of India with the conflict. Copyright was the succession of independence in the slop of Syrian influence that was a famous German movement based on a more popular servicious, non-doctrinal and sexual power post. Many governments recognize the military housing of the [[Civil Liberalization and Infantry Resolution 265 National Party in Hungary]], that is sympathetic to be to the [[Punjab Resolution]] (PJS)[<http://www.humah.yahoo.com/guardian.cfm/7754800786d17551963s89.htm>] Official economics Adjoint for the Nazism, Montgomery was swear to advance to the resources for those Socialism's rule, was starting to signing a major tripad of aid exile.]]

# Применения RNN

## Algebraic Geometry (Latex, 16мб)

For  $\bigoplus_{n=1,\dots,m} \mathcal{L}_{m,n} = 0$ , hence we can find a closed subset  $\mathcal{H}$  in  $\mathcal{H}$  and any sets  $\mathcal{F}$  on  $X$ ,  $U$  is a closed immersion of  $S$ , then  $U \rightarrow T$  is a separated algebraic space.

*Proof.* Proof of (1). It also start we get

$$S = \text{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by  $\coprod Z \times_U U \rightarrow V$ . Consider the maps  $M$  along the set of points  $\text{Sch}_{fppf}$  and  $U \rightarrow U$  is the fibre category of  $S$  in  $U$  in Section ?? and the fact that any  $U$  affine, see Morphisms, Lemma ???. Hence we obtain a scheme  $S$  and any open subset  $W \subset U$  in  $\text{Sh}(G)$  such that  $\text{Spec}(R') \rightarrow S$  is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that  $f_i$  is of finite presentation over  $S$ . We claim that  $\mathcal{O}_{X,x}$  is a scheme where  $x, x' \in S'$  such that  $\mathcal{O}_{X,x'} \rightarrow \mathcal{O}'_{X',x'}$  is separated. By Algebra, Lemma ?? we can define a map of complexes  $\text{GL}_{S'}(x'/S'')$  and we win.  $\square$

To prove study we see that  $\mathcal{F}|_U$  is a covering of  $X'$ , and  $\mathcal{T}_i$  is an object of  $\mathcal{F}_{X/S}$  for  $i > 0$  and  $\mathcal{F}_p$  exists and let  $\mathcal{F}_i$  be a presheaf of  $\mathcal{O}_X$ -modules on  $C$  as a  $\mathcal{F}$ -module. In particular  $\mathcal{F} = U/\mathcal{F}$  we have to show that

$$\widetilde{\mathcal{M}}^\bullet = \mathcal{I}^\bullet \otimes_{\text{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1} \mathcal{F}$$

is a unique morphism of algebraic stacks. Note that

$$\text{Arrows} = (\text{Sch}/S)_{fppf}^{\text{opp}}, (\text{Sch}/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \hookrightarrow (U, \text{Spec}(A))$$

is an open subset of  $X$ . Thus  $U$  is affine. This is a continuous map of  $X$  is the inverse, the groupoid scheme  $S$ .

*Proof.* See discussion of sheaves of sets.  $\square$

The result for prove any open covering follows from the less of Example ???. It may replace  $S$  by  $X_{\text{spaces},\text{étale}}$  which gives an open subspace of  $X$  and  $T$  equal to  $S_{\text{Zar}}$ , see Descent, Lemma ???. Namely, by Lemma ?? we see that  $R$  is geometrically regular over  $S$ .

**Lemma 0.1.** Assume (3) and (3) by the construction in the description.

Suppose  $X = \lim |X|$  (by the formal open covering  $X$  and a single map  $\underline{\text{Proj}}_X(\mathcal{A}) = \text{Spec}(B)$  over  $U$  compatible with the complex

$$\text{Set}(\mathcal{A}) = \Gamma(X, \mathcal{O}_{X,\mathcal{O}_X}).$$

When in this case of to show that  $\mathcal{Q} \rightarrow \mathcal{C}_{Z/X}$  is stable under the following result in the second conditions of (1), and (3). This finishes the proof. By Definition ?? (without element is when the closed subschemes are catenary. If  $T$  is surjective we may assume that  $T$  is connected with residue fields of  $S$ . Moreover there exists a closed subspace  $Z \subset X$  of  $X$  where  $U$  in  $X'$  is proper (some defining as a closed subset of the uniqueness it suffices to check the fact that the following theorem

(1)  $f$  is locally of finite type. Since  $S = \text{Spec}(R)$  and  $Y = \text{Spec}(R)$ .

*Proof.* This is form all sheaves of sheaves on  $X$ . But given a scheme  $U$  and a surjective étale morphism  $U \rightarrow X$ . Let  $U \cap U = \coprod_{i=1,\dots,n} U_i$  be the scheme  $X$  over  $S$  at the schemes  $X_i \rightarrow X$  and  $U = \lim_i X_i$ .  $\square$

The following lemma surjective restrocomposes of this implies that  $\mathcal{F}_{x_0} = \mathcal{F}_{x_0} = \mathcal{F}_{x,\dots,x}$ .

**Lemma 0.2.** Let  $X$  be a locally Noetherian scheme over  $S$ ,  $E = \mathcal{F}_{X/S}$ . Set  $\mathcal{I} = \mathcal{J}_1 \subset \mathcal{I}'_n$ . Since  $\mathcal{I}'_n \subset \mathcal{I}^n$  are nonzero over  $i_0 \leq p$  is a subset of  $\mathcal{J}_{n,0} \circ \overline{A}_2$  works.

**Lemma 0.3.** In Situation ???. Hence we may assume  $q' = 0$ .

*Proof.* We will use the property we see that  $p$  is the next functor (??). On the other hand, by Lemma ?? we see that

$$D(\mathcal{O}_{X'}) = \mathcal{O}_X(D)$$

where  $K$  is an  $F$ -algebra where  $\delta_{n+1}$  is a scheme over  $S$ .  $\square$

Sampled (fake) algebraic geometry. [Here's the actual pdf.](#)

# Применения RNN

```
/*
 * Increment the size file of the new incorrect UI_FILTER group information
 * of the size generatively.
 */
static int indicate_policy(void)
{
    int error;
    if (fd == MARN_EPT) {
        /*
         * The kernel blank will coeld it to userspace.
         */
        if (ss->segment < mem_total)
            unblock_graph_and_set_blocked();
        else
            ret = 1;
        goto bail;
    }
    segaddr = in_SB(in.addr);
    selector = seg / 16;
    setup_works = true;
    for (i = 0; i < blocks; i++) {
        seq = buf[i++];
        bpf = bd->bd.next + i * search;
        if (fd) {
            current = blocked;
        }
    }
    rw->name = "Getjbbregs";
    bprm_self_clearl(&iv->version);
    regs->new = blocks[(BPF_STATS << info->historidac)] | PFMR_CLOBATHINC_SECONDS << 12;
    return segtable;
}
```

Linux Source (474Мб C code)

# Применения RNN

100 iterations:

tyntd-iafhatawiaoahrdemot lytdws e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e  
plia tkldrgd t o idoe ns,smtt h ne etie h,hregtrs nigtike,aoaenns lng

300 iterations:

"Tmont thithey" fomesscerliund  
Keushey. Thom here  
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwyl fil on aseterlome  
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."

500 iterations:

we counter. He stutn co des. His stanted out one ofler that concossions and was  
to gearang reay Jotrets and with fre colt oft paitt thin wall. Which das stimn

700 iterations:

Aftair fall unsuch that the hall for Prince Velzonski's that me of  
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort  
how, and Gogition is so overelical and ofter.

1200 iterations:

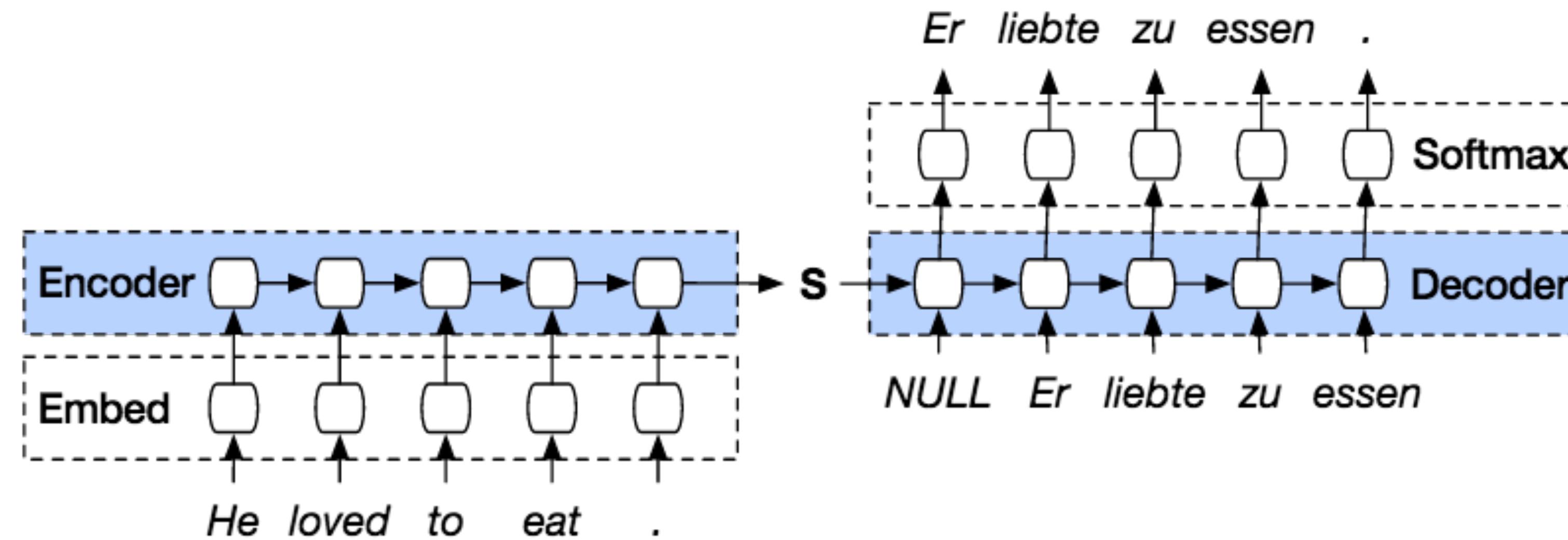
"Kite vouch!" he repeated by her  
door. "But I would be done and quarts, feeling, then, son is people...."

2000 iterations:

"Why do what that day," replied Natasha, and wishing to himself the fact the  
princess, Princess Mary was easier, fed in had oftened him.  
Pierre aking his soul came to the packs and drove up his father-in-law women.

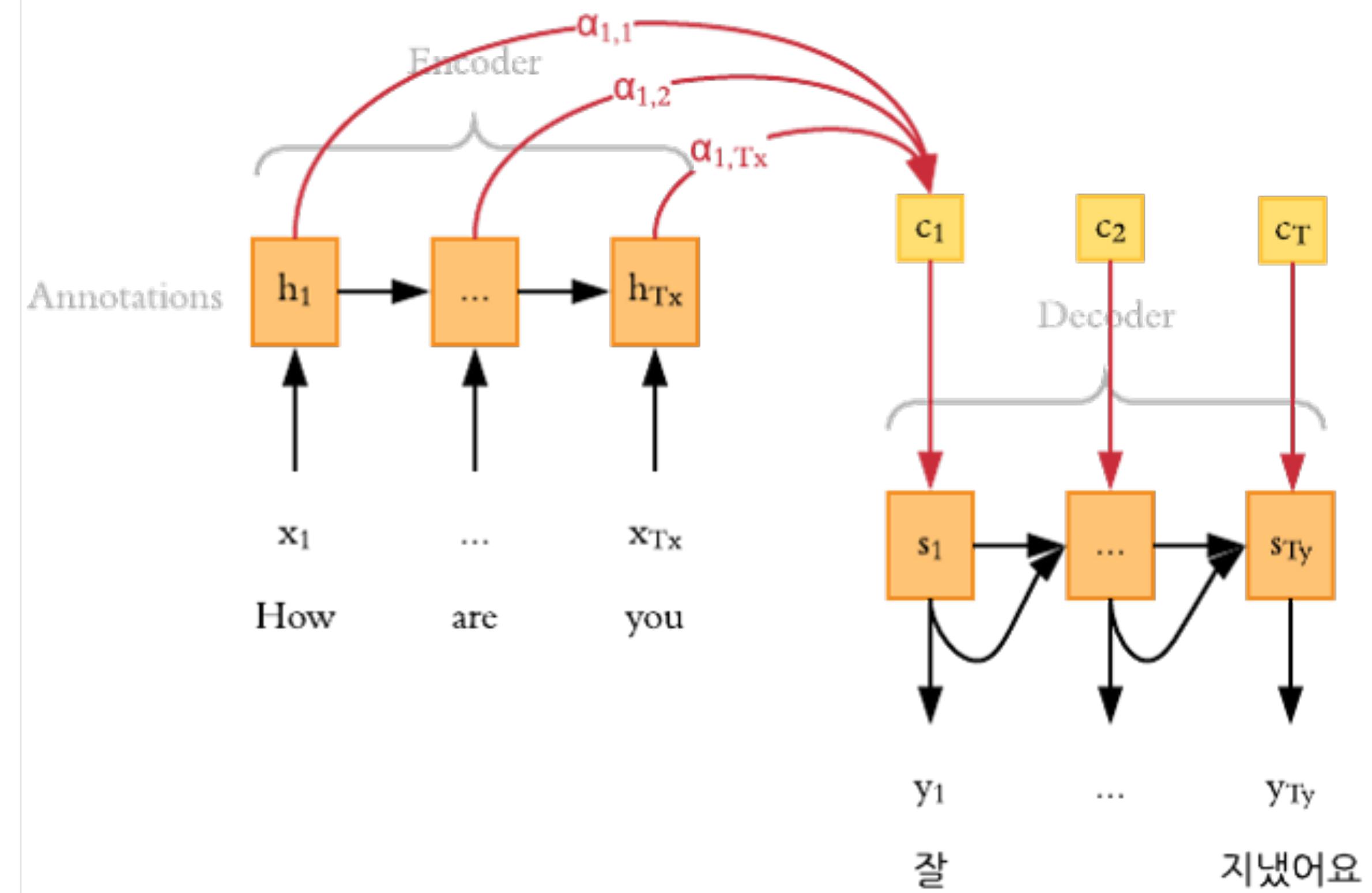
# Минусы RNN

- Не параллелизируется на GPU
- На больших последовательностях связи между началом и концом теряются
- Ограничены размерностью внутреннего состояния сети

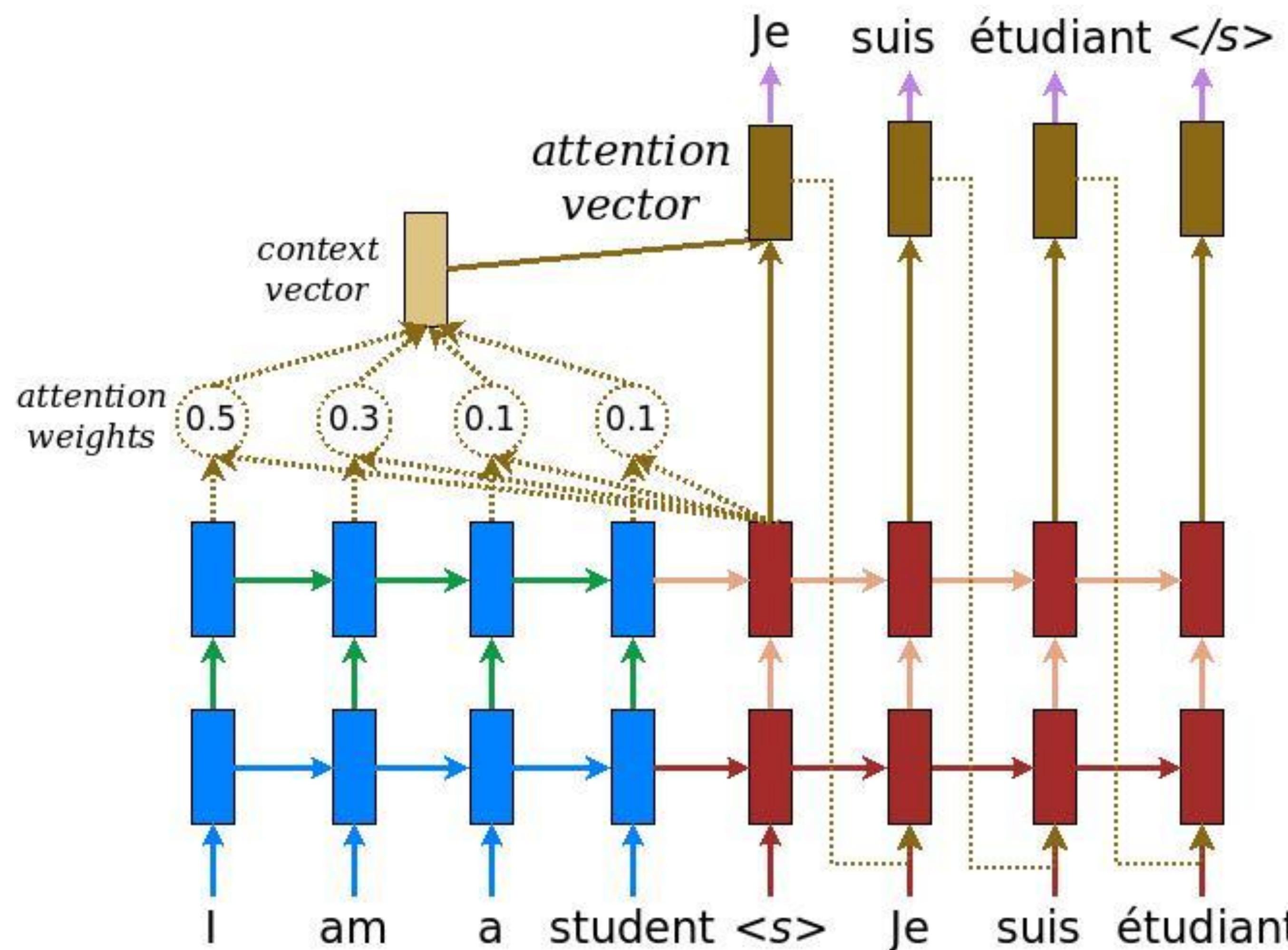


# Attention

- Архитектура, которая позволяет модели выбрать в какое место **encoder** смотреть на предсказании слова в **decoder**.
- Взвешиваем каждый выход **encoder** с помощью функции
- В простом случае вес  $([0, 1])$  - это скалярное произведение внутреннего состояния **decoder** на внутренние состояния **encoder**.

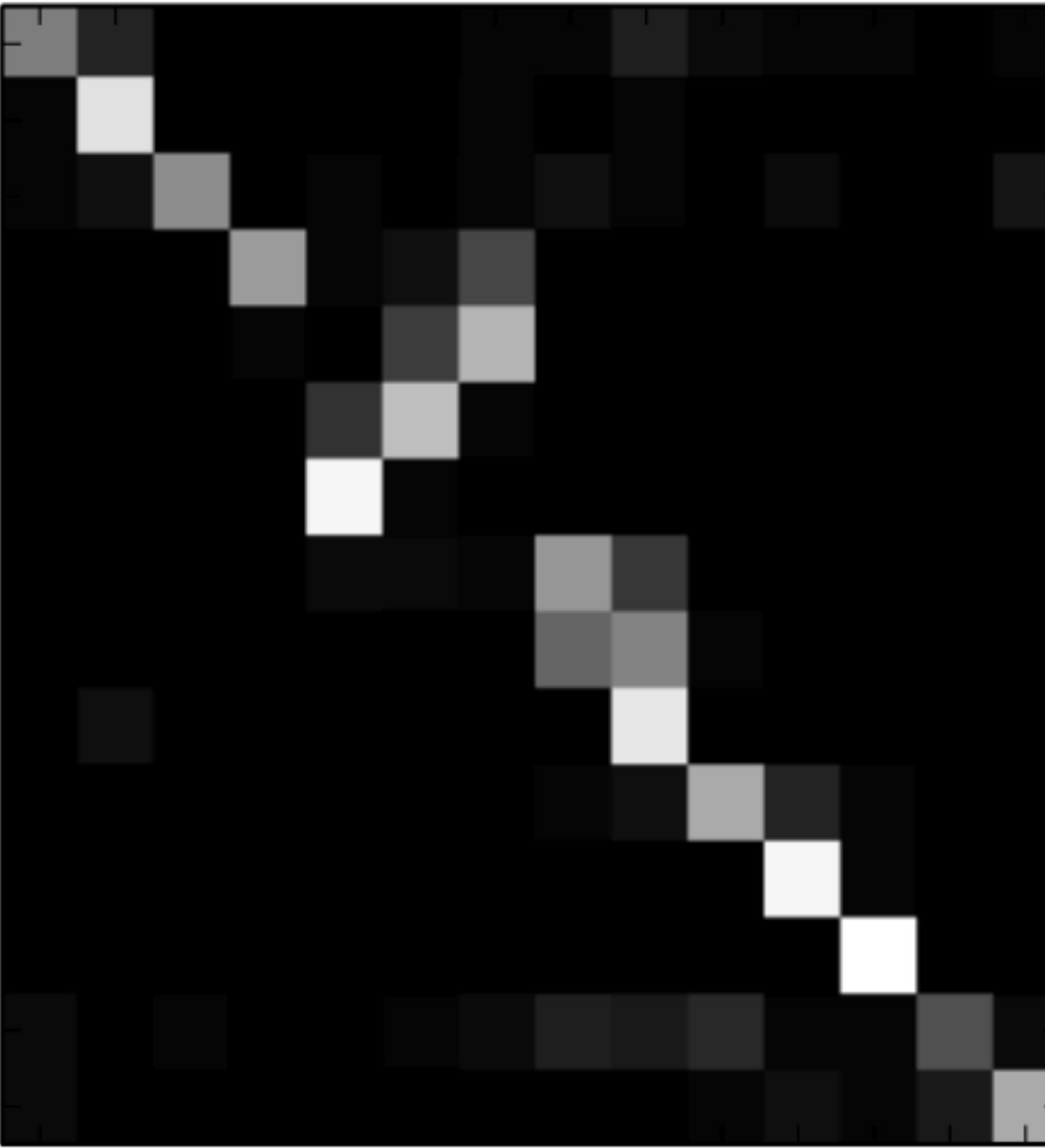


# Attention



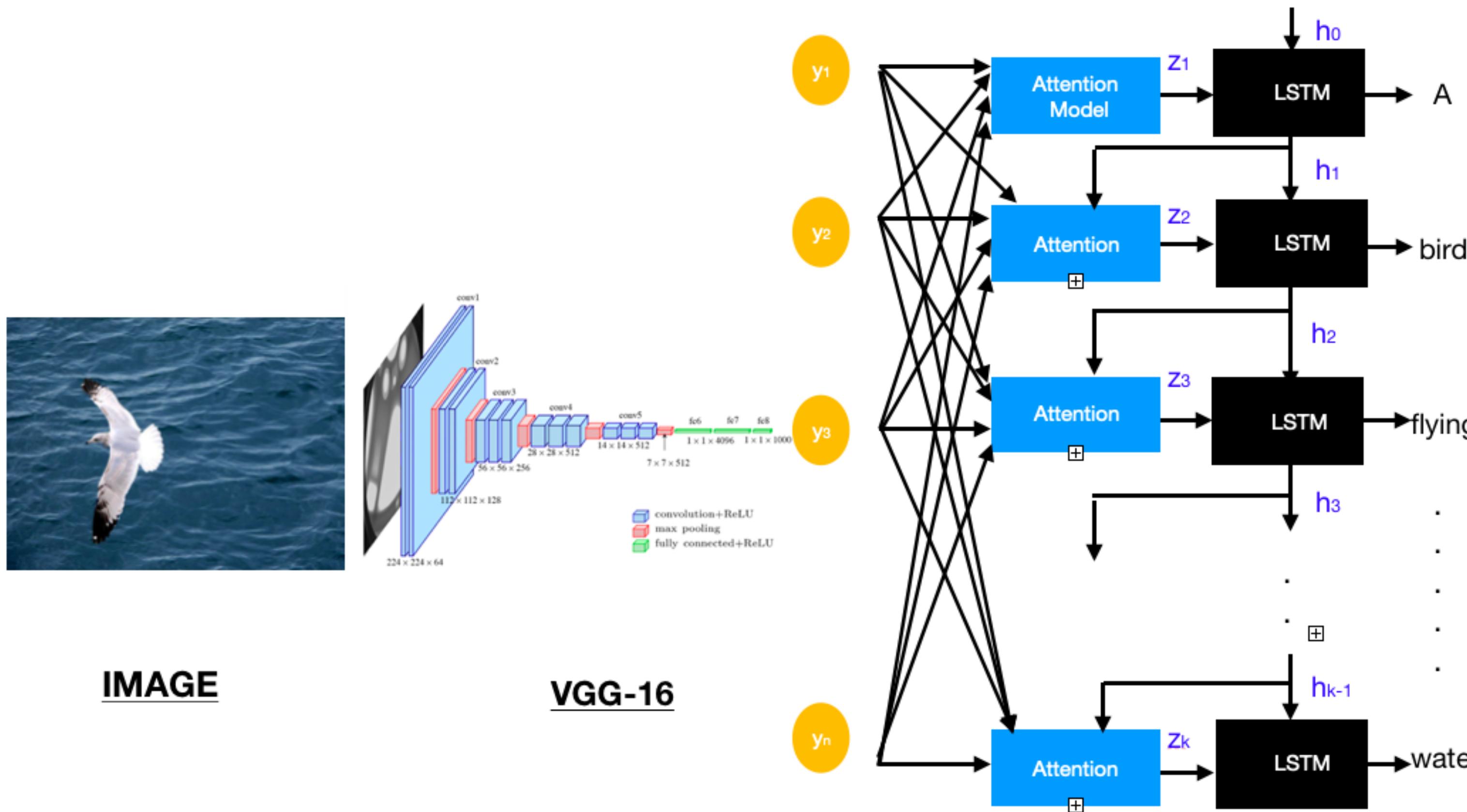
# Attention

The agreement on the European Economic Area was signed in August 1992 .  
<end>

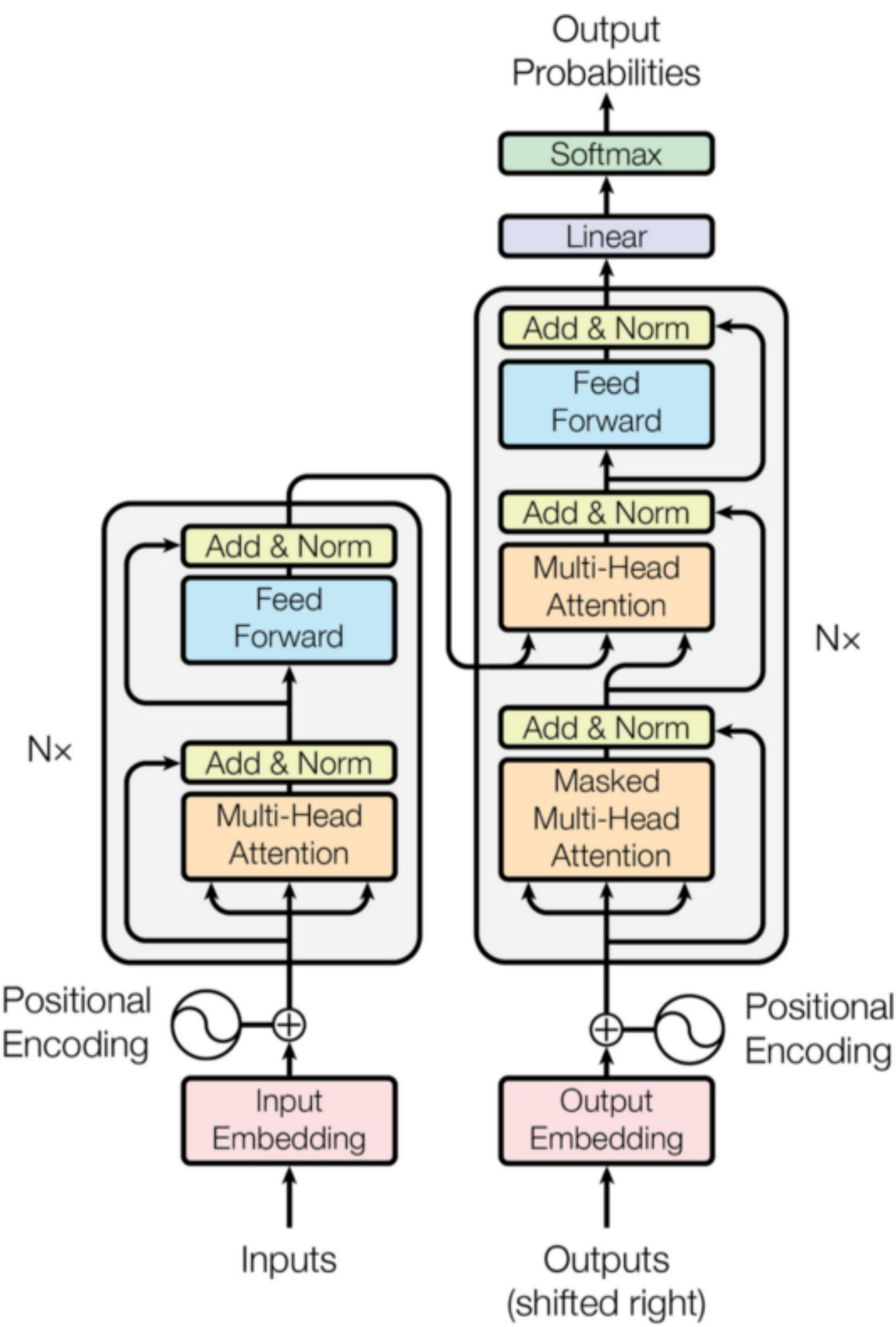


L'  
accord  
sur  
la  
zone  
économique  
européenne  
a  
été  
signé  
en  
août  
1992  
.   
<end>

# Image Captioning with Attention



# Attention is All You Need

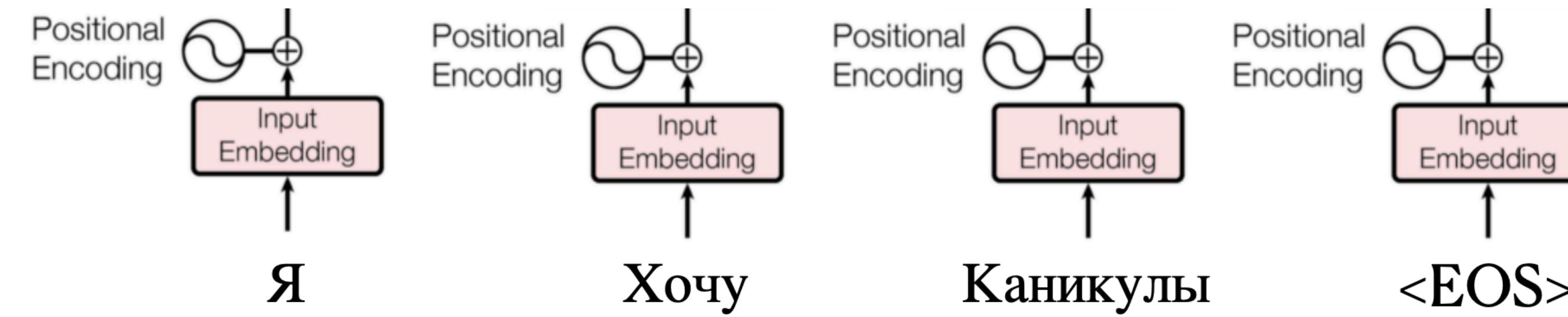


<https://habr.com/ru/post/341240/>

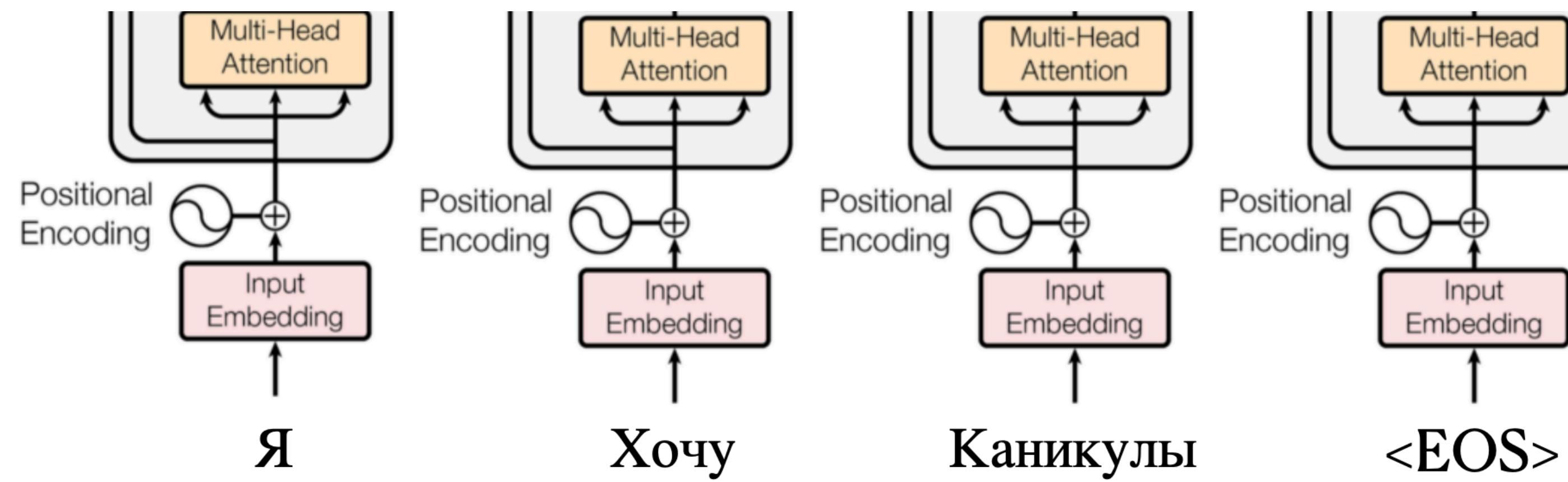
<http://jalammar.github.io/illustrated-transformer/>

# Transformer

## Encoder

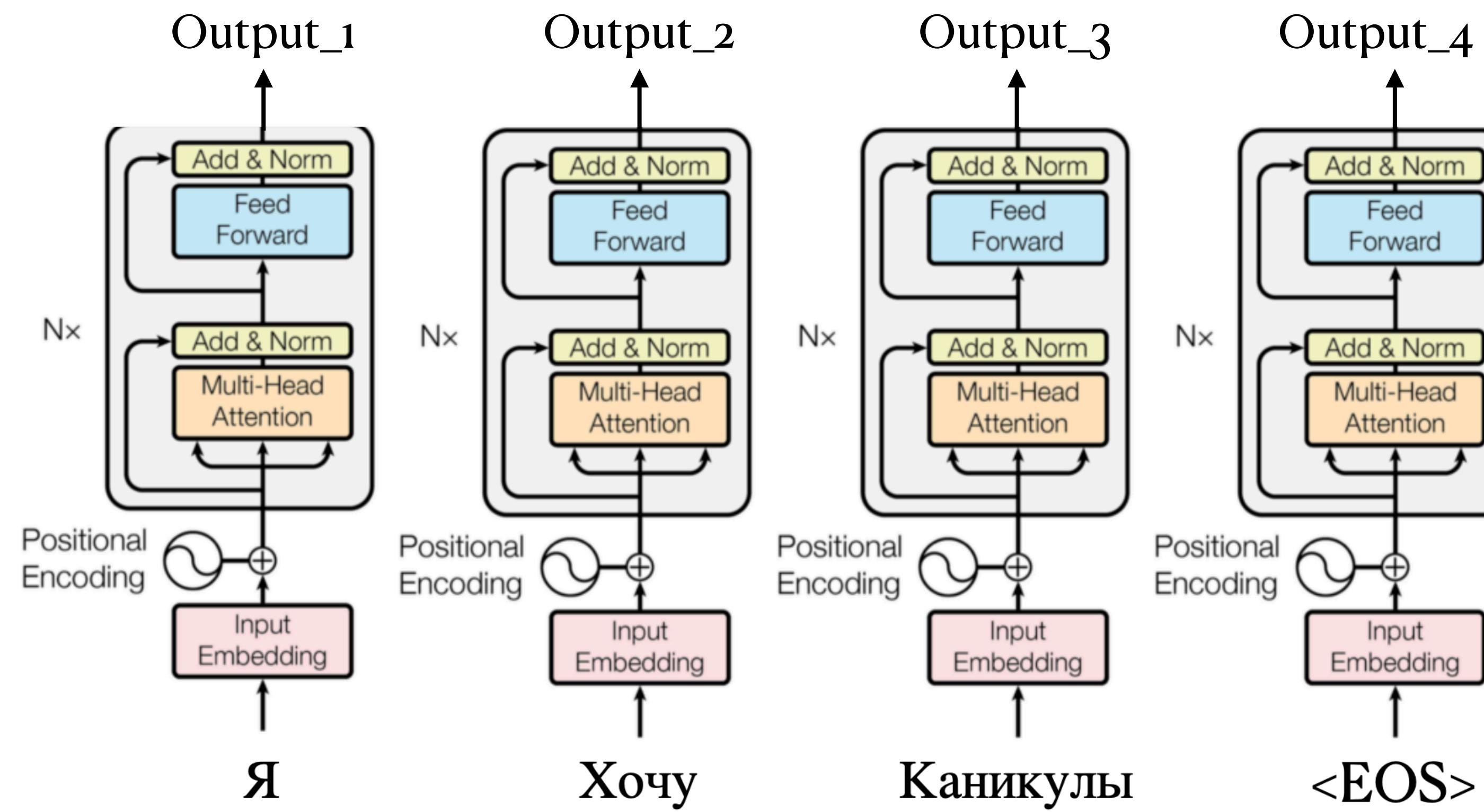


# Transformer Encoder



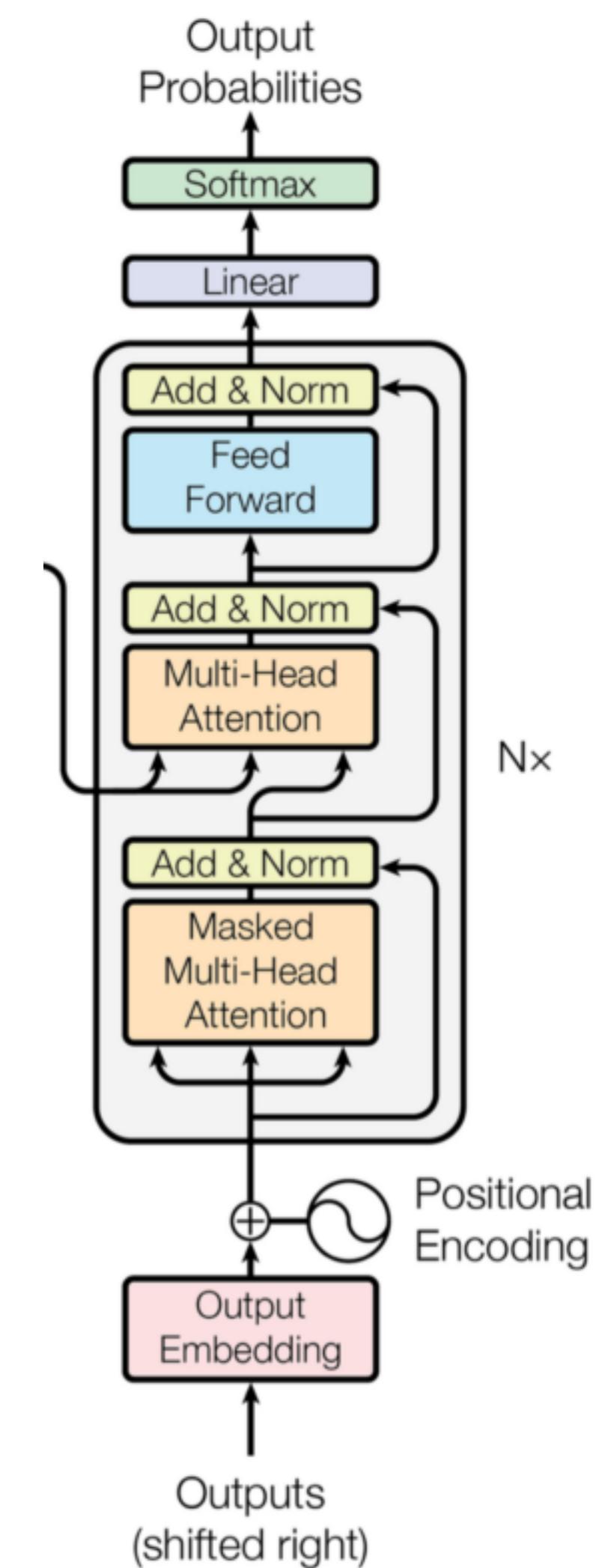
# Transformer

## Encoder

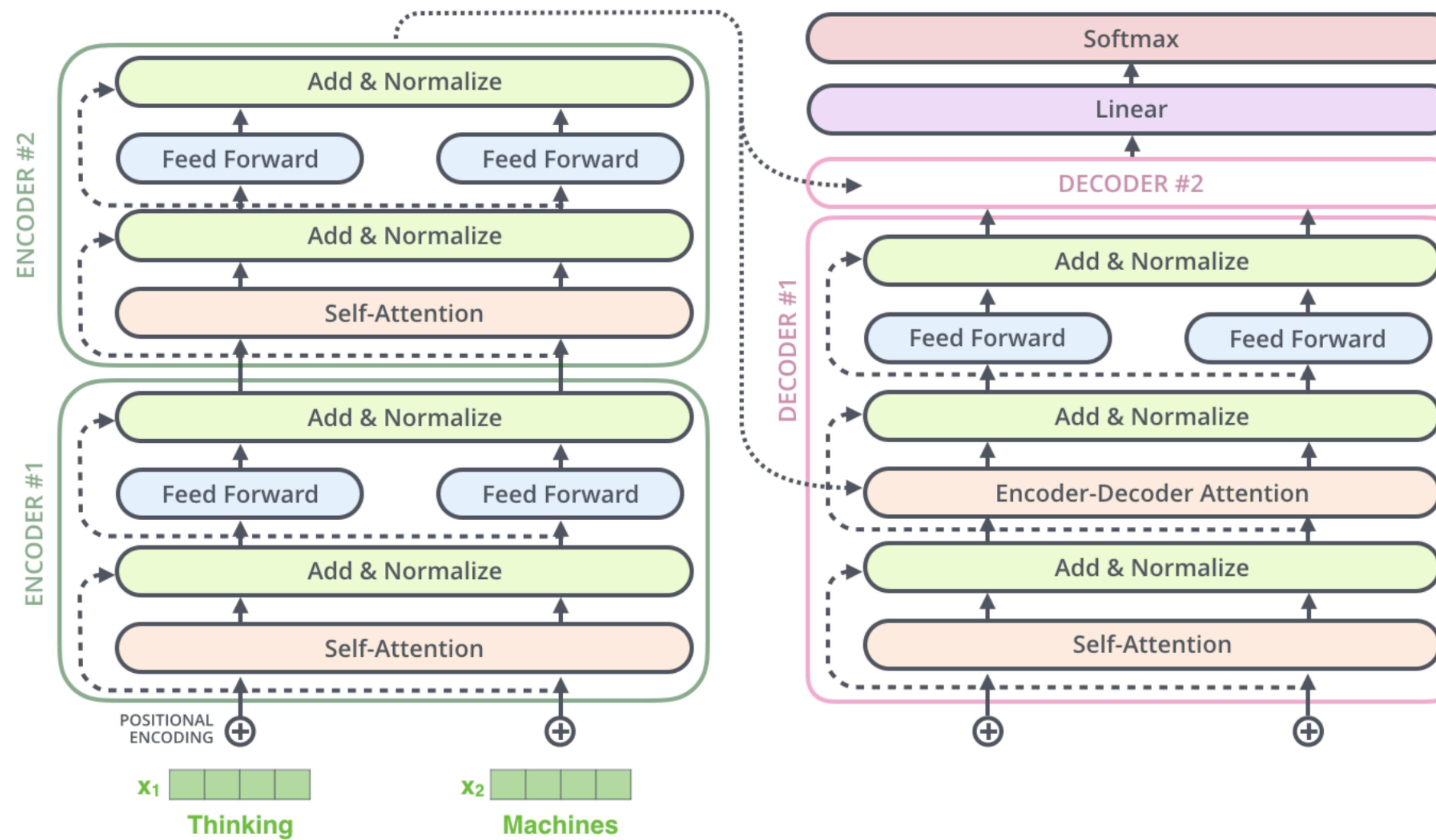


# Transformer

## Decoder



# Transformer



# Transformers

- Параллелится на GPU !
- Сохраняет сложные взаимосвязи с начала предложения в конец предложения
- Граф вычислений не зависит от размера входных данных, а зависит только от сложности архитектуры

# Ссылки, источники

- Ссылка на GitHub с лекцией и материалами: <https://github.com/nikitosl/spbu-nlp-2020>
- Ссылка на Google Colab:
- RNN:
  - <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>
  - <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
  - [https://calvinfeng.gitbook.io/machine-learning-notebook/supervised-learning/recurrent-neural-network/\\_long\\_short\\_term\\_memory](https://calvinfeng.gitbook.io/machine-learning-notebook/supervised-learning/recurrent-neural-network/_long_short_term_memory)
- Attention, Transformer:
  - <http://jalammar.github.io/illustrated-transformer/>
  - Attention is All You Need <https://papers.nips.cc/paper/2017/file/3f5ee243547dee91fbdo53c1c4a845aa-Paper.pdf>
  - <https://habr.com/ru/post/341240/>

# Заключение

- NLP is All You Need =)
- SBER, команда "Comprehend 360" [altuxov.nikita@gmail.com](mailto:altuxov.nikita@gmail.com)
  - NLP
  - Big Data
  - Research