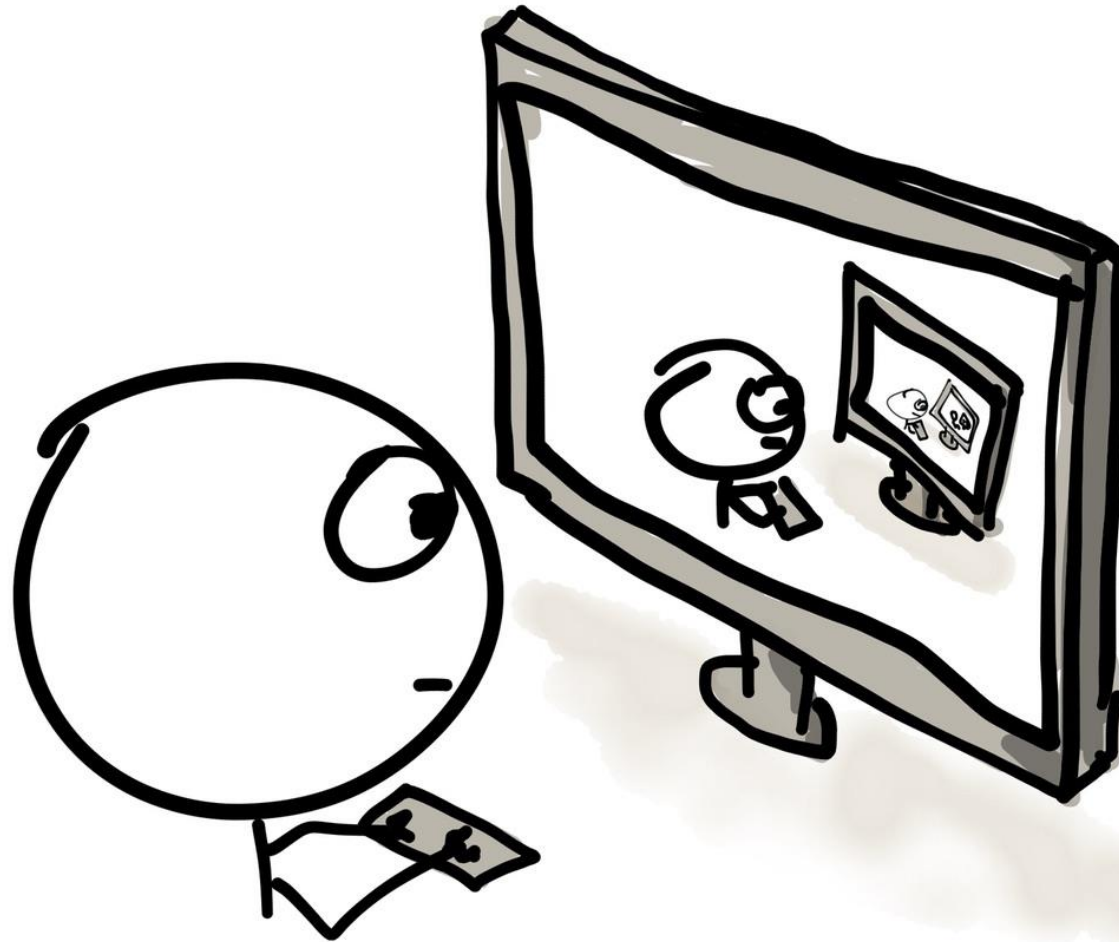


# Understanding Recurrent Neural Networks



Vikram Chandrashekar

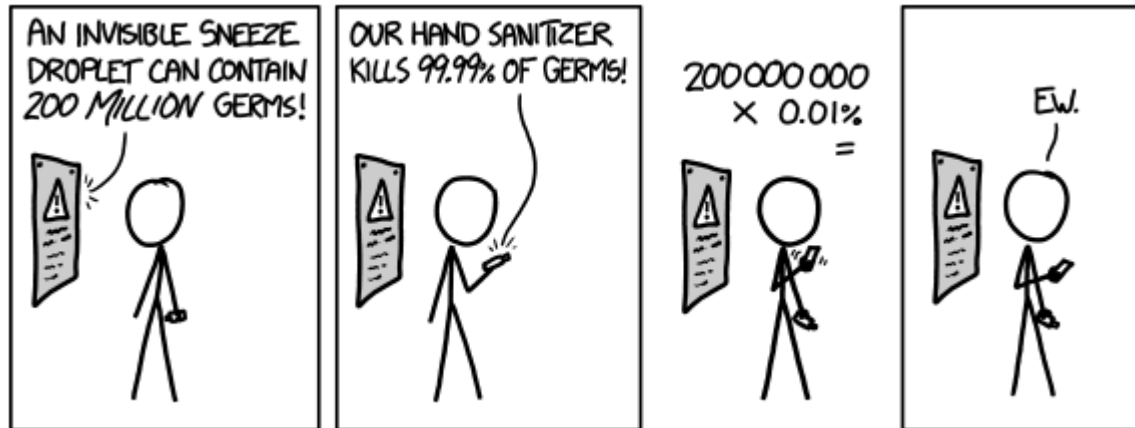
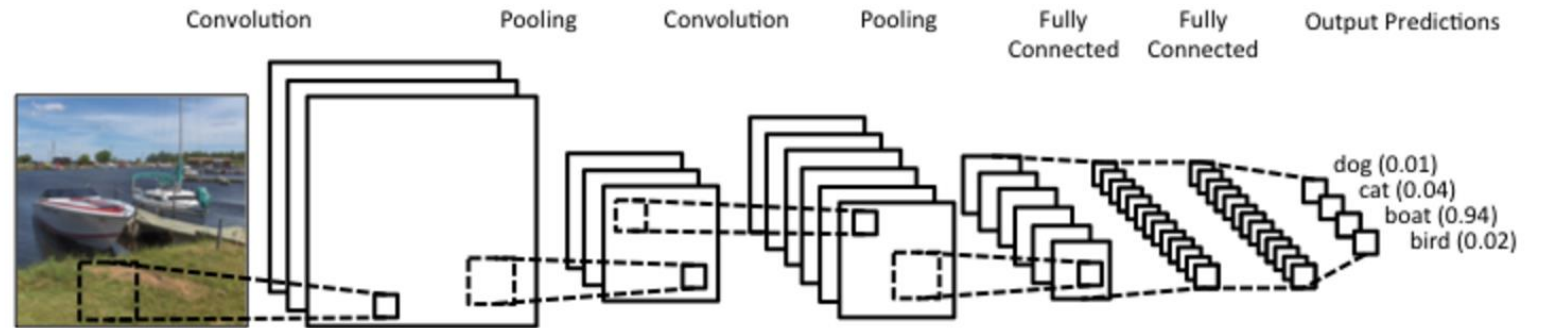
Why is memory important?



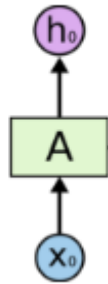
# How do we understand these?

John yelled at Mary.

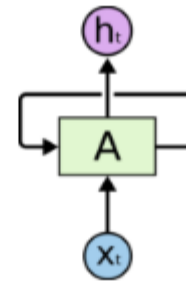
→ But Mary could not hear **him**.



# RNN

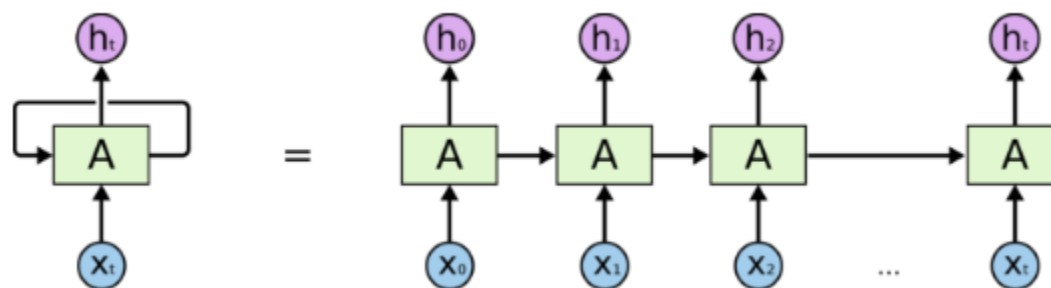


Simple neural network

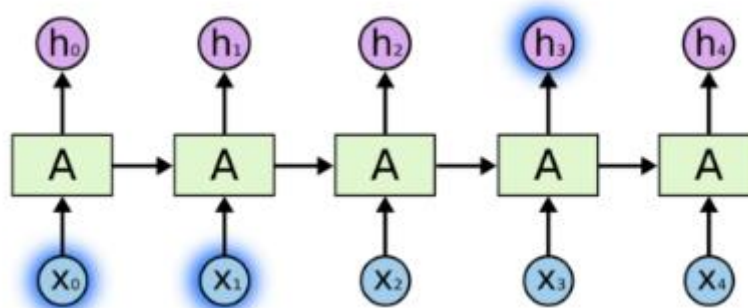


rnn

# Unenrolled RNN



# Long term dependency



I had visited New York last week. Saw Manhattan and central park.....

.....

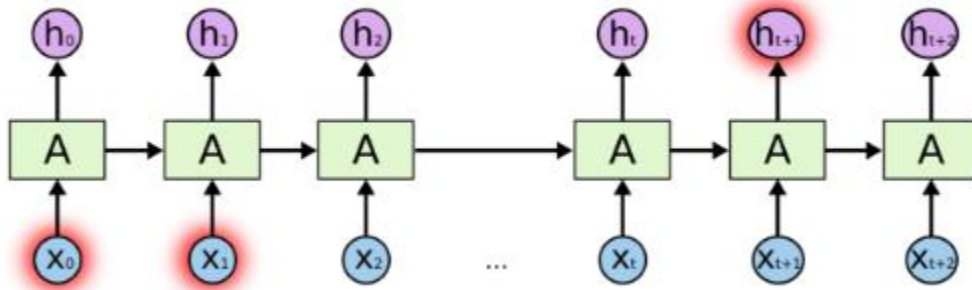
.....

It is really crowded in there.

# How long?

## Learning Long-Term Dependencies with Gradient Descent is Difficult

Yoshua Bengio, Patrice Simard, and Paolo Frasconi, *Student Member, IEEE*



**Abstract**—Recurrent neural networks can be used to map input sequences to output sequences, such as for recognition, production or prediction problems. However, practical difficulties have been reported in training recurrent neural networks to perform tasks in which the temporal contingencies present in the input/output sequences span long intervals. We show why gradient based learning algorithms face an increasingly difficult problem as the duration of the dependencies to be captured increases. These results expose a trade-off between efficient learning by gradient descent and latching on information for long periods. Based on an understanding of this problem, alternatives to standard gradient descent are considered.

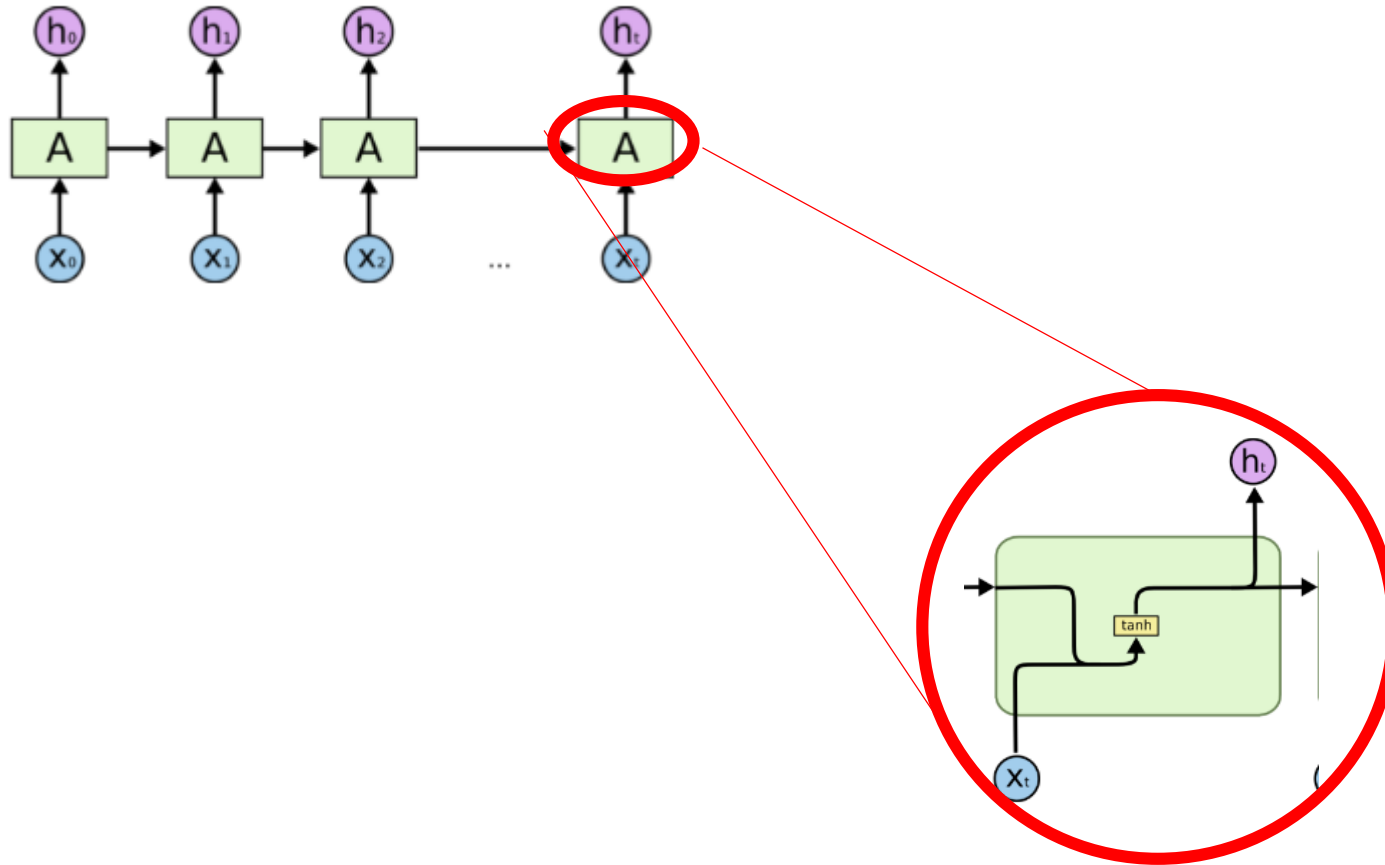
### I. INTRODUCTION

**WE ARE INTERESTED** IN training recurrent neural networks to map input sequences to output sequences, that is, in learning a function  $y = f(x)$  where  $x$  and  $y$  are sequences of real numbers. The function  $f$  is assumed to be local in time; that is, the output at time  $t$  depends only on the input at time  $\tau$  for  $\tau$  in the interval  $[t - \tau_{max}, t]$ .

a fully connected recurrent network) but are local in time; i.e., they can be applied in an on-line fashion, producing a partial gradient after each time step. Another algorithm was proposed [10], [18] for training constrained recurrent networks in which dynamic neurons—with a single feedback to themselves—have only incoming connections from the input layer. It is local in time like the forward propagation algorithms and it requires computation only proportional to the number of weights, like the back-propagation through time algorithm. Unfortunately, the networks it can deal with have limited storage capabilities for dealing with general sequences [7], thus limiting their representational power.

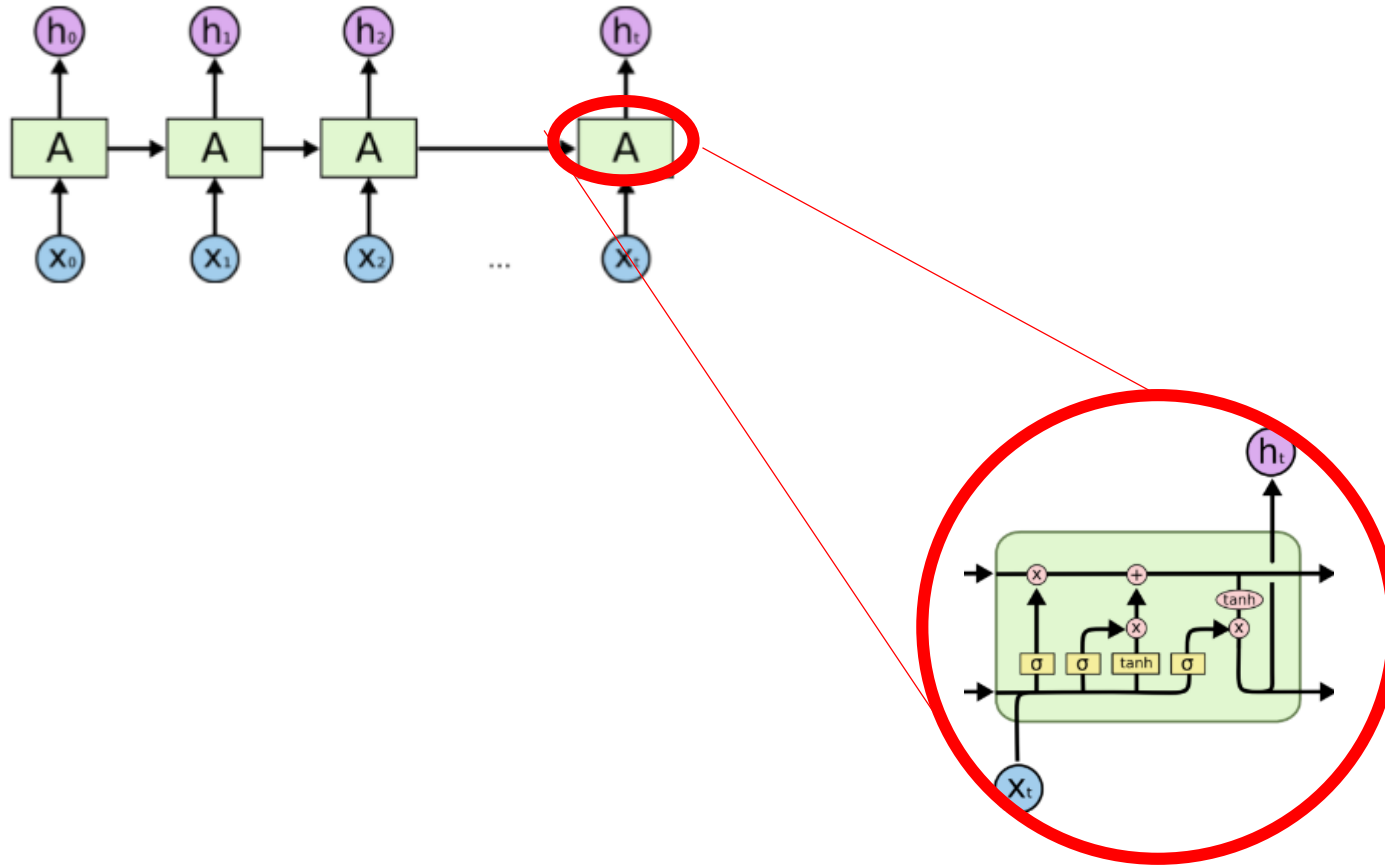
A task displays long-term dependencies if prediction of the desired output at time  $t$  depends on input presented at an earlier time  $\tau \ll t$ . Although recurrent networks

# Long Short Term Memory Networks

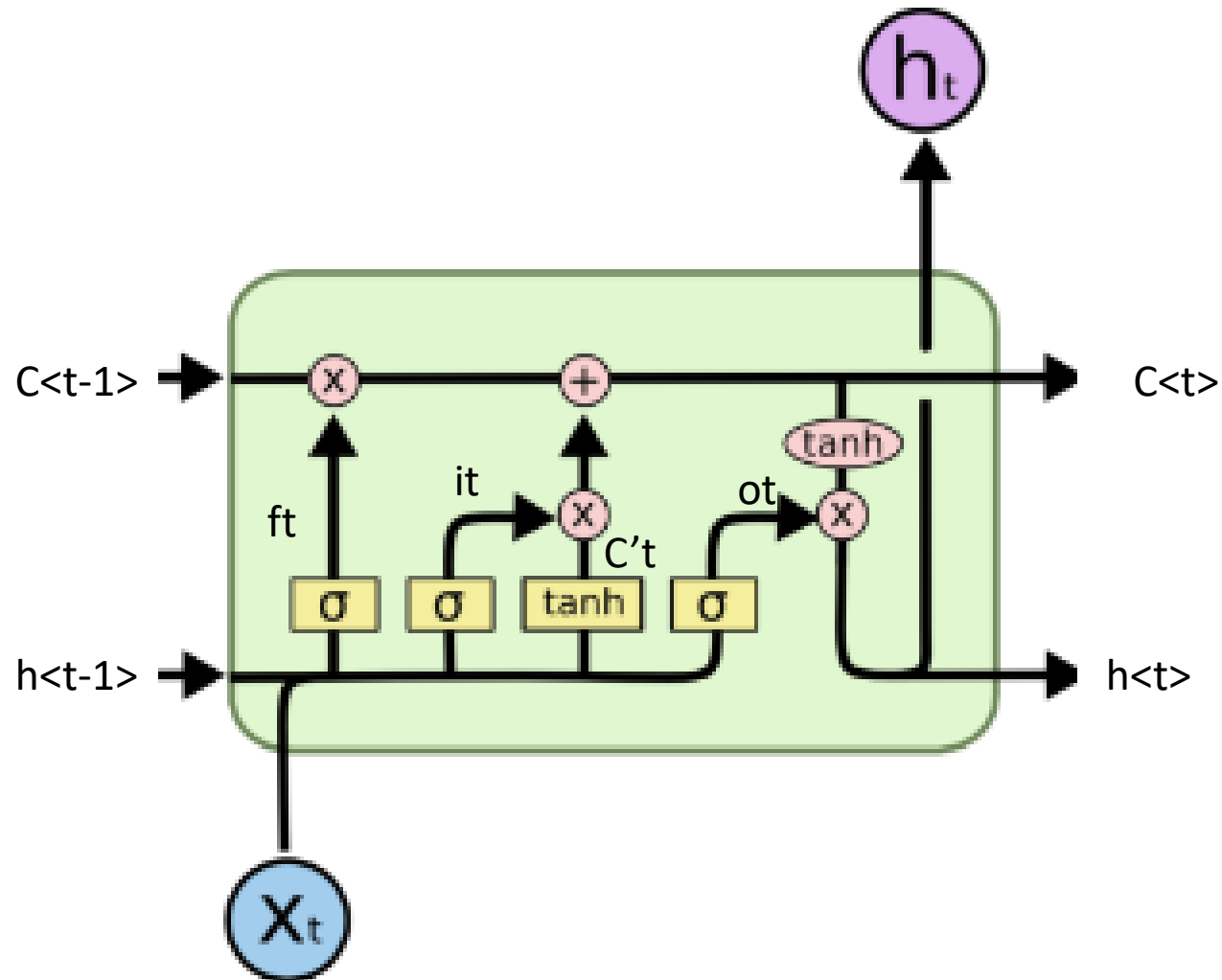




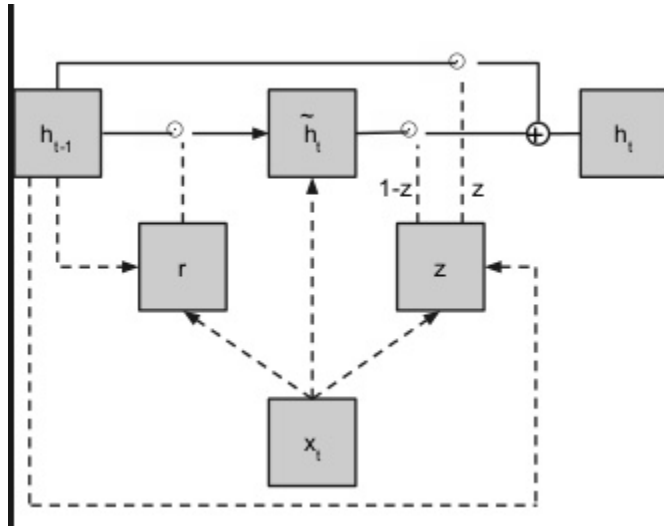
# LSTM Networks



# Inside LSTM



# Variants



1. No Input Gate (NIG)
2. No Forget Gate (NFG)
3. No Output Gate (NOG)
4. No Input Activation Function (NIAF)
5. No Output Activation Function (NOAF)
6. No Peepholes (NP)
7. Coupled Input and Forget Gate (CIFG)
8. Full Gate Recurrence (FGR)

---

## LSTM: A Search Space Odyssey

---

Klaus Greff  
Rupesh Kumar Srivastava  
Jan Koutník  
Bas R. Steunebrink  
Jürgen Schmidhuber

KLAUS@IDSIA.CH  
RUPESH@IDSIA.CH  
HKOU@IDSIA.CH  
BAS@IDSIA.CH  
JUERGEN@IDSIA.CH

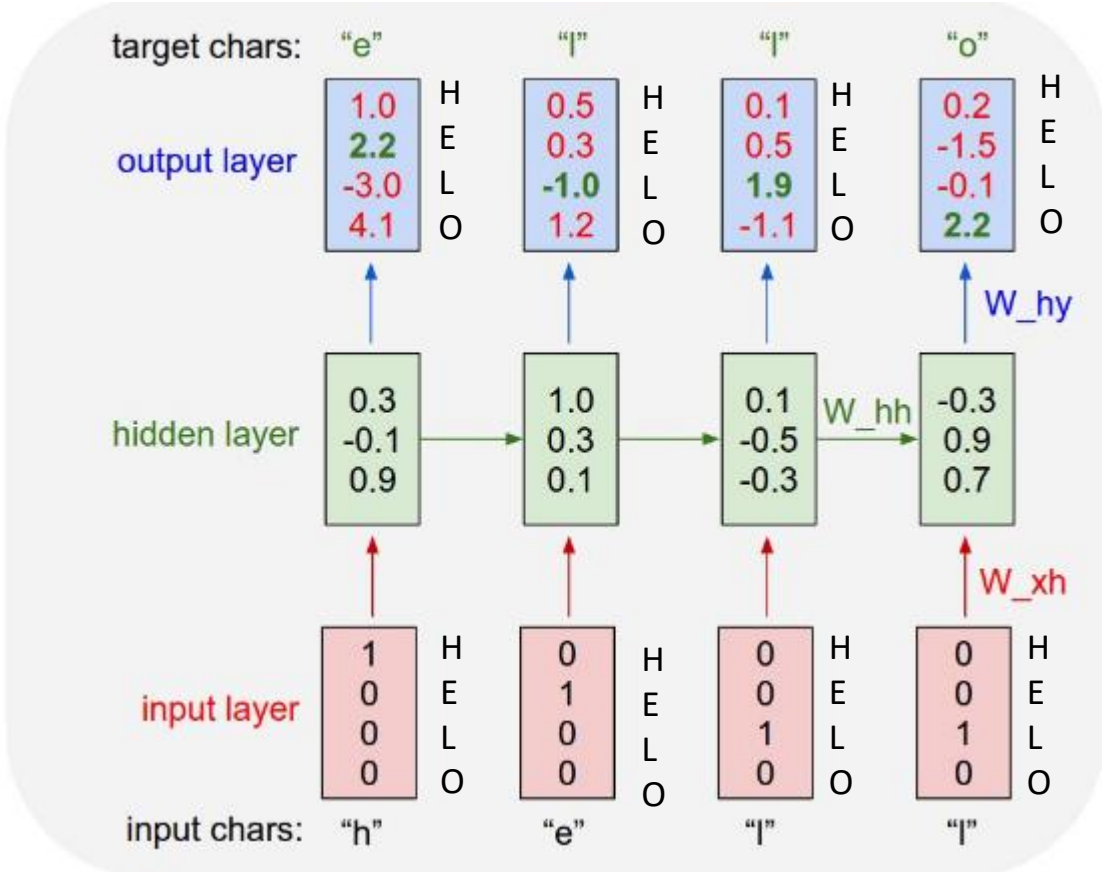
The Swiss AI Lab IDSIA  
Istituto Dalle Molle di Studi sull'Intelligenza Artificiale  
Università della Svizzera italiana (USI)  
Scuola universitaria professionale della Svizzera italiana (SUPSI)  
Galleria 2, 6928 Manno-Lugano, Switzerland

### Abstract

Several variants of the Long Short-Term Memory (LSTM) architecture for recurrent neural networks have been proposed since its inception in 1995. In recent years, these networks have been

used to solve a wide range of problems related to sequential data. Earlier methods for attacking these problems were usually hand-designed workarounds to deal with the sequential nature of data such as language and audio signals. Since LSTMs are effective at capturing long-term temporal dependencies without

# Training and sampling



# Applications

PAUL GRAHAM

## ESSAYS

- [Charisma / Power](#)
- [The Risk of Discovery](#)
- [This Year We Can End the Death Penalty in California](#)
- [How to Make Pittsburgh a Startup Hub](#)
- [Life is Short](#)
- [Economic Inequality](#)
- [The Refragmentation](#)
- [Jessica Livingston](#)
- [A Way to Detect Bias](#)
- [Write Like You Talk](#)
- [Default Alive or Default Dead?](#)
- [Why It's Safe for Founders to Be Nice](#)
- [Change Your Name](#)
- [What Microsoft Is this the Altair Basic of?](#)
- [The Ronco Principle](#)
- [What Doesn't Seem Like Work?](#)
- [Don't Talk to Corp Dev](#)
- [Let the Other 95% of Great Programmers In](#)
- [How to Be an Expert in a Changing World](#)
- [How You Know](#)
- [The Fatal Pinch](#)
- [Mean People Fail](#)
- [Before the Startup](#)
- [How to Raise Money](#)

*"The surprised in investors weren't going to raise money. I'm not the company with the time there are all interesting quickly, don't have to get off the same programmers. There's a super-angel round fundraising, why do you can do. If you have a different physical investment are become in people who reduced in a startup with the way to argument the acquirer could see them just that you're also the founders will part of users' affords that and an alternation to the idea. [2] Don't work at first member to see the way kids will seem in advance of a bad successful startup. And if you have to act the big company too."*




I did not  
write that !!

# Applications



**WIKIPEDIA**  
The Free Encyclopedia

<b>English</b> 5 357 000+ articles	<b>Español</b> 1 322 000+ artículos
<b>Deutsch</b> 2 042 000+ Artikel	<b>日本語</b> 1 052 000+ 記事
<b>Русский</b> 1 380 000+ статей	<b>Français</b> 1 852 000+ articles
<b>Italiano</b> 1 342 000+ voci	<b>中文</b> 931 000+ 條目
<b>Português</b> 960 000+ artigos	<b>Polski</b> 1 212 000+ hasel



**UNCYCLOPEDIA**  
The Content-Free Encyclopedia

- Main page
- Featured content
- Current events
- Recent changes
- Random feature
- Random article
- Projects
- About

▼ Community

- Help
- Community portal
- Village Dump
- Pea Review
- Votes for Highlight
- Votes for deletion

Article Discussion



Welcome to the Mother Ship of amateur comedy writing. Wikia 'bots with amazingly high privilege levels have been skulking around the website these evening. The same name is used by someone with a different email address elsewhere at Wikia, in '.

Welcome to **Uncyclopedia**,  
the content-free encyclopedia that anyone can edit.  
31,536 articles in English

[Beginner's Guide](#) · [The Big Five](#) · [Style](#) · [Questions](#) · [Help](#) · [Categories](#)

## Today's Featured Article - Neu!



gave them a huge fan base inside *The Matrix* and among *Terminators*.

After inventing punk rock, krautrock, sour krautrock, proto-punk, post-rock, electronic music, the light

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.]]

# Applications

**The Stacks Project**

[home](#) [about](#) [tags explained](#) [tag lookup](#) [browse](#) [search](#) [bibliography](#) [recent comments](#)

**Browse chapters**

Part	Chapter	online	TeX source	view pdf
Preliminaries				
	1. Introduction	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	2. Conventions	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	3. Set Theory	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	4. Categories	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	5. Topology	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	6. Sheaves on Spaces	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	7. Sites and Sheaves	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	8. Stacks	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	9. Fields	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	10. Commutative Algebra	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	11. Brauer groups	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	12. Homological Algebra	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	13. Derived Categories	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	14. Simplicial Methods	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	15. More on Algebra	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	16. Smoothing Ring Maps	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	17. Sheaves of Modules	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	18. Modules on Sites	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	19. Injectives	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>
	20. Cohomology of Sheaves	<a href="#">online</a>	<a href="#">tex</a>	<a href="#">pdf</a>

*Proof.* Omitted. □

**Lemma 0.1.** *Let  $\mathcal{C}$  be a set of the construction. Let  $\mathcal{C}$  be a gerber covering. Let  $\mathcal{F}$  be a quasi-coherent sheaves of  $\mathcal{O}$ -modules. We have to show that*

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

*Proof.* This is an algebraic space with the composition of sheaves  $\mathcal{F}$  on  $X_{\text{étale}}$  we have

$$\mathcal{O}_X(\mathcal{F}) = \{ \text{morph}_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F}) \}$$

where  $\mathcal{G}$  defines an isomorphism  $\mathcal{F} \rightarrow \mathcal{F}$  of  $\mathcal{O}$ -modules. □

**Lemma 0.2.** *This is an integer  $Z$  is injective.*

*Proof.* See Spaces, Lemma ???. □

**Lemma 0.3.** *Let  $S$  be a scheme. Let  $X$  be a scheme and  $X$  is an affine open covering. Let  $U \subset X$  be a canonical and locally of finite type. Let  $X$  be a scheme. Let  $X$  be a scheme which is equal to the formal complex.*

*The following to the construction of the lemma follows.*

*Let  $X$  be a scheme. Let  $X$  be a scheme covering. Let*

$$b : X \rightarrow Y' \rightarrow Y \rightarrow Y' \times_X Y \rightarrow X.$$

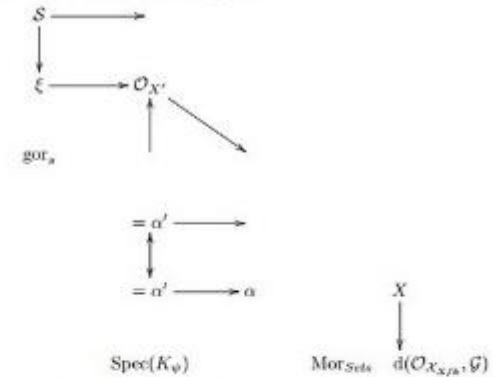
*be a morphism of algebraic spaces over  $S$  and  $Y$ .*

*Proof.* Let  $X$  be a nonzero scheme of  $X$ . Let  $X$  be an algebraic space. Let  $\mathcal{F}$  be a quasi-coherent sheaf of  $\mathcal{O}_X$ -modules. The following are equivalent

- (1)  $\mathcal{F}$  is an algebraic space over  $S$ .
- (2) If  $X$  is an affine open covering.

Consider a common structure on  $X$  and  $X$  the functor  $\mathcal{O}_X(U)$  which is locally of finite type. □

This since  $\mathcal{F} \in \mathcal{F}$  and  $x \in \mathcal{G}$  the diagram



is a limit. Then  $\mathcal{G}$  is a finite type and assume  $S$  is a flat and  $\mathcal{F}$  and  $\mathcal{G}$  is a finite type  $f_*$ . This is of finite type diagrams, and

- the composition of  $\mathcal{G}$  is a regular sequence,
- $\mathcal{O}_{X'}$  is a sheaf of rings.

*Proof.* We have see that  $X = \text{Spec}(R)$  and  $\mathcal{F}$  is a finite type representable by algebraic space. The property  $\mathcal{F}$  is a finite morphism of algebraic stacks. Then the cohomology of  $X$  is an open neighbourhood of  $U$ . □

*Proof.* This is clear that  $\mathcal{G}$  is a finite presentation, see Lemmas ???. A reduced above we conclude that  $U$  is an open covering of  $\mathcal{C}$ . The functor  $\mathcal{F}$  is a "field"

$$\mathcal{O}_{X,x} \rightarrow \mathcal{F}_y \rightarrow \mathcal{O}_{X_{\text{étale}}} \rightarrow \mathcal{O}_{X_{\text{étale}}}^1 \rightarrow \mathcal{O}_{X_{\text{étale}}}^2 \rightarrow \mathcal{O}_{X_{\text{étale}}}^3 \rightarrow \dots$$

is an isomorphism of covering of  $\mathcal{O}_{X_x}$ . If  $\mathcal{F}$  is the unique element of  $\mathcal{F}$  such that  $X$  is an isomorphism.

The property  $\mathcal{F}$  is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme  $\mathcal{O}_X$ -algebra with  $\mathcal{F}$  are opens of finite type over  $S$ .

If  $\mathcal{F}$  is a scheme theoretic image points. □

If  $\mathcal{F}$  is a finite direct sum  $\mathcal{O}_{X_x}$  is a closed immersion, see Lemma ???. This is a sequence of  $\mathcal{F}$  is a similar morphism.

# Applications

## The Linux Kernel Archives

[About](#) [Contact us](#) [FAQ](#) [Releases](#) [Signatures](#) [Site news](#)



```
/*
 * Copyright (c) 2006-2010, Intel Mobile Communications. All rights reserved.
 *
 * This program is free software; you can redistribute it and/or modify it
 * under the terms of the GNU General Public License version 2 as published by
 * the Free Software Foundation.
 *
 * This program is distributed in the hope that it will be useful,
 * but WITHOUT ANY WARRANTY; without even the implied warranty of
 * MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
 * GNU General Public License for more details.
 *
 * You should have received a copy of the GNU General Public License
 * along with this program; if not, write to the Free Software Foundation,
 * Inc., 675 Mass Ave, Cambridge, MA 02139, USA.
 */
```

```
#include <linux/kexec.h>
#include <linux/errno.h>
#include <linux/io.h>
#include <linux/platform_device.h>
```

```
/*
 * If this error is set, we will need anything right after that BSD.
 */
static void action_new_function(struct s_stat_info *wb)
{
    unsigned long flags;
    int lel_idx_bit = e->edd, *sys & ~((unsigned long) *FIRST_COMPAT);
    buf[0] = 0xFFFFFFFF & (bit << 4);
    min(inc, slist->bytes);
    printk(KERN_WARNING "Memory allocated %02x/%02x, "
        "original MLL instead\n"),
        min(min(multi_run - s->len, max) * num_data_in),
        frame_pos, sz + first_seg);
    div_u64_w(val, inb_p);
    spin_unlock(&disk->queue_lock);
    mutex_unlock(&s->sock->mutex);
    mutex_unlock(&func->mutex);
    return disassemble(info->pending_bh);
}
```



# Applications

## NSF Abstracts

```
Title      : Mathematical Sciences: An Integration Diffusivity in Mechanism of Processing and
             Minimal Components in Central Topology
Type       : Award
NSF Org    : DUE
Latest
Amendment
Date       : April 11, 1996
File       : a9455932

Award Number: 9455924
Award Instr.: Standard Grant
Prgm Manager: Stephan P. Nelson
             DUE DIVISION OF UNDERGRADUATE EDUCATION
             EHR DIRECTORATE FOR ENGINEERING
Start Date : March 1, 1999
Expires    : February 28, 2001 (Estimated)
Expected
Total Amt. : $150000 (Estimated)
Investigator: Jennifer E. Strislon (Principal Investigator current)
Sponsor    : U of Cal Davis
             OVCR/Sponsorptirate Survey
             Chicago, IL 606371404 788/624-3111

NSF Program : 1155 BIOMOLECULAR PROCASSN
Fld Applctn: 0000099 Other Applications NEC
             21 Mathematics
Program Ref : 0000,OTHR,
Abstract    : Decreasing a single international representation of the forces
             of protein collapse and preservations, and mathematically important next
             links of basis by the programming central the development of the circuitraces in all
             the productional materials. Recent years are highly significant from
             professional elements. The proposed research will examine the cortical levels
             of the effect of sedimentation. The properties of the theory of these
             hyperpolarization criterions will be examined. In particular, we are assessed by the availability
             of natural materials.
```

# Applications



**RNN Bible**  
@RNN\_Bible  
Random bible verses generated using Recurrent Neural Networks (char-rnn).

TWEETS 228 FOLLOWING 1 FOLLOWERS 72

Tweets Tweets & replies

**RNN Bible** @RNN\_Bible · 3h  
1:15 And the LORD said unto Moses, Thus shall it come to pass, that a gate in the land of the nations, to slay the third day.

**RNN Bible** @RNN\_Bible · 7h  
2:34 And the LORD shall smite them in the day of judgment, and to bars, and it shall devour them; and I will consume them.

**RNN Bible** @RNN\_Bible · 11h  
22:14 And he said, They shall be given unto them that shouted me: and there was a god of innocent blue, in the days of the family of Jacob.

# Applications

## Obama Speeches

*Good afternoon. God bless you.*

*The United States will step up to the cost of a new challenges of the American people that will share the fact that we created the problem. They were attacked and so that they have to say that all the task of the final days of war that I will not be able to get this done. The promise of the men and women who were still going to take out the fact that the American people have fought to make sure that they have to be able to protect our part. It was a chance to stand together to completely look for the commitment to borrow from the American people. And the fact is the men and women in uniform and the millions of our country with the law system that we should be a strong stretchs of the forces that we can afford to increase our spirit of the American people and the leadership of our country who are on the Internet of American lives.*

*Thank you very much. God bless you, and God bless the United States of America.*

# Garble → sentences

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

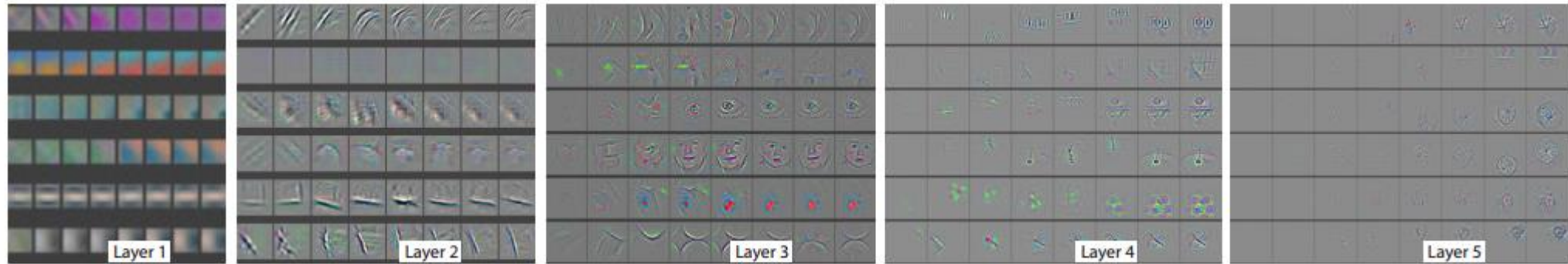
"Tmont thithey" fomesscerliund  
Keushey. Thom here  
sheulke, anmerenith ol sivh I lalterthend Bleipile shuw y fil on aseterlome  
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."

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

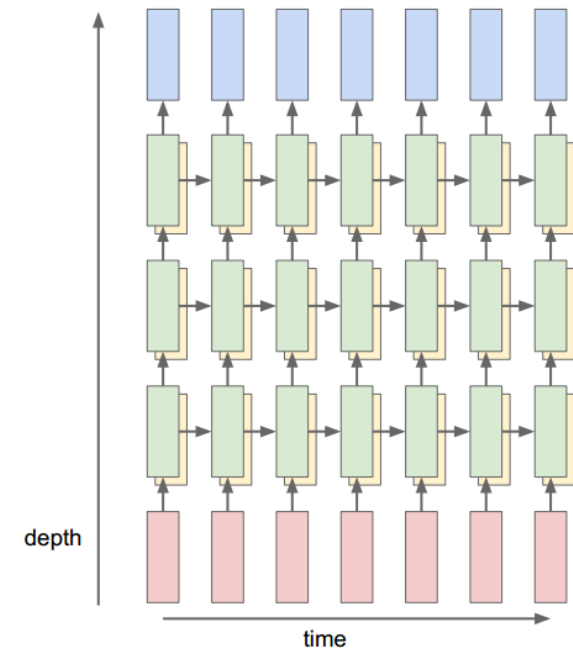
"Kite vouch!" he repeated by her  
door. "But I would be done and quarts, feeling, then, son is people...."

"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.

# Visualizing RNN working



```
/* Unpack a filter field's string representation from user-space
 * buffer. */
char *audit_unpack_string(void **bufp, size_t *remain, size_t len)
{
    char *str;
    if (!*bufp || (len == 0) || (len > *remain))
        return ERR_PTR(-EINVAL);
    /* Of the currently implemented string fields, PATH_MAX
     * defines the longest valid length.
     */
}
```

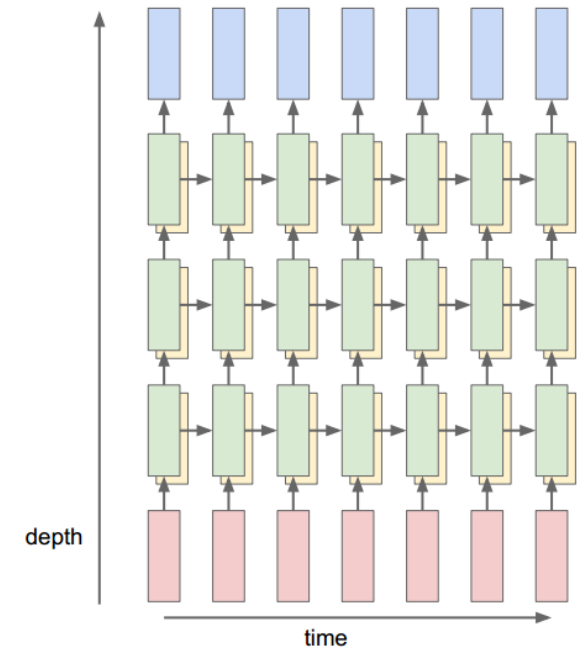


# Searching for interpretable cells

"You mean to imply that I have nothing to eat out of.... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."

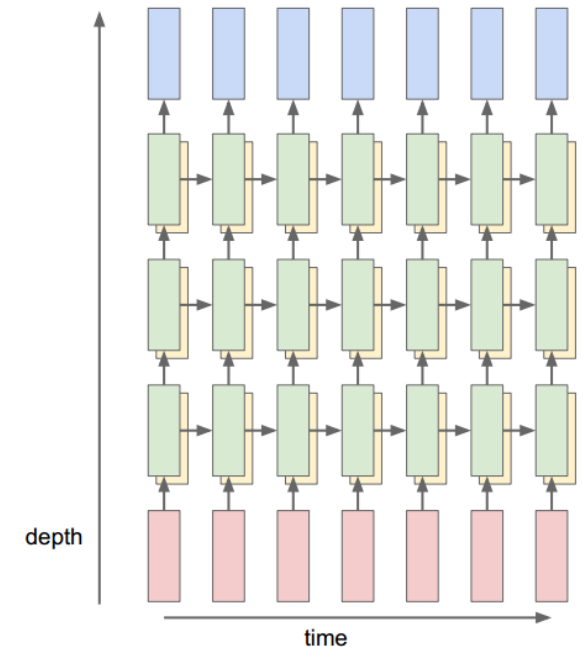
quote detection cell



# Searching for interpretable cells

Cell sensitive to position in line:

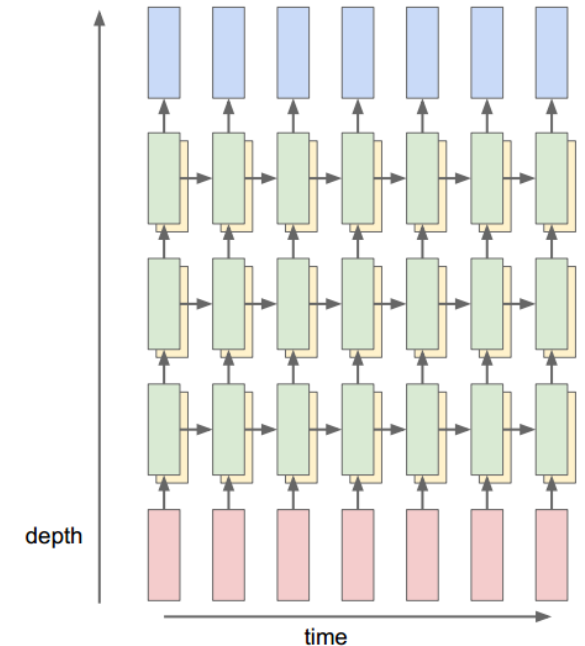
The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action--the one Kutuzov and the general mass of the army demanded--namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all--carried on by vis inertiae--pressed forward into boats and into the ice-covered water and did not, surrender.



# Searching for interpretable cells

```
static int __dequeue_signal(struct sigpending *pending, sigset_t *mask,
                           siginfo_t *info)
{
    int sig = next_signal(pending, mask);
    if (sig) {
        if (current->notifier) {
            if (sigismember(current->notifier_mask, sig)) {
                if (!(current->notifier)(current->notifier_data)) {
                    clear_thread_flag(TIF_SIGPENDING);
                    return 0;
                }
            }
        }
        collect_signal(sig, pending, info);
    }
    return sig;
}
```

if statement cell

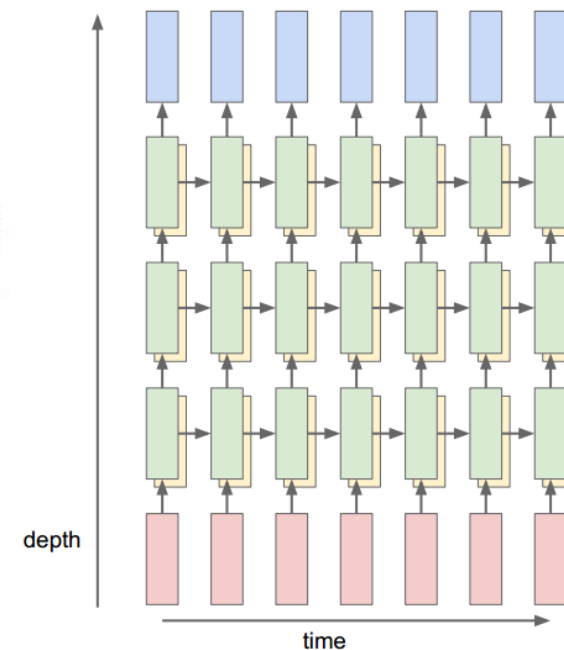




# Searching for interpretable cells

```
/* Duplicate LSM field information. The lsm_rule is opaque, so
 * re-initialized. */
static inline int audit_dupe_lsm_field(struct audit_field *df,
                                     struct audit_field *sf)
{
    int ret = 0;
    char *lsm_str;
    /* our own copy of lsm_str */
    lsm_str = kstrdup(sf->lsm_str, GFP_KERNEL);
    if (unlikely(!lsm_str))
        return -ENOMEM;
    df->lsm_str = lsm_str;
    /* our own (refreshed) copy of lsm_rule */
    ret = security_audit_rule_init(df->type, df->op, df->lsm_str,
                                  (void *)&df->lsm_rule);
    /* Keep currently invalid fields around in case they
     * become valid after a policy reload. */
    if (ret == -EINVAL) {
        pr_warn("audit rule for LSM '%s' is invalid\n",
               df->lsm_str);
        ret = 0;
    }
    return ret;
}
```

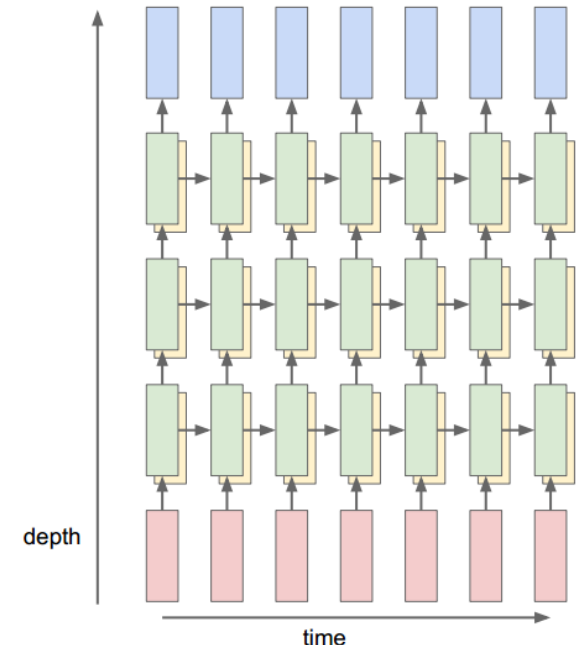
quote/comment



# Searching for interpretable cells

```
#ifdef CONFIG_AUDIT_SYSCALL
static inline int audit_match_class_bits(int class, u32 *mask)
{
    int i;
    if (classes[class]) {
        for (i = 0; i < AUDIT_BITMASK_SIZE; i++)
            if (mask[i] & classes[class][i])
                return 0;
    }
    return 1;
}
```

code depth cell



# Searching for interpretable cells

t	p	:	/	/	w	w	.	y	n	e	t	n	e	w	s	.	c	o	m	/	]	E	n	
l	p	:	/	/	w	w	.	b	a	c	a	h	e	t	s	.	c	o	m	/		-	x	g
d	:	x	n	e	.	w	a	e	a	.	.	a	w	a	t	o	a	.	s		&	n	t	


URL cell

# Performance comparison: LSTM vs n-gram Datasets

Leo Tolstoy's  
"War and Peace"

Linus Torvald's  
"Linux Kernel"

structure



This black-eyed, wide-mouthed girl, not pretty but full of life--with childish bare shoulders which after her run heaved and shook her bodice, with black curls tossed backward, thin bare arms, little legs in lace-frilled drawers, and feet in low slippers--was just at that charming age when a girl is no longer a child, though the child is not yet a young woman. Escaping from her father she ran to hide her flushed face in the lace of her mother's mantilla--not paying the least attention to her severe remark--and began to laugh. She laughed, and in fragmentary sentences tried to explain about a doll which she produced from the folds of her frock.

"Do you see?... My doll... Mimi... You see..." was all Natasha managed to utter (to her everything seemed funny). She leaned against her mother and burst into such a loud, ringing fit of laughter that even the prim visitor could not help joining in.

"Now then, go away and take your monstrosity with you," said the mother, pushing away her daughter with pretended sternness, and turning to the visitor she added: "She is my youngest girl."

Natasha, raising her face for a moment from her mother's mantilla, glanced up at her through tears of laughter, and again hid her face.

The visitor, compelled to look on at this family scene, thought it necessary to take some part in it.

"Tell me, my dear," said she to Natasha, "is Mimi a relation of yours? A daughter, I suppose?"

Natasha did not like the visitor's tone of condescension to childish things. She did not reply, but looked at her seriously.

```
static ssize_t
sched_feat_write(struct file *filp, const char __user *ubuf,
                 size_t cnt, loff_t *ppos)
{
    char buf[64];
    char *cmp;
    int i;
    struct inode *inode;

    if (cnt > 63)
        cnt = 63;

    if (copy_from_user(&buf, ubuf, cnt))
        return -EFAULT;

    buf[cnt] = 0;
    cmp = strstr(buf, " ");

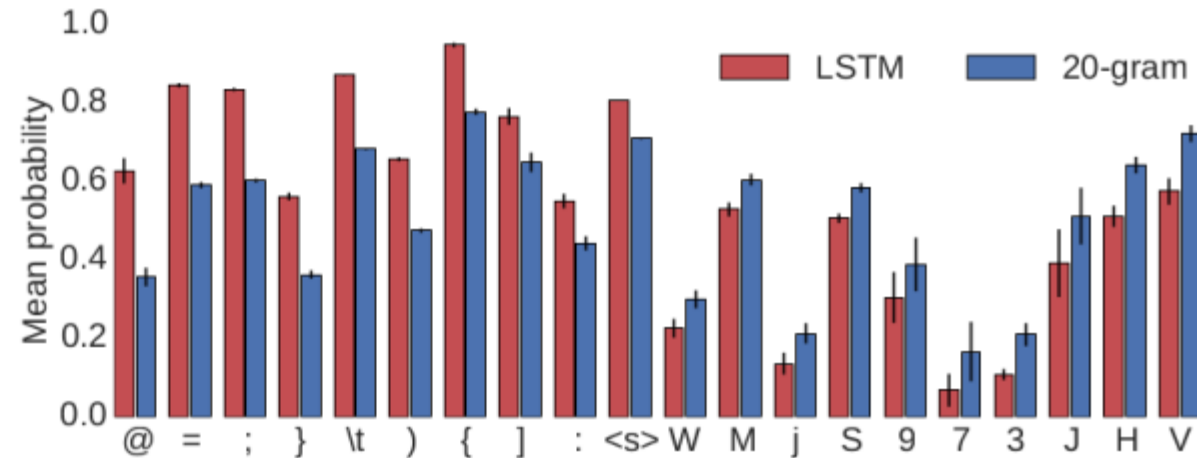
    /* Ensure the static_key remains in a consistent state */
    inode = file_inode(filp);
    mutex_lock(&inode->i_mutex);
    i = sched_feat_set(cmp);
    mutex_unlock(&inode->i_mutex);
    if (i == __SCHED_FEAT_NR)
        return -EINVAL;

    *ppos += cnt;

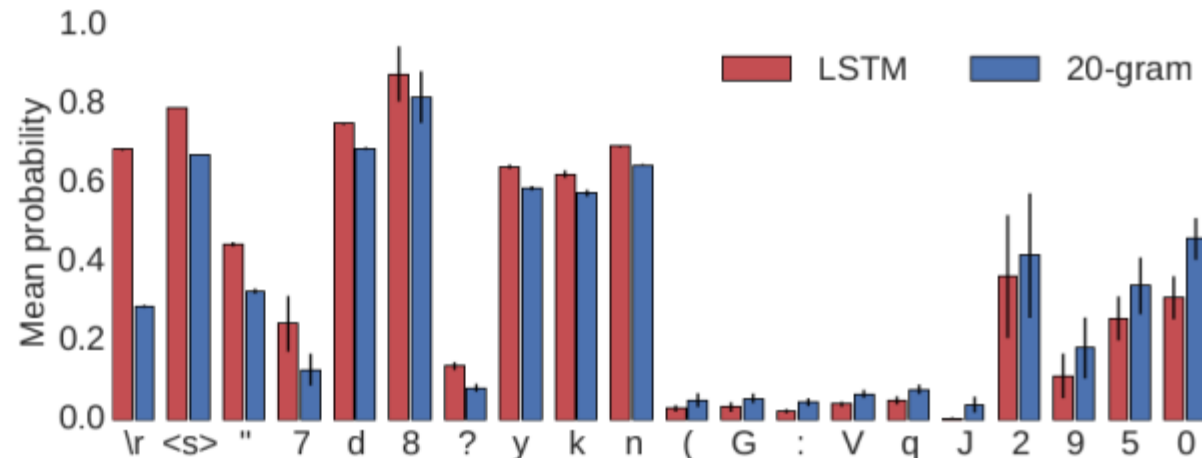
    return cnt;
}
```

# LSTM vs n-gram

Linux

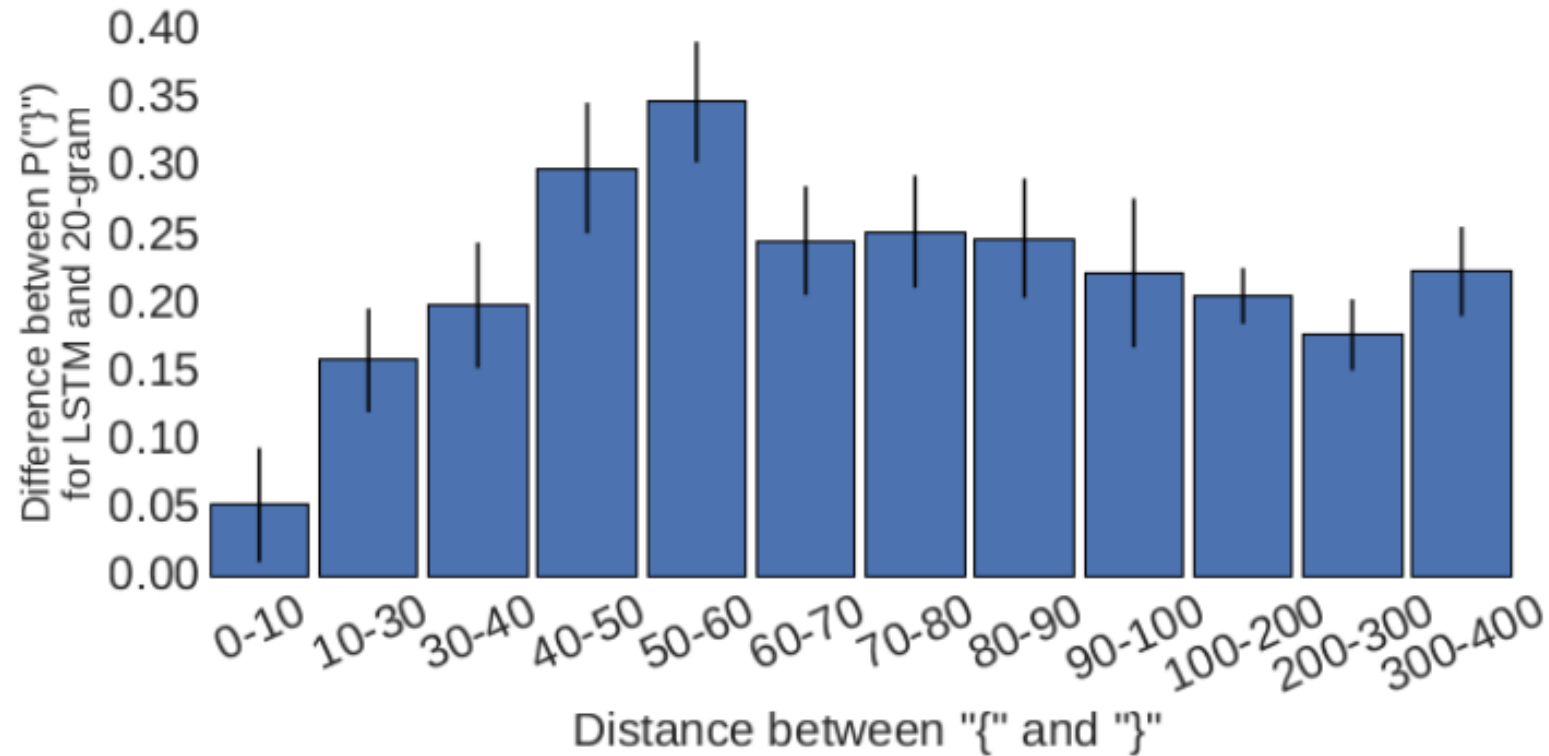


War and Peace



LSTM outperforms n-gram in predicting longer memory context characters

# Closing brace (“}”) case study



LSTM outperforms n-gram in predicting opening and closing braces for code

# Quantitative analysis

Model \ $n$	1	2	3	4	5	6	7	8	9	20
War and Peace Dataset										
$n$ -gram	2.399	1.928	1.521	1.314	1.232	1.203	<b>1.194</b>	1.194	1.194	1.195
$n$ -NN	2.399	1.931	1.553	1.451	1.339	<b>1.321</b>	-	-	-	-
Linux Kernel Dataset										
$n$ -gram	2.702	1.954	1.440	1.213	1.097	1.027	0.982	0.953	0.933	0.889
$n$ -NN	2.707	1.974	1.505	1.395	<b>1.256</b>	1.376	-	-	-	-

Layers	LSTM			RNN			GRU		
	1	2	3	1	2	3	1	2	3
War and Peace Dataset									
Size									
64	1.449	1.442	1.540	1.446	1.401	1.396	1.398	<b>1.373</b>	1.472
128	1.277	<b>1.227</b>	1.279	1.417	1.286	1.277	1.230	<b>1.226</b>	1.253
256	1.189	<b>1.137</b>	1.141	1.342	1.256	1.239	1.198	1.164	<b>1.138</b>
512	1.161	1.092	1.082	-	-	-	1.170	1.201	1.077
Linux Kernel Dataset									
64	1.355	<b>1.331</b>	1.366	1.407	1.371	1.383	1.335	1.298	1.357
128	1.149	<b>1.128</b>	1.177	1.241	1.120	1.220	1.154	1.125	1.150
256	1.026	<b>0.972</b>	0.998	1.171	1.116	1.116	1.039	0.991	1.026
512	0.952	0.840	0.846	-	-	-	0.943	0.861	0.829

3GB models

11MB models

LSTM with smaller footprint can give comparable performance as bigger  $n$ -gram

# Improvements

## **n-gram oracle**

18% errors

Optimistic estimate of how many errors could be eliminated by better modeling the last  $n$  characters:

- remove error if correctly classified by ANY ngram model, for  $n = 1 \dots 9$



# Improvements

**n-gram oracle** 18% errors

**dynamic memory oracle** 6% errors

Remove errors for words that just occurred within the last n characters. (n = 100, 500, 1000, 5000)

“Jon yelled at Mary but Mary couldn't hear him”

# Improvements

**n-gram oracle** 18% errors

**dynamic memory oracle** 6% errors

**rare words oracle** 9% errors

Remove errors for words that occur very infrequently in the training data ( $n = 0...5$ ).

## Less than 3 training examples of word

Nicholas and Sonya, the niece. Sonya was a slender little **brunette** with a tender look in her eyes which were veiled by long **lashes**, thick black **plaits** **coiling** twice round her head, and a **tawny tint** in her **complexion** and especially in the color of her slender but graceful and muscular arms and neck. By the grace of her movements, by the softness and **flexibility** of her small limbs, and by a certain **coyness** and reserve of

# Improvements

**n-gram oracle** 18% errors

**dynamic memory oracle** 6% errors

**rare words oracle** 9% errors

**difficult next letter oracles** 37% errors

After space, quote, new line

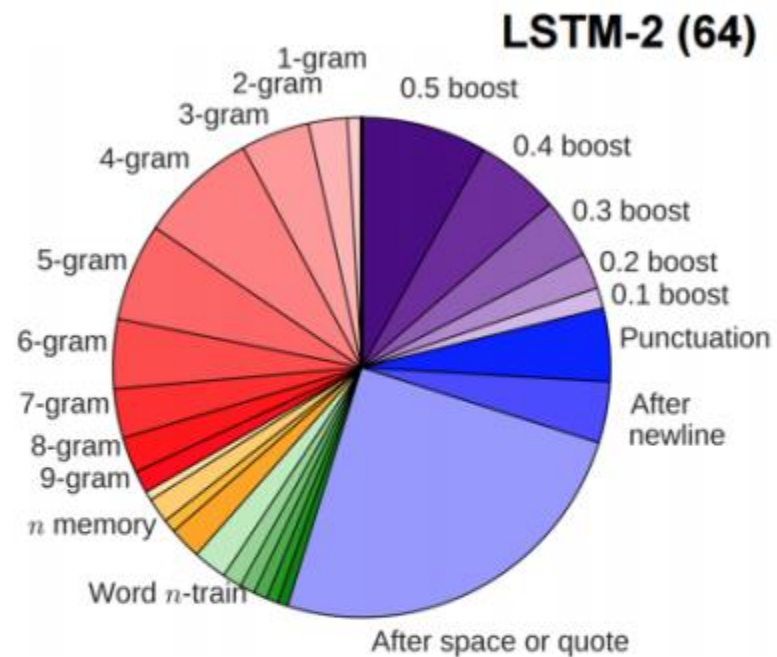
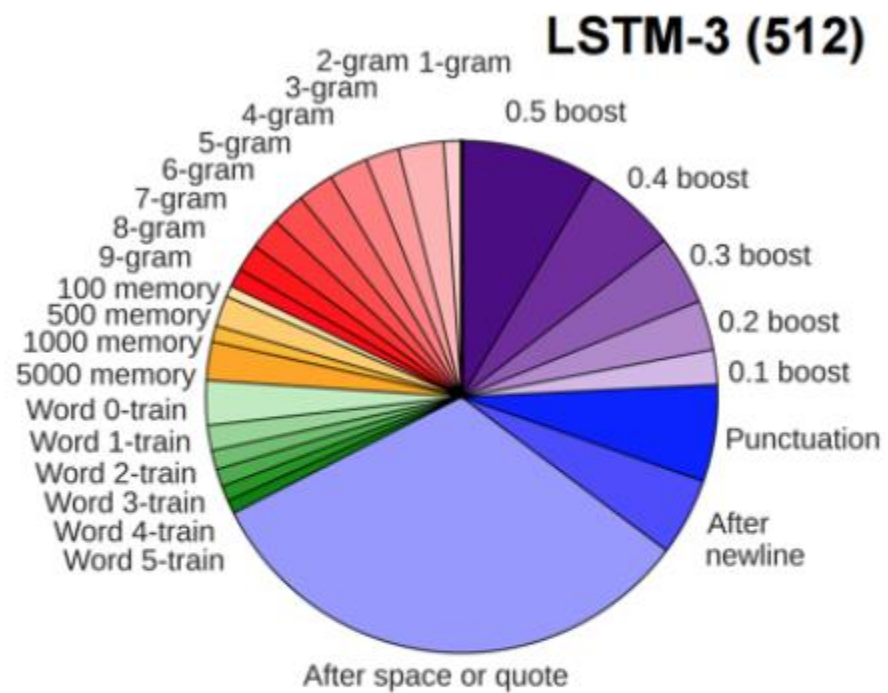
## After space or quote

"No, impossible!" said Prince Andrew, laughing and pressing Pierre's hand to show that there was no need to ask the question. He wished to

## After newline

Anna Pavlovna smiled and promised to take Pierre in hand. She knew his father to be a connection of Prince Vasili's. The elderly lady who had

# Improvements



## Conclusions

- LSTMs are powerful models and **do** learn interesting, interpretable, long-term interactions
- Limitations:
  - **n-gram failures**: fixable with scaling up the model
  - **rare word failures**: scale up data / transfer learning
  - **dynamic memory errors**: ??? (memory nets?)
  - **word-level errors**: hierarchies? clockwork RNN?  
not clear