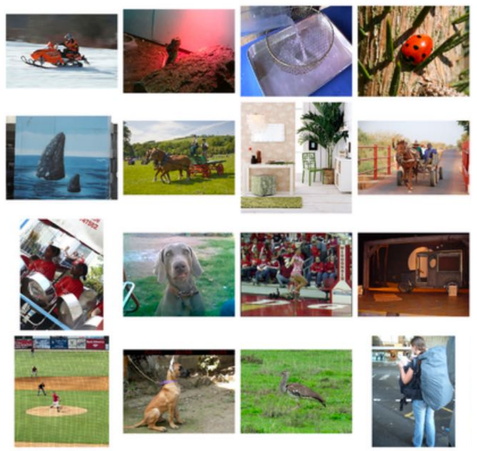
Towards Visually Interactive Neural Probabilistic Models

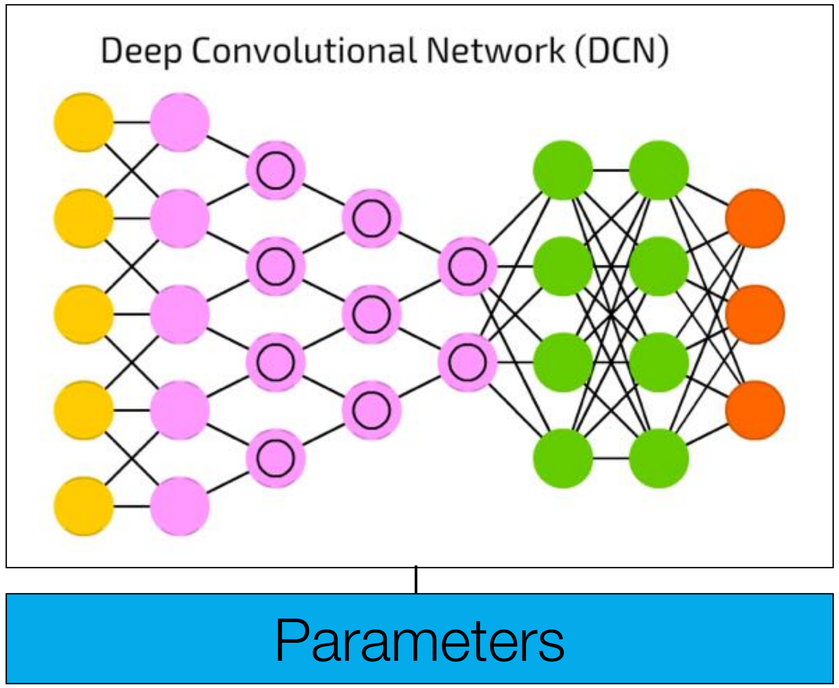
Hanspeter Pfister
pfister@seas.harvard.edu

Harvard John A. Paulson
School of Engineering
and Applied Sciences

Deep Neural Networks



Input



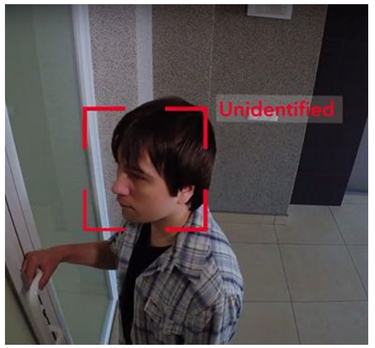
Cat
Dog
Sheep
…

Output

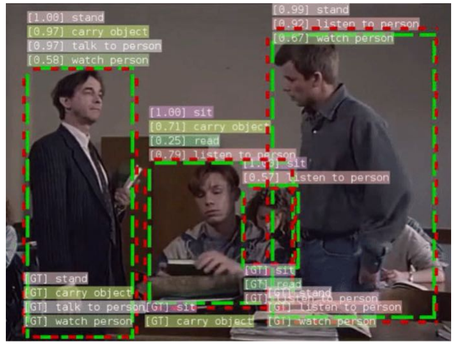
Computer Vision

Object Tracking

Object Recognition



Face Recognition

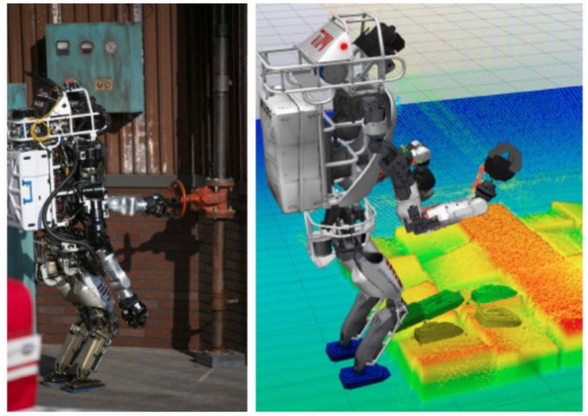


Activity Recognition

"construction worker in orange safety vest is working on road."

Image Captioning

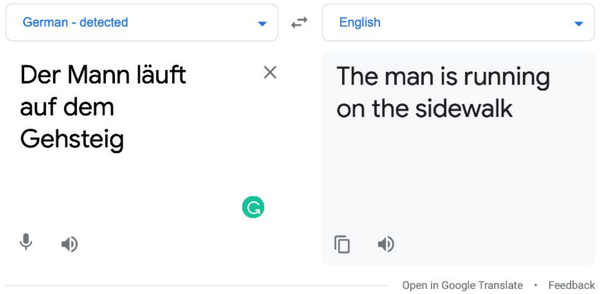
Live Video Captioning



Robotics

Self-Driving Cars

Natural Language Processing



Language Translation

Speech Recognition & Synthesis

TEAM	WIN	LOSS	PTS	FG_PCT	RB	AST	...
Pacers	4	6	99	42	40	17	...
Celtics	5	4	105	44	47	22	...

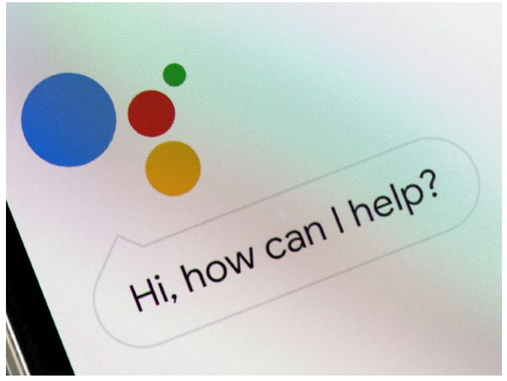
PLAYER	H/V	AST	RB	PTS	FG	CITY	...
Jeff Teague	H	4	3	20	4	Indiana	...
Miles Turner	H	1	8	17	6	Indiana	...
Isaiah Thomas	V	5	0	23	4	Boston	...
Kelly Olynyk	V	4	6	16	6	Boston	...
Amir Johnson	V	3	9	14	4	Boston	...
...	...	...	...	...	...	...	...

PTS: points, FT_PCT: free throw percentage, RB: rebounds, AST: assists, H/V: home or visiting, FG: field goals, CITY: player team city.

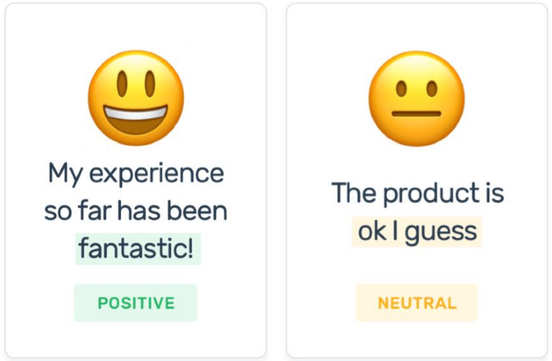
The **Boston Celtics** defeated the host house on Saturday. In a battle betwe able to prevail with a much needed ro as the **Celtics** outshot the **Pacers** fron free-throw line. Boston also held India from long distance. The Celtics also while tying the Pacers in turnovers. T game went down to the final seconds of injuries, but they had the fortunate **iah** Thomas led the team in scoring, shooting. He got most of those point **Kelly Olynyk** got a rare start and finis **rebounds and four assists**.

Table 1: Example of data-records and document summary. Entities and values corres

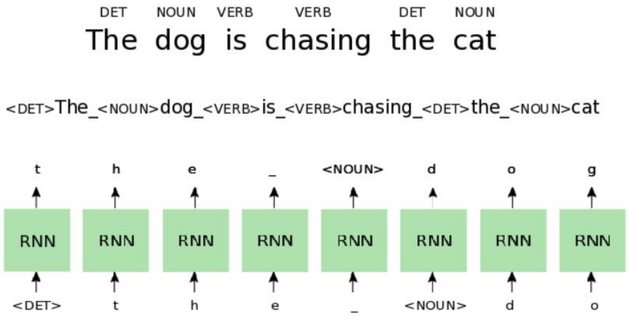
Data to Text



Conversational Bots



Sentiment Analysis



Language Modeling

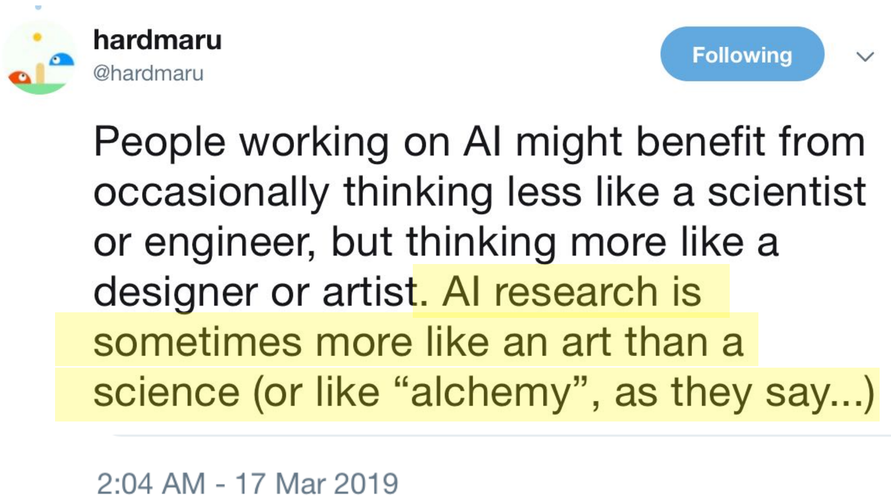
Do we completely understand
how deep neural networks
achieve these amazing results?


SPOILER
ALERT!

No!

Art vs. Science

- In general, bigger neural network models with millions of parameters yield better performance
- Difficult to understand and debug (art vs. science)



(Matt Rourke/AP) Three Mile Island nuclear power plant, Middletown, Pa.

Human-Centered Design

Technology has to be designed in such a way that humans can use it – even when we are stressed out or tired or just liable to make mistakes.

The interfaces around us need to be accepting of human limitations. They have to present a world comprehensible to us.

This is the idea of **empathy**.

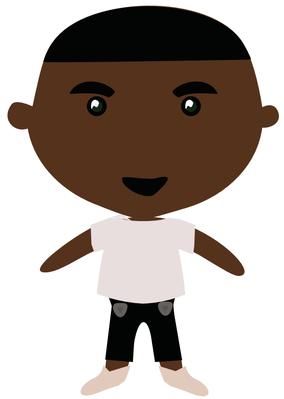
Good Design is Human, Wireframe Podcast, October 9, 2018

Human-Centered AI

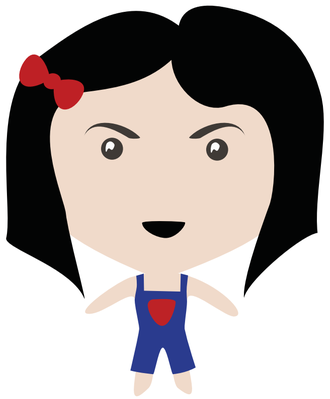
Amplify, Augment, Empower & Enhance People

Ben Shneiderman, 2021

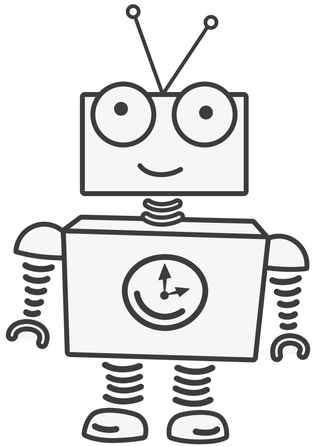
Human-AI Collaboration



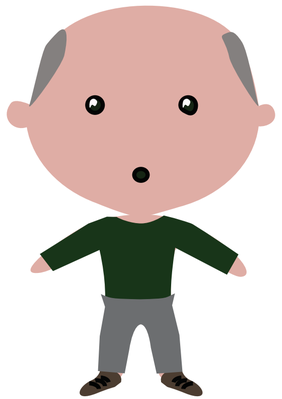
Cesar



Anna



AI-reen



Bert

AI-reen is from http://robotodex.com/

Sasha
Rush

Sebastian
Gehrmann

Hendrik
Strobelt

Johanna
Beyer

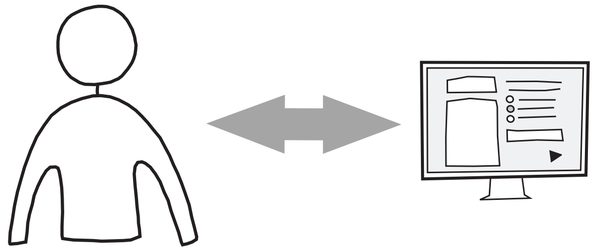
Jambay
Kinley

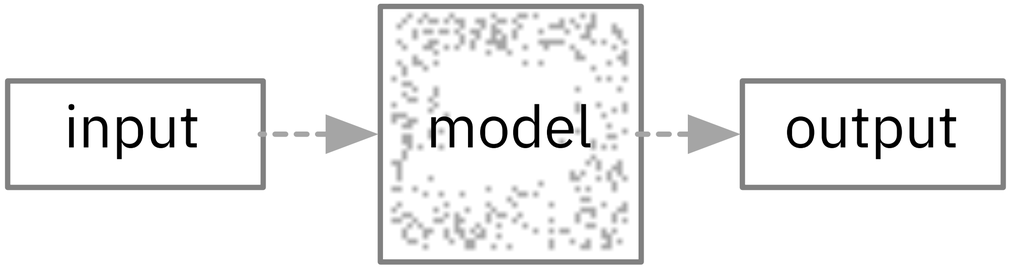
Adam
Perer

Robert
Krueger

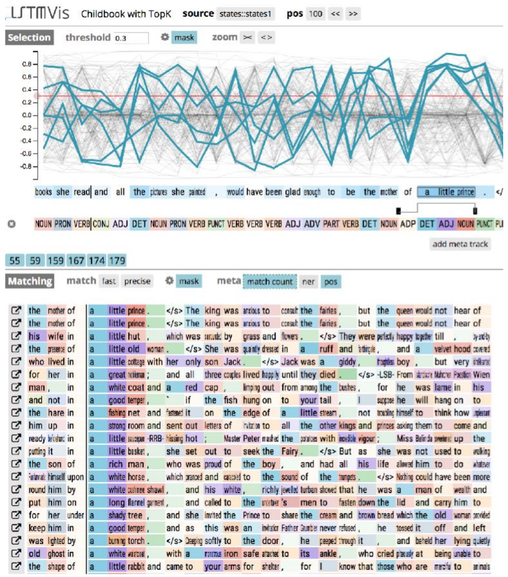
Michael
Behrisch

Act 1 - Observing



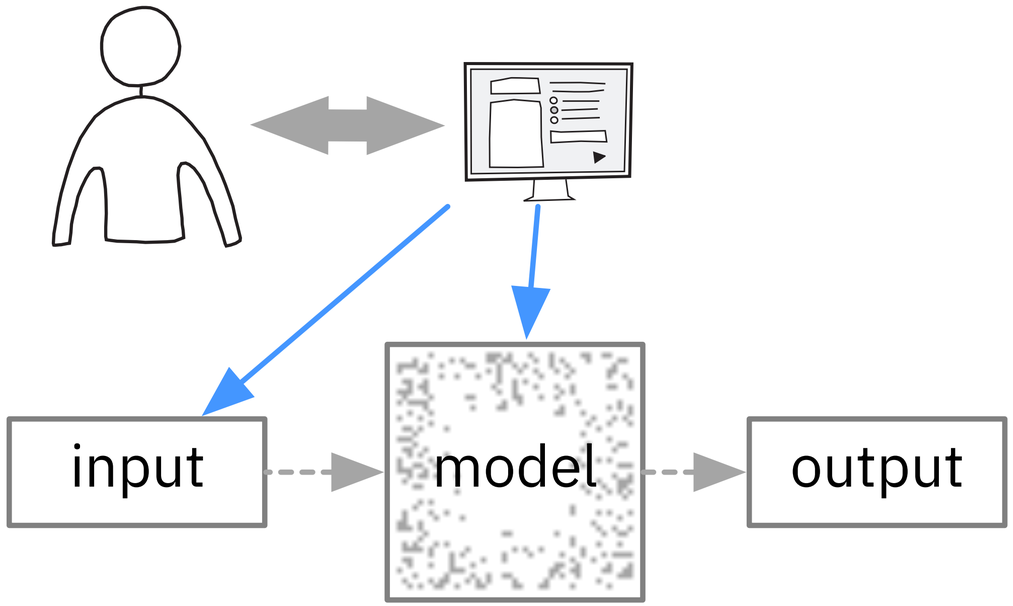


LSTMVis

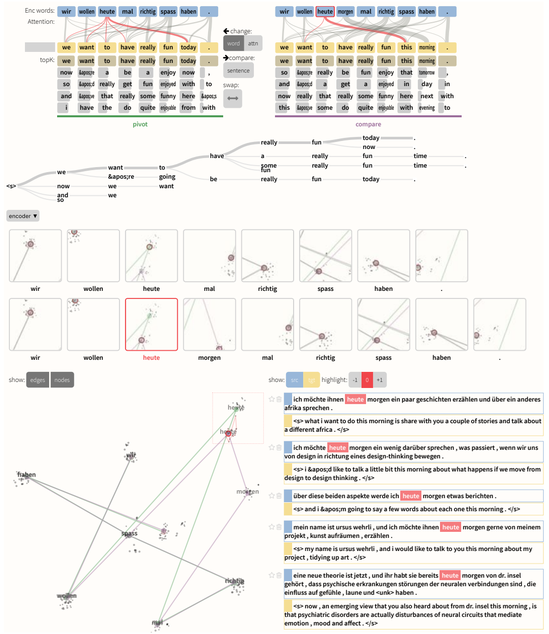


[Strobelt et al., TVCG 2017]

Act 2 - Interacting

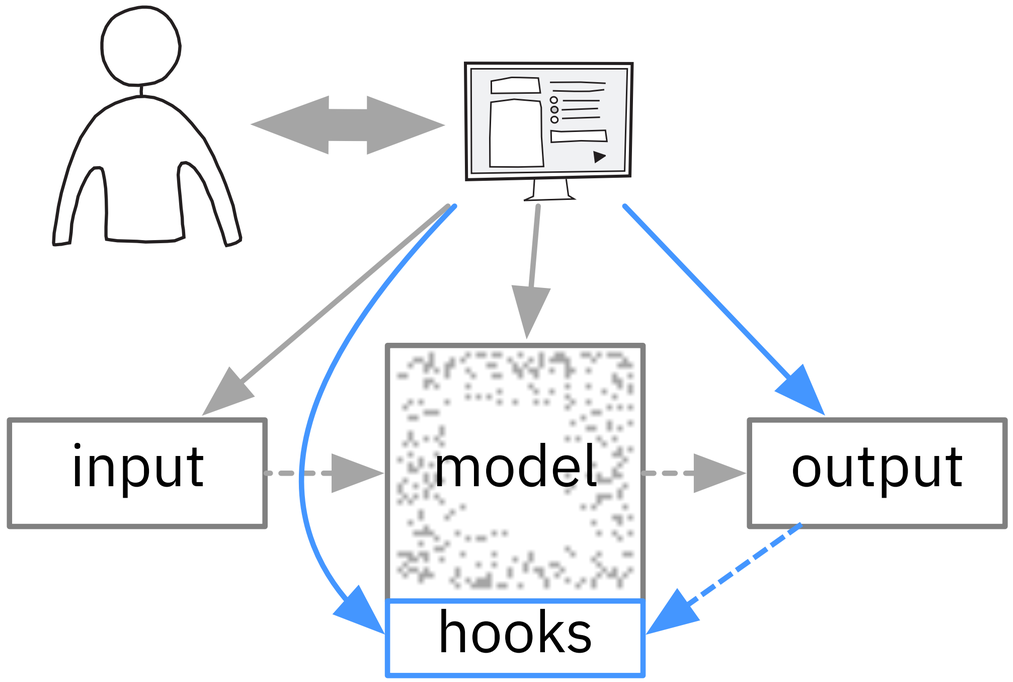


Seq2Seq-Vis

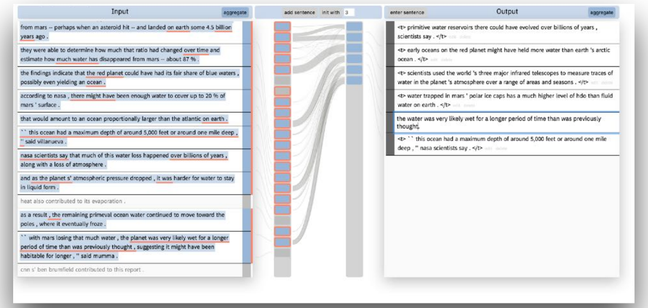


[Strobelt et al., TVCG 2018]

Act 3 - Collaborating

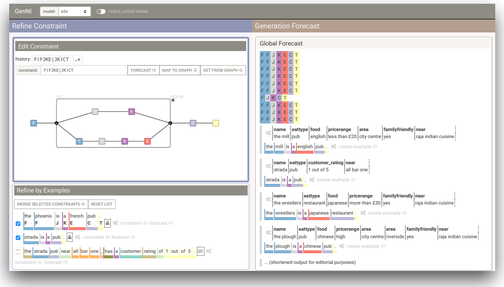


CSI



[Gehrmann et al., TVCG 2019]

GenNI



[Strobelt et al., TVCG 2021]

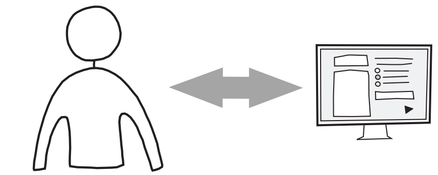
Three Messages

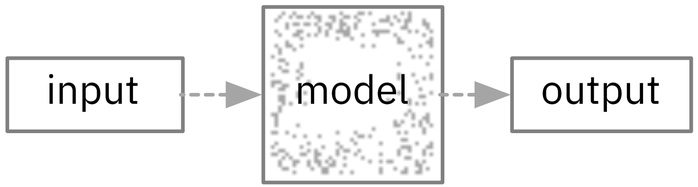
1. Interfaces to deep learning systems need to follow the principles of **human-centered design**.

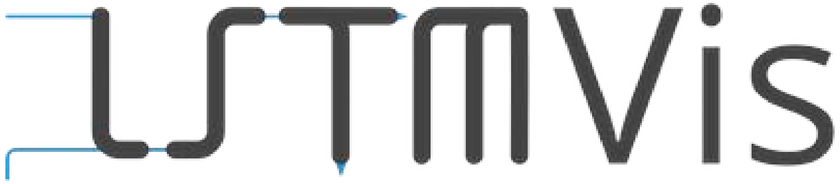
2. To re-establish the human agency over deep learning systems we introduce the concept of **interactive collaboration**.

3. Neural network models need to be extended to include **intermediate representations (“hooks”)** that can be **understood and acted upon by humans.**

Act 1: Observing







A Tool for Visual Analysis of Hidden State Dynamics in RNNs

Hendrik Strobelt, Sebastian Gehrmann,
Hanspeter Pfister, Alexander Rush

[TVCG 2017]

Harvard John A. Paulson
School of Engineering
and Applied Sciences

CORNELL
TECH

IBM Research AI
Visual AI Lab

Machine translation

Text classification

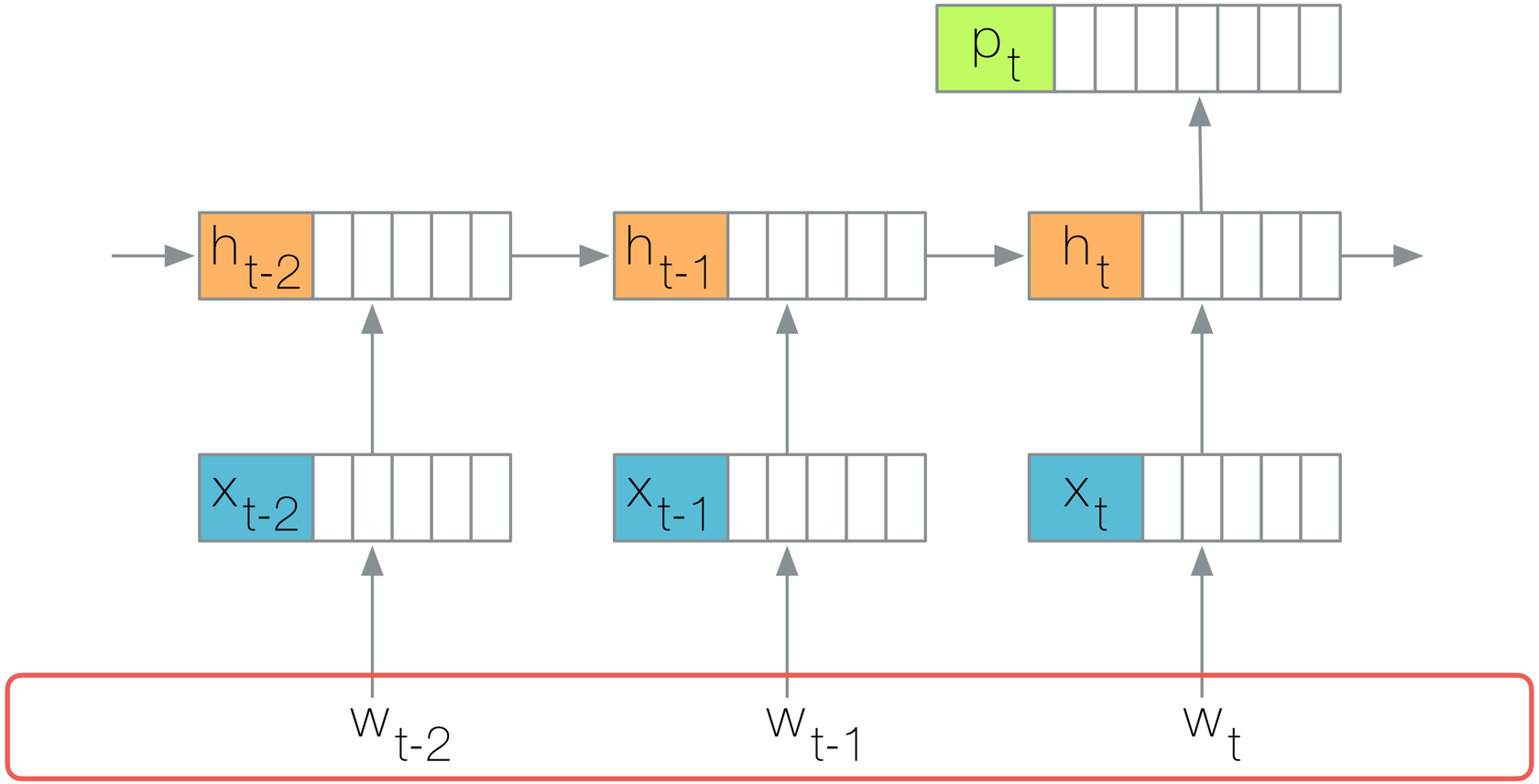
Image captioning

RNNs

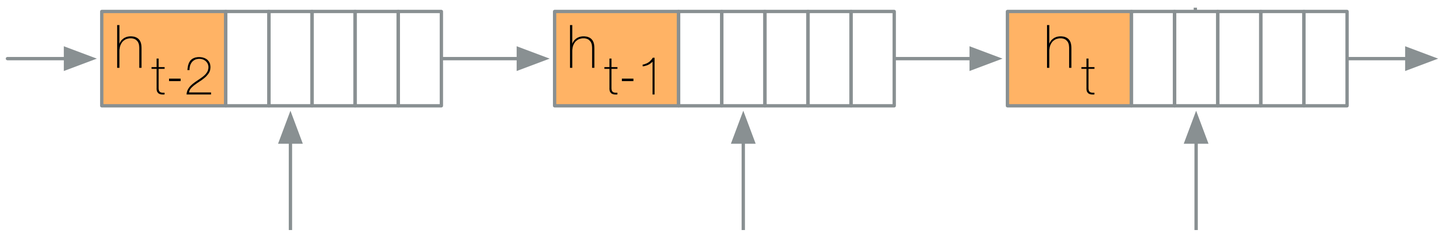
...

Music generation

Speech recognition


p_t
h_{t-2}
h_{t-1}
h_t
x_{t-2}
x_{t-1}
x_t
w_{t-2}
w_{t-1}
w_t

Why are RNNs so effective at capturing the history of words?



What does an RNN capture in its hidden states?

Visual Analysis Tool

Formulate Hypothesis - What properties do groups of hidden states learn to capture?

Refine Hypothesis - What textual similarities do the hidden states represent?

Compare models and datasets - to allow early generalization about the insights the representations provide

Simple Language with Memory

alphabet: ( ) 0 1 2 3 4

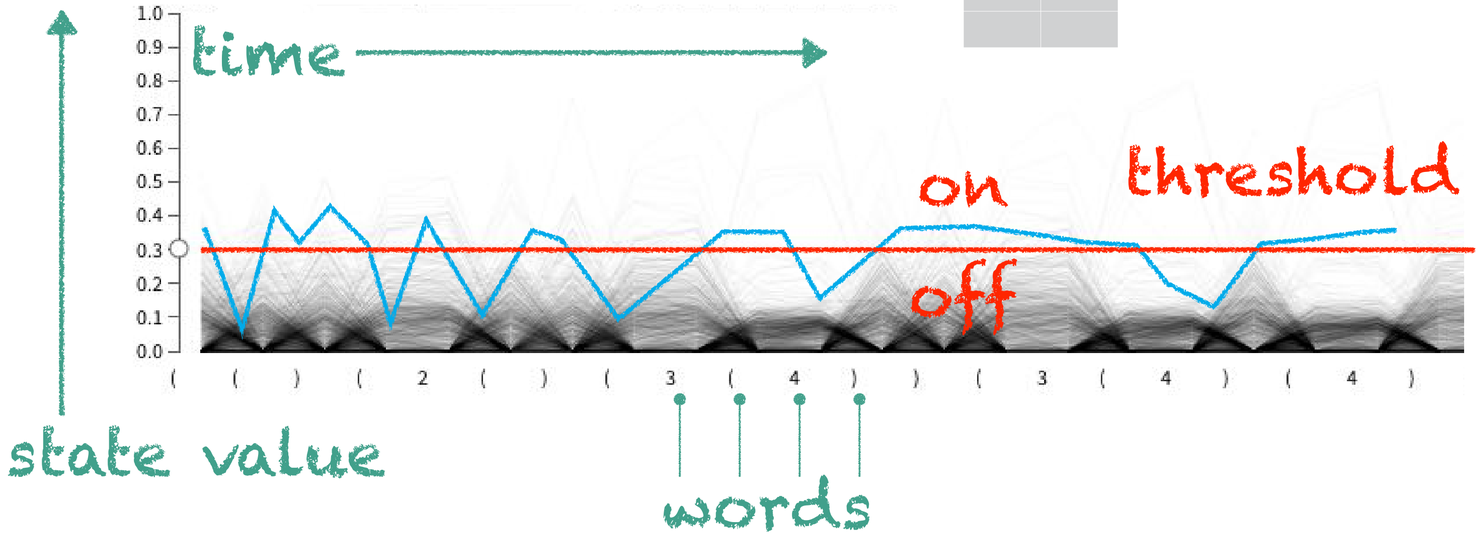
corpus: (1 (2) ()) 0 (((3)) 1)

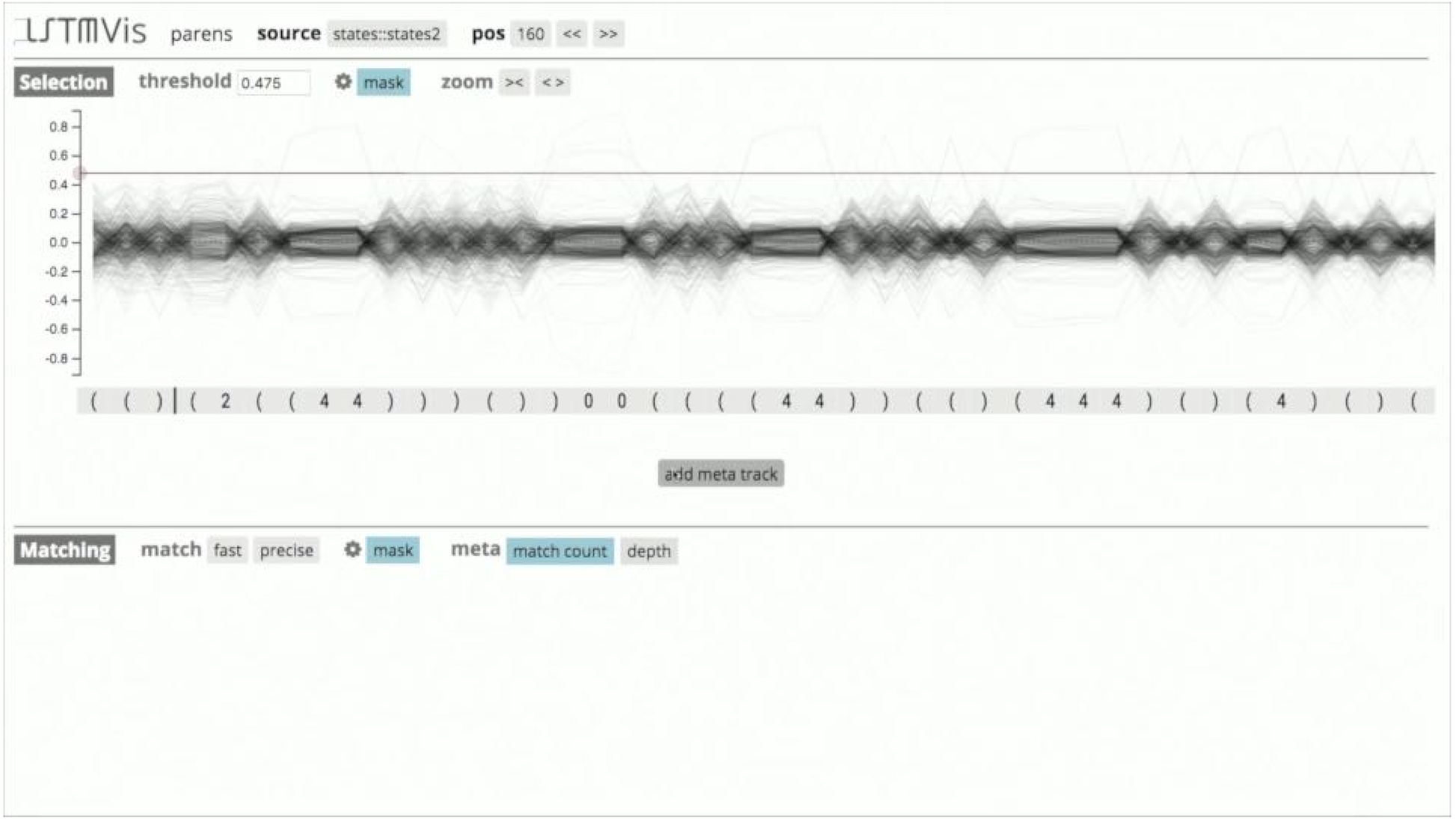
An opening parenthesis increases nesting level; a closing parenthesis decreases it

Randomly, numbers are inserted depicting the nesting level

$h_{i\ldots j}$

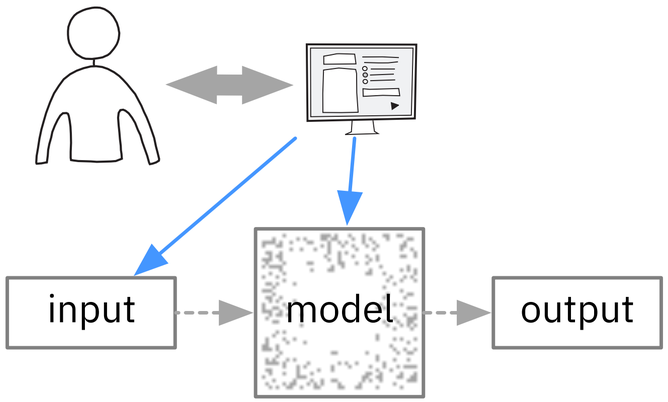
h_{t-1}	h_t
0.12	-0.4
0.01	0.07
0.1	0.3
…	




LSTMVis
parens
source
states::states2
pos
160
<<
>>
Selection
threshold
0.475
mask
zoom
><
< >
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
(( ) (2 ((4 4))) ( )) 0 0 ((((4 4)) (( ) (4 4 4) ( ) (4) ( ) (
add meta track
Matching
match
fast
precise
mask
meta
match count
depth

http://lstm.seas.harvard.edu/

Act 2: Interacting



Seq2Seq-Vis: A visual debugging tool for sequence-to-sequence models

Hendrik Strobelt, Sebastian Gehrmann, Michael Behrisch,
Adam Perer, Hanspeter Pfister, Alexander Rush

[TVCG 2018]

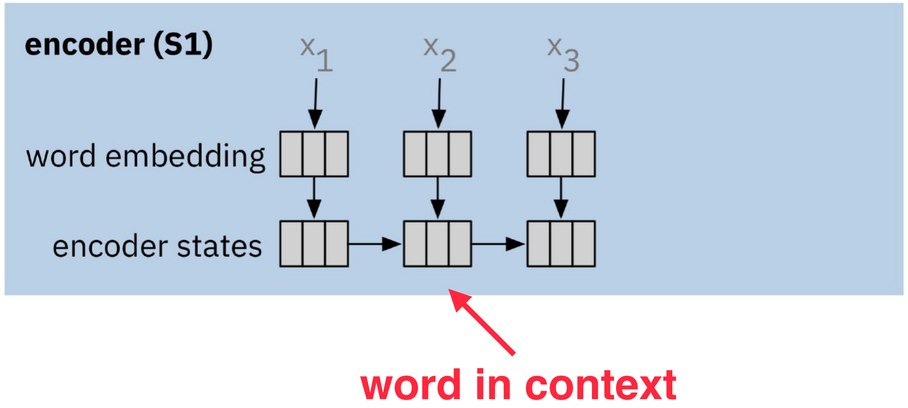
Harvard John A. Paulson
School of Engineering
and Applied Sciences

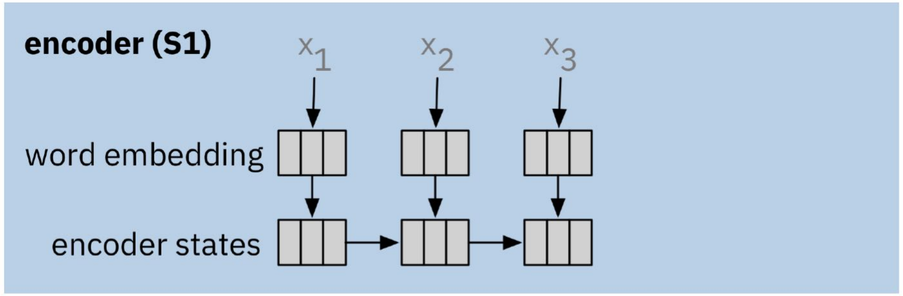
CORNELL TECH

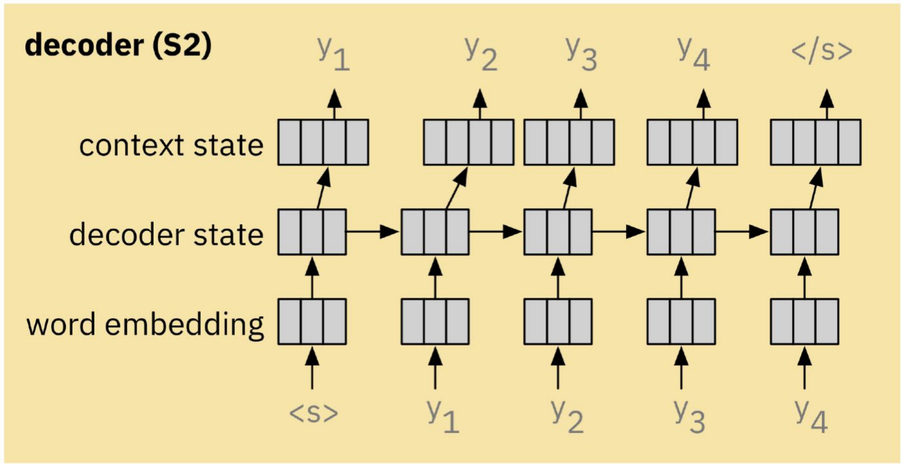
IBM **Research AI**
Visual AI Lab

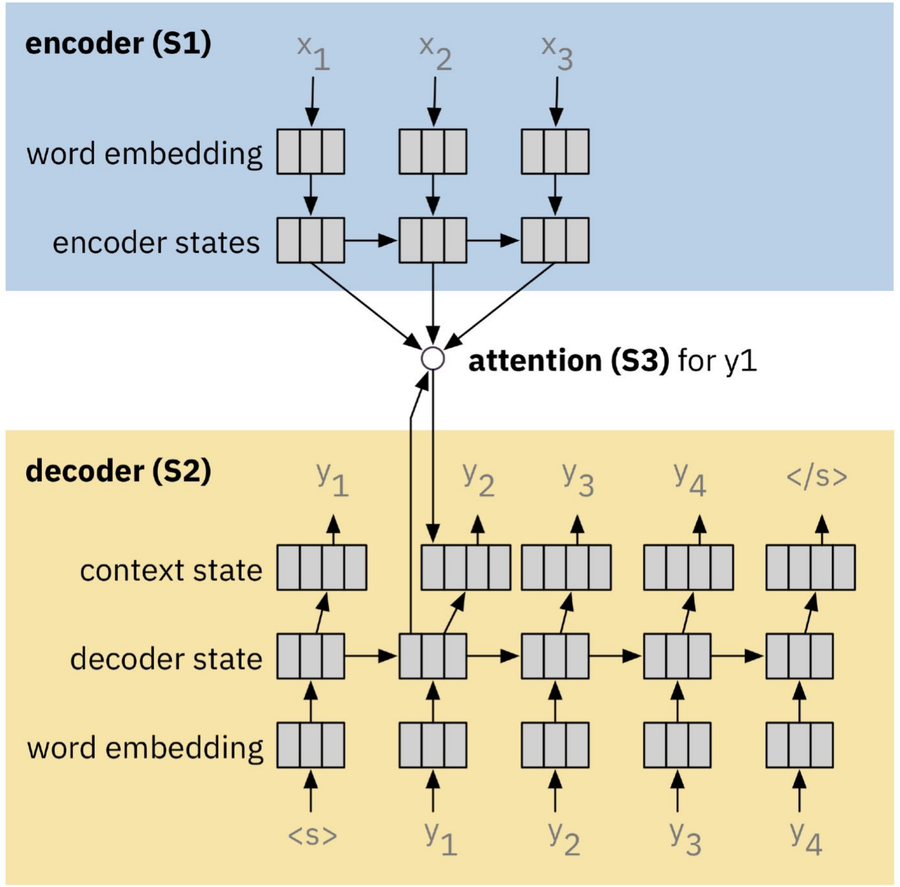
Seq2seq models can learn to transform an **arbitrary length input sequence** into an **arbitrary length output sequence** and are state-of-art for tasks like …

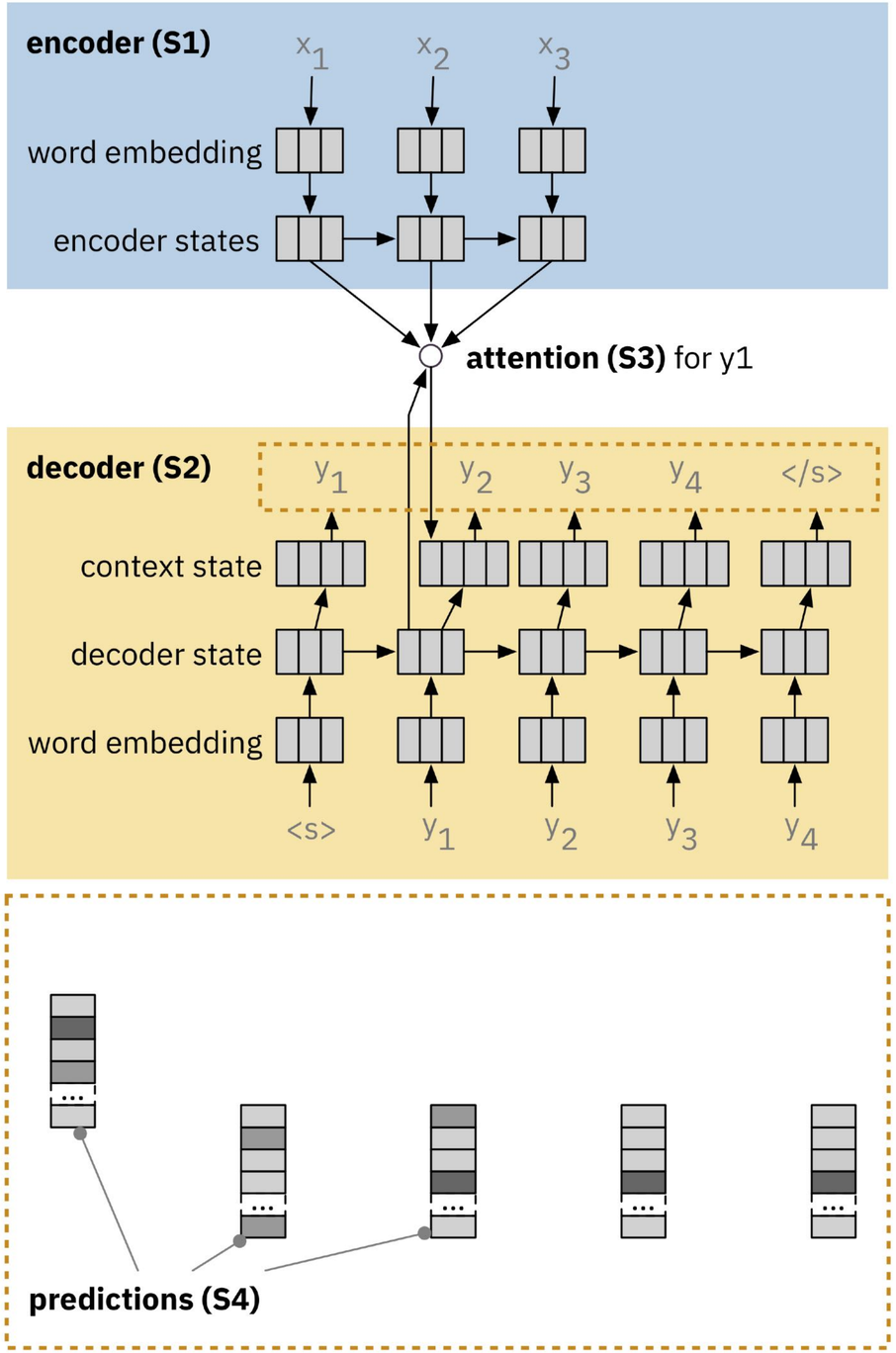
- Machine translation
- Natural language generation
- Question answering
- Speech recognition
- Image captioning
- Summarization
- …

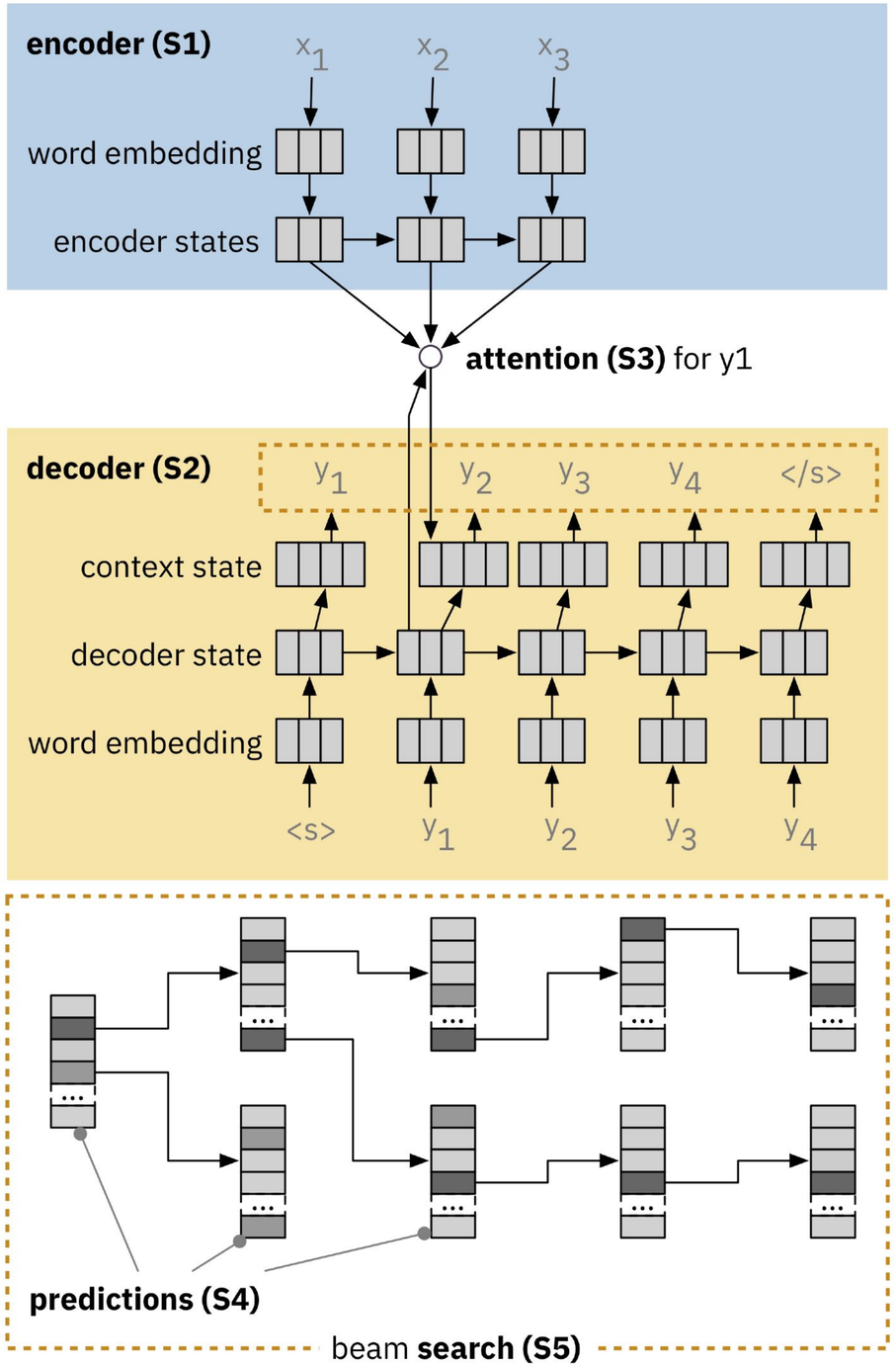

encoder (S1)
x1
x2
x3
word embedding
encoder states
word in context

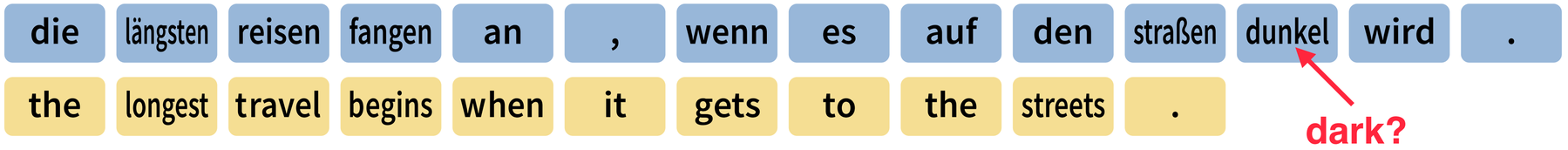

encoder (S1)
x1
x2
x3
word embedding
encoder states


decoder (S2)
y1
y2
y3
y4
</s>
context state
decoder state
word embedding
<s>
y1
y2
y3
y4


encoder (S1)
x1
x2
x3
word embedding
encoder states
attention (S3) for y1
decoder (S2)
y1
y2
y3
y4
</s>
context state
decoder state
word embedding
<s>
y1
y2
y3
y4

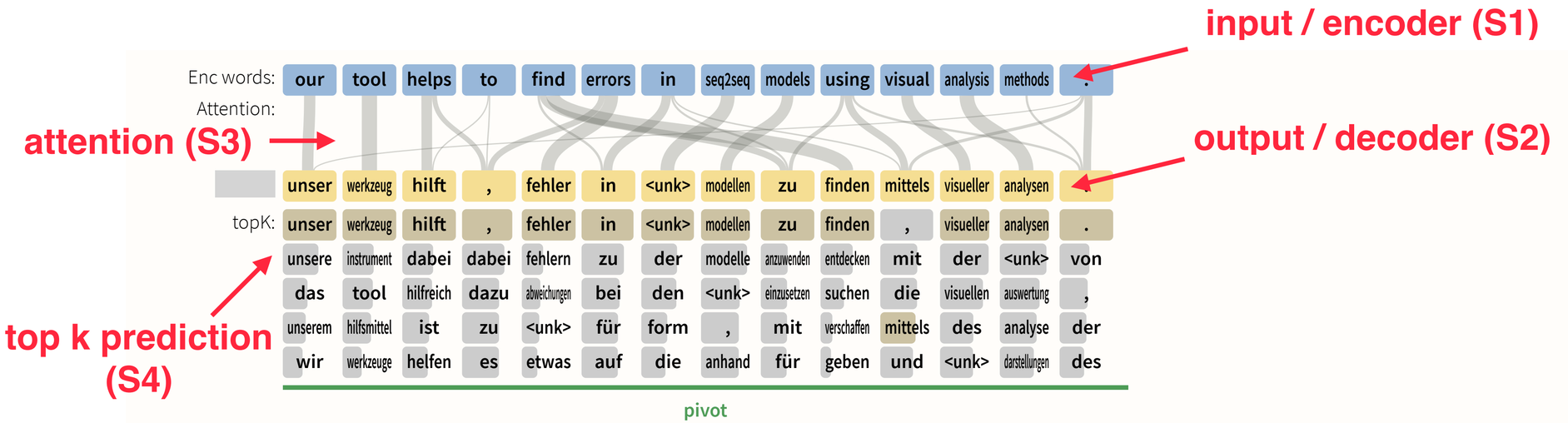

encoder (S1)
x1
x2
x3
word embedding
encoder states
attention (S3) for y1
decoder (S2)
y1
y2
y3
y4
</s>
context state
decoder state
word embedding
<s>
y1
y2
y3
y4
predictions (S4)

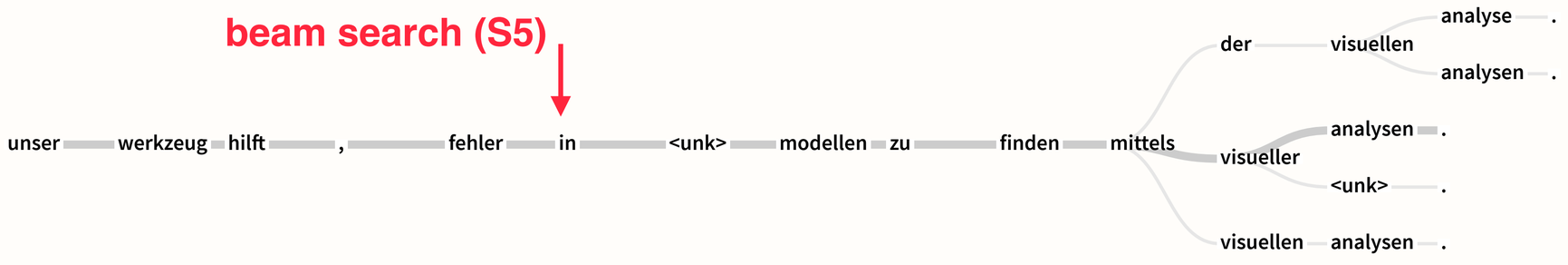

encoder (S1)
x1
x2
x3
word embedding
encoder states
attention (S3) for y1
decoder (S2)
y1
y2
y3
y4
</s>
context state
decoder state
word embedding
<s>
y1
y2
y3
y4
predictions (S4)
beam search (S5)


die längsten reisen fangen an , wenn es auf den straßen dunkel wird .
the longest travel begins when it gets to the streets .
dark?

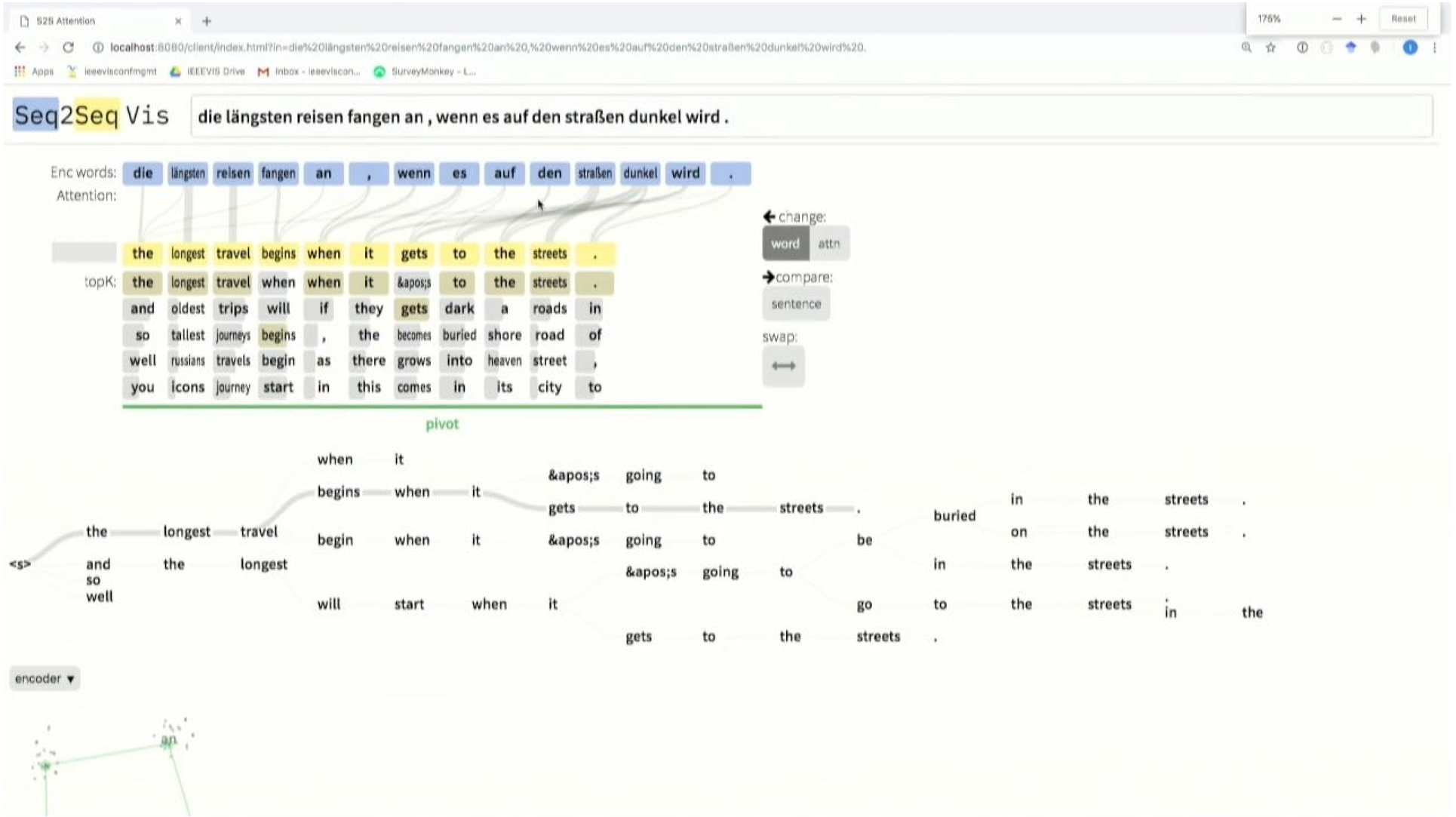
How to identify what went wrong?

Examine Model Decisions

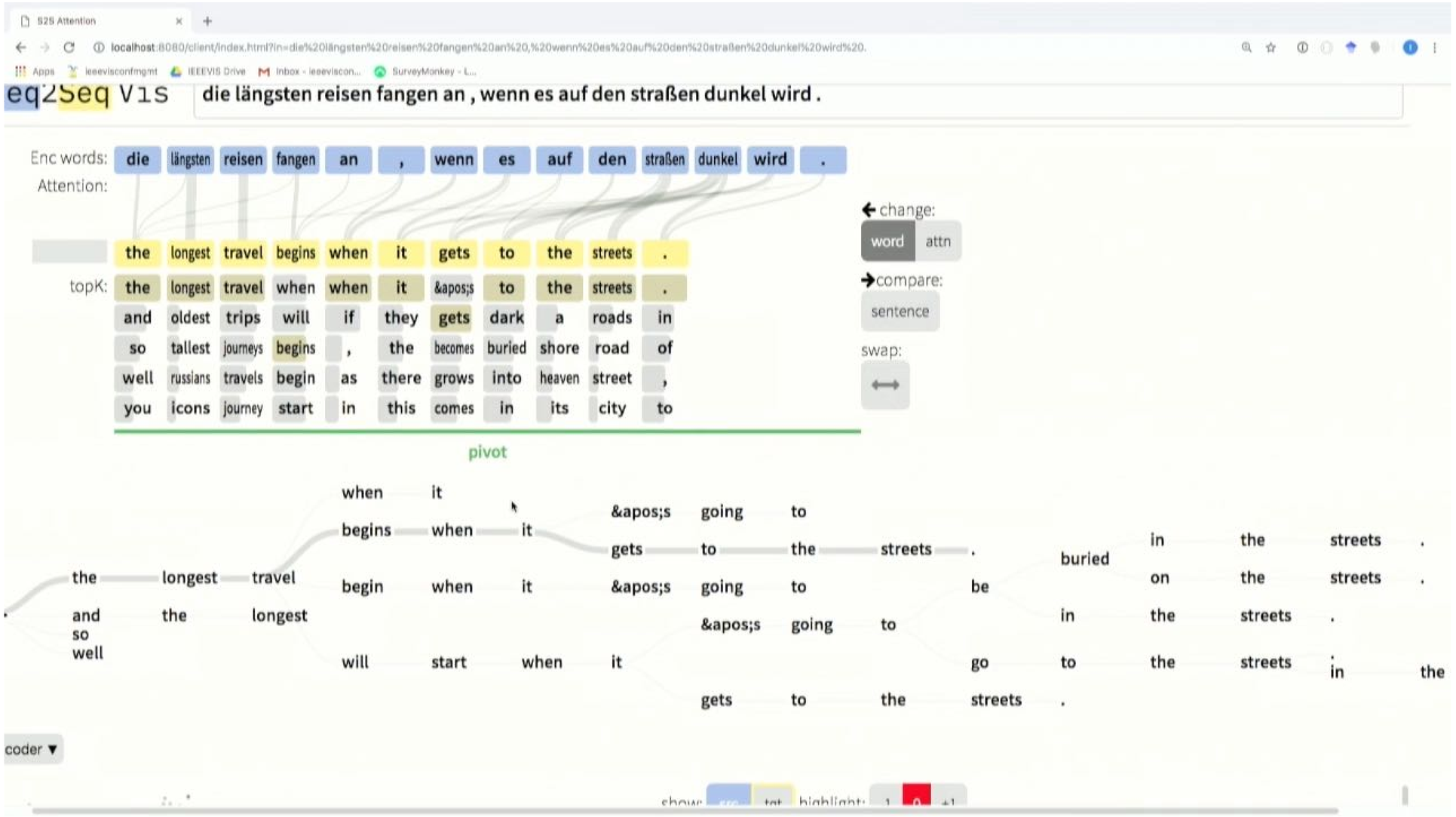




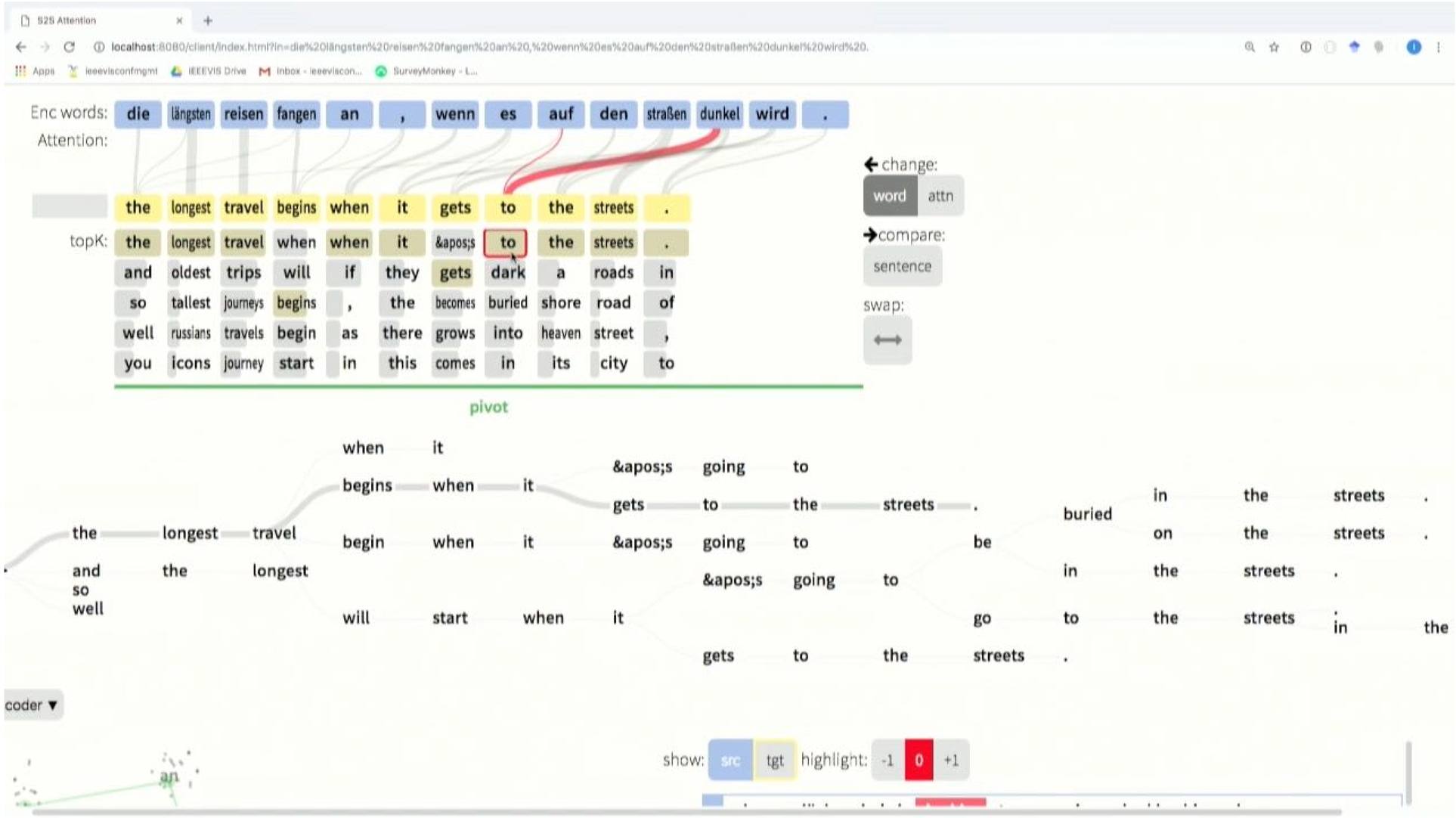
Check Encoder



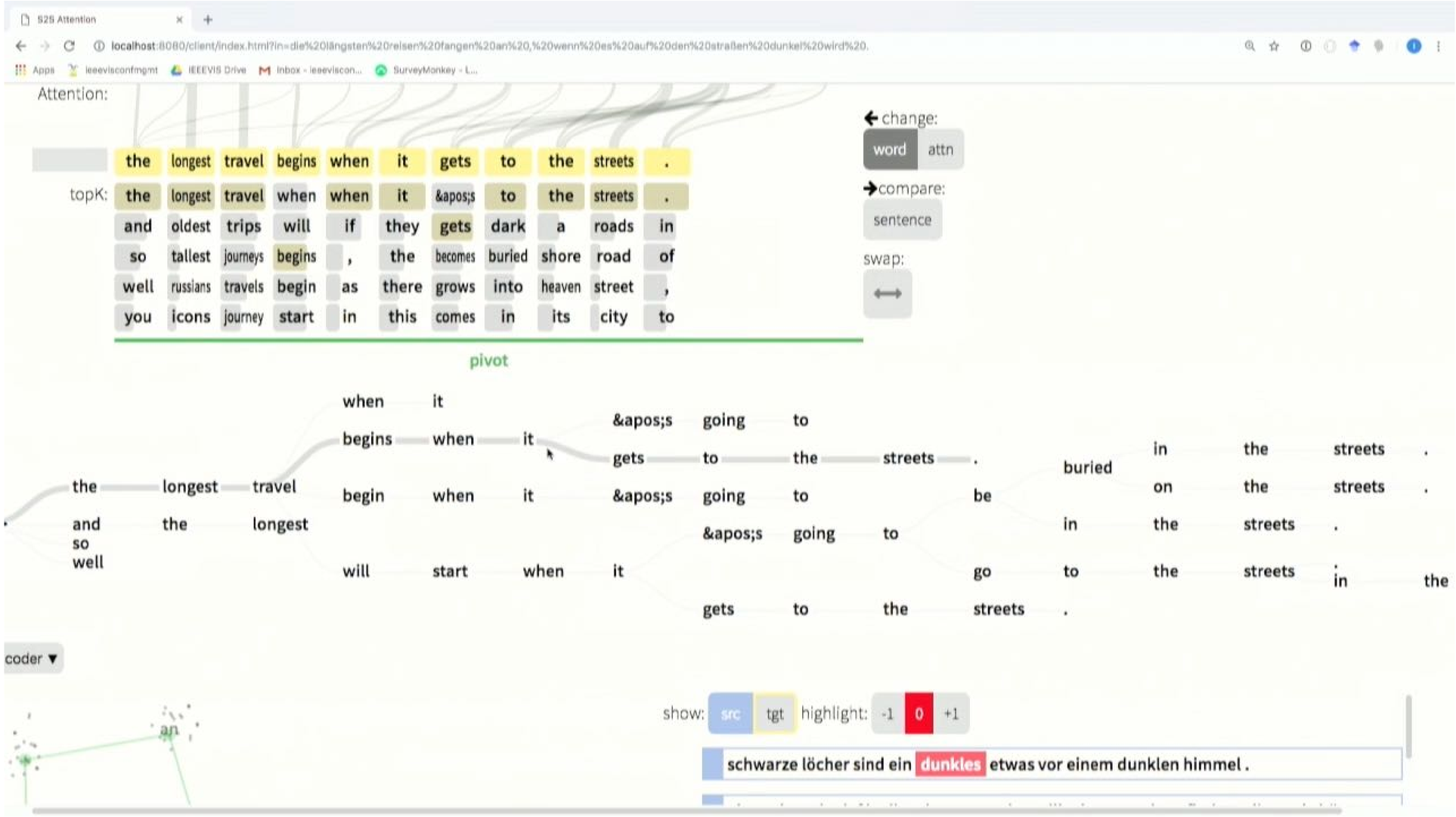
Check Attention



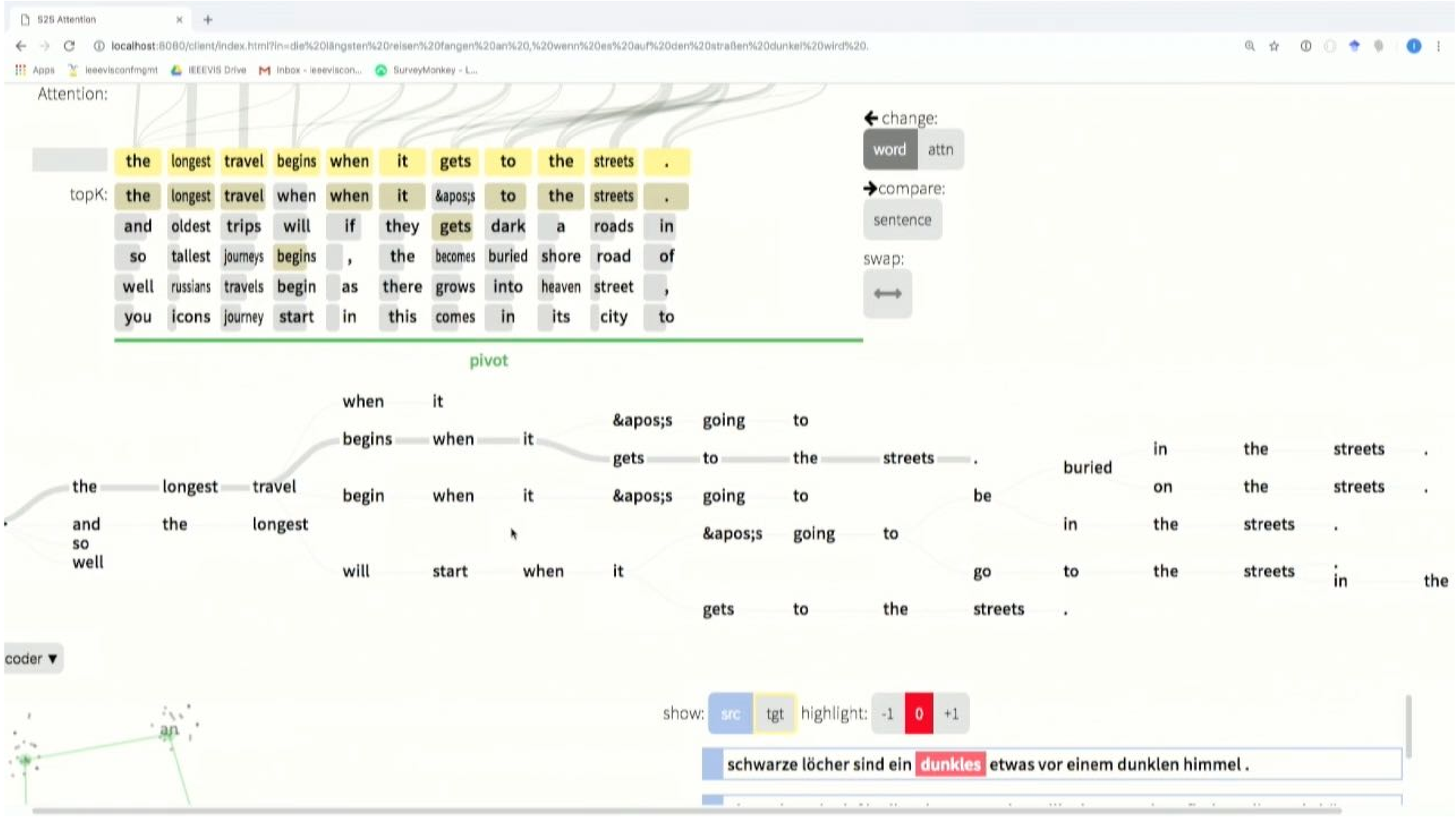
Check Prediction Model



Check Beam Search

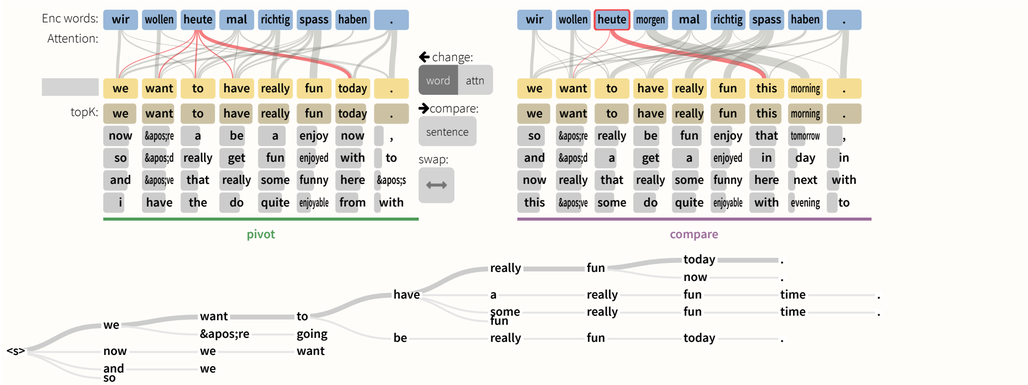


What-If Testing

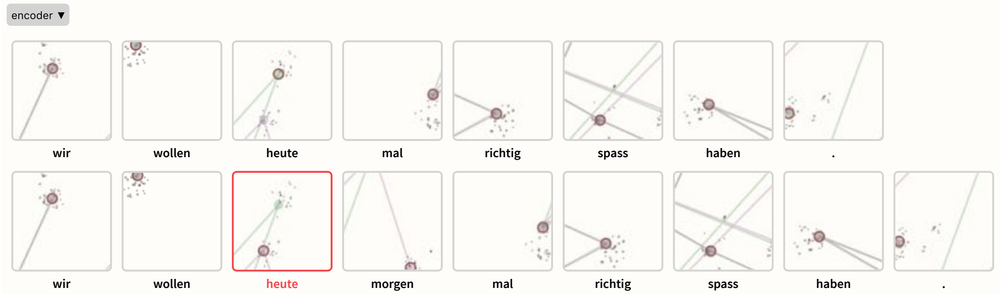


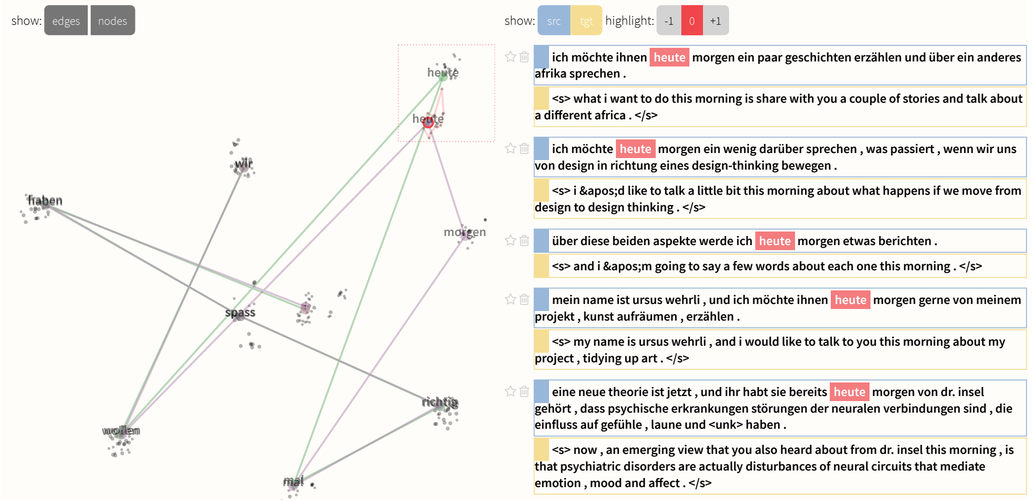
http://seq2seq-vis.io/

Translation View (a)

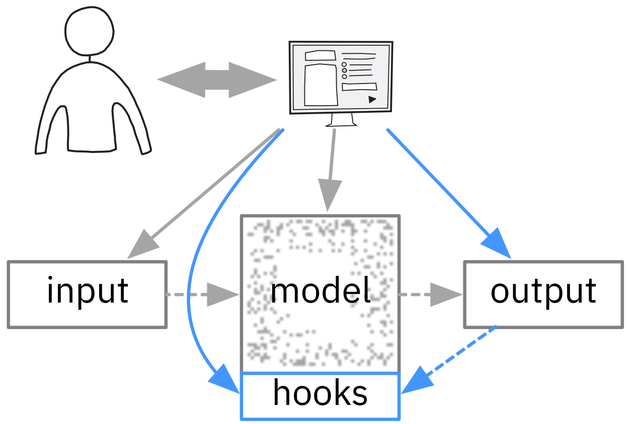


Neighborhood View (b)



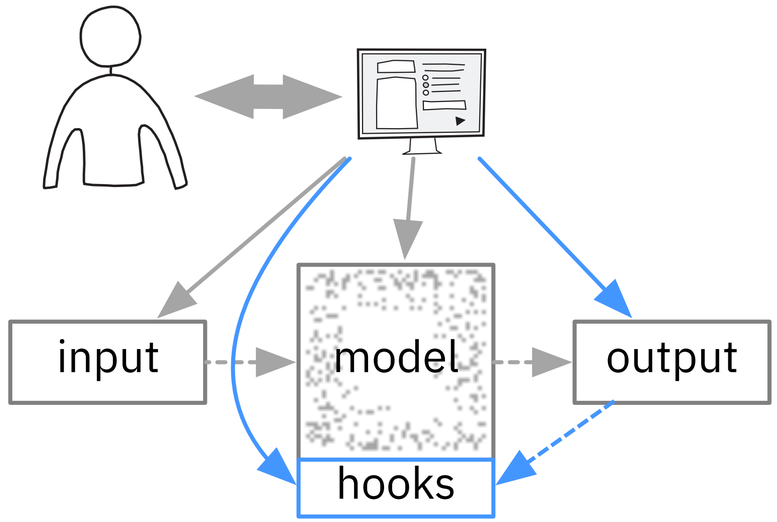


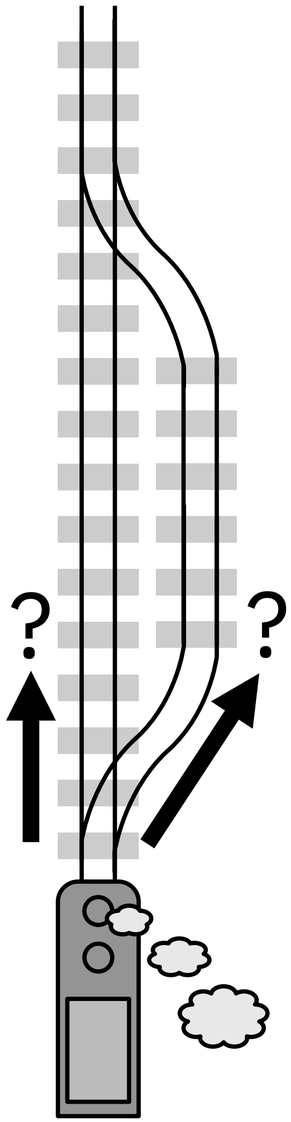
Act 3: Collaborating



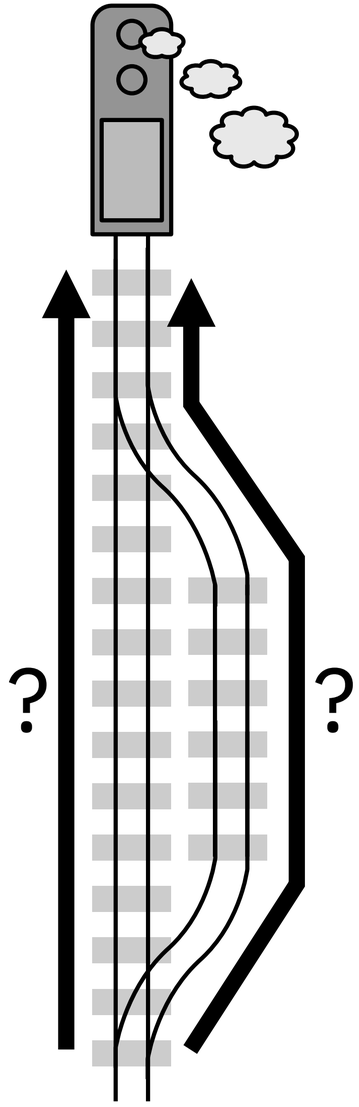
Model “Hooks”

- Forward Path: Prediction
- Backward Path: Examine and correct model decisions interactively

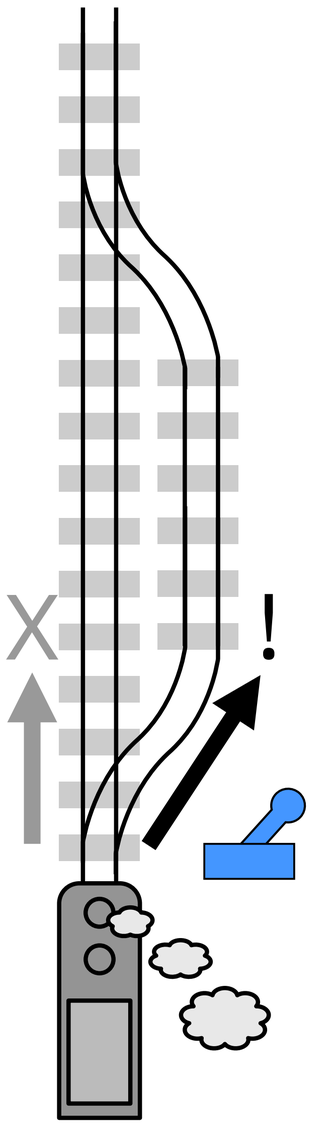




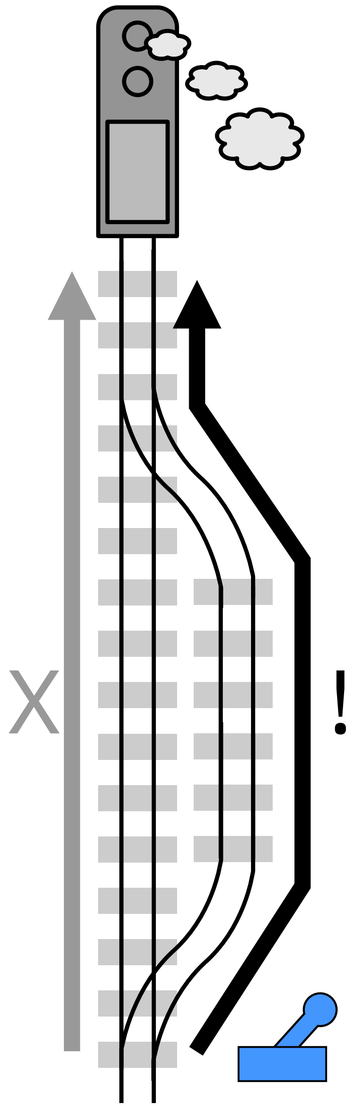
Forward



Backward

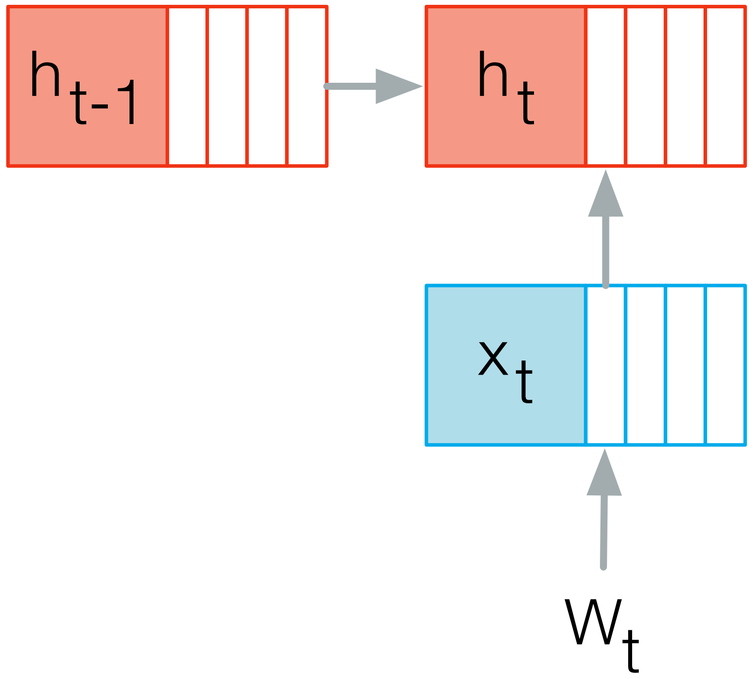


Forward

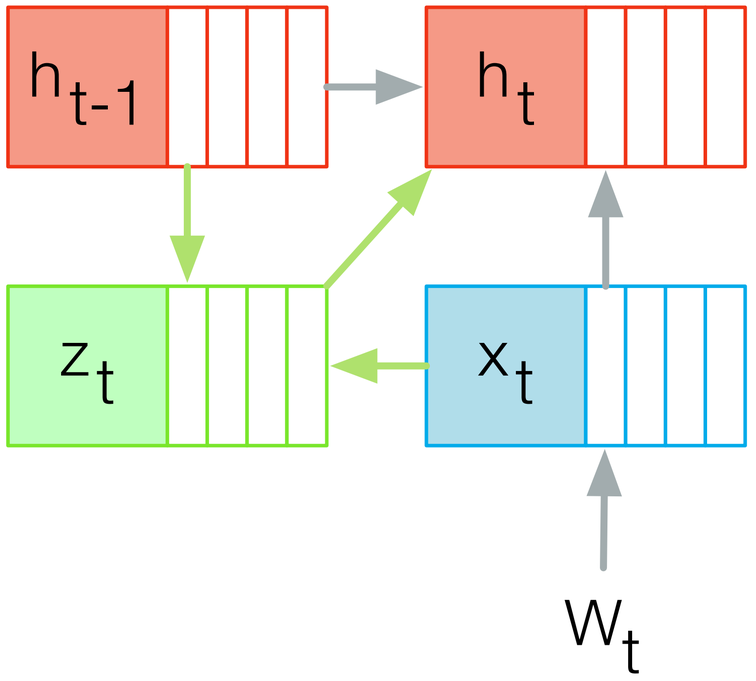


Backward

Traditional Prediction Model



Latent-Variable Model



Visual Interaction with Deep Learning Models through Collaborative Semantic Inference

Sebastian Gehrmann, Hendrik Strobelt, Kathryn Hite,
Robert Krueger, Hanspeter Pfister, Alexander Rush

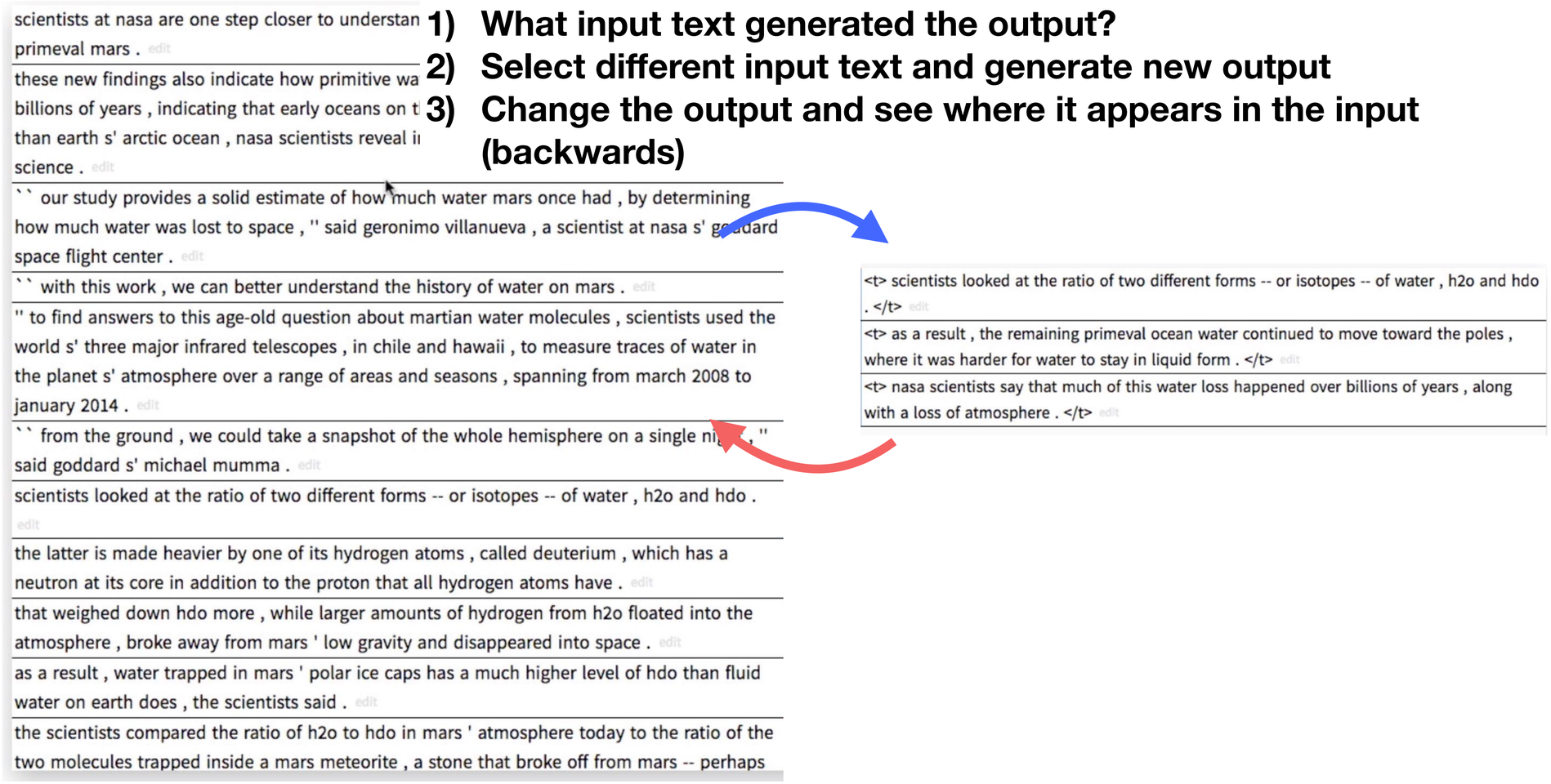
[TVCG 2019]

Harvard John A. Paulson
School of Engineering
and Applied Sciences

CORNELL
TECH

IBM Research AI
Visual AI Lab

Collaborative Summarization



Generate initial 3 sentence summary from all inputs

Input aggregate

scientists at nasa are one step closer to understanding how much water could have existed on primeval mars .

these new findings also indicate how primitive water reservoirs there could have evolved over billions of years , indicating that early oceans on the red planet might have held more water than earth s' arctic ocean , nasa scientists reveal in a study published friday in the journal science .

`` our study provides a solid estimate of how much water mars once had , by determining how much water was lost to space , '' said geronimo villanueva , a scientist at nasa s' goddard space flight center .

`` with this work , we can better understand the history of water on mars .

'' to find answers to this age-old question about martian water molecules , scientists used the world s' three major infrared telescopes , in chile and hawaii , to measure traces of water in the planet s' atmosphere over a range of areas and seasons , spanning from march 2008 to january 2014 .

`` from the ground , we could take a snapshot of the whole hemisphere on a single night , '' said goddard s' michael mumma .

scientists looked at the ratio of two different forms -- or isotopes -- of water , h2o and hdo .

the latter is made heavier by one of its hydrogen atoms , called deuterium , which has a neutron at its core in addition to the proton that all hydrogen atoms have .

that weighed down hdo more , while larger amounts of hydrogen from h2o floated into the atmosphere , broke away from mars ' low gravity and disappeared into space .

add sentence init with 3

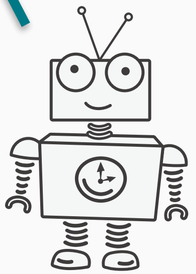
(d) (d)

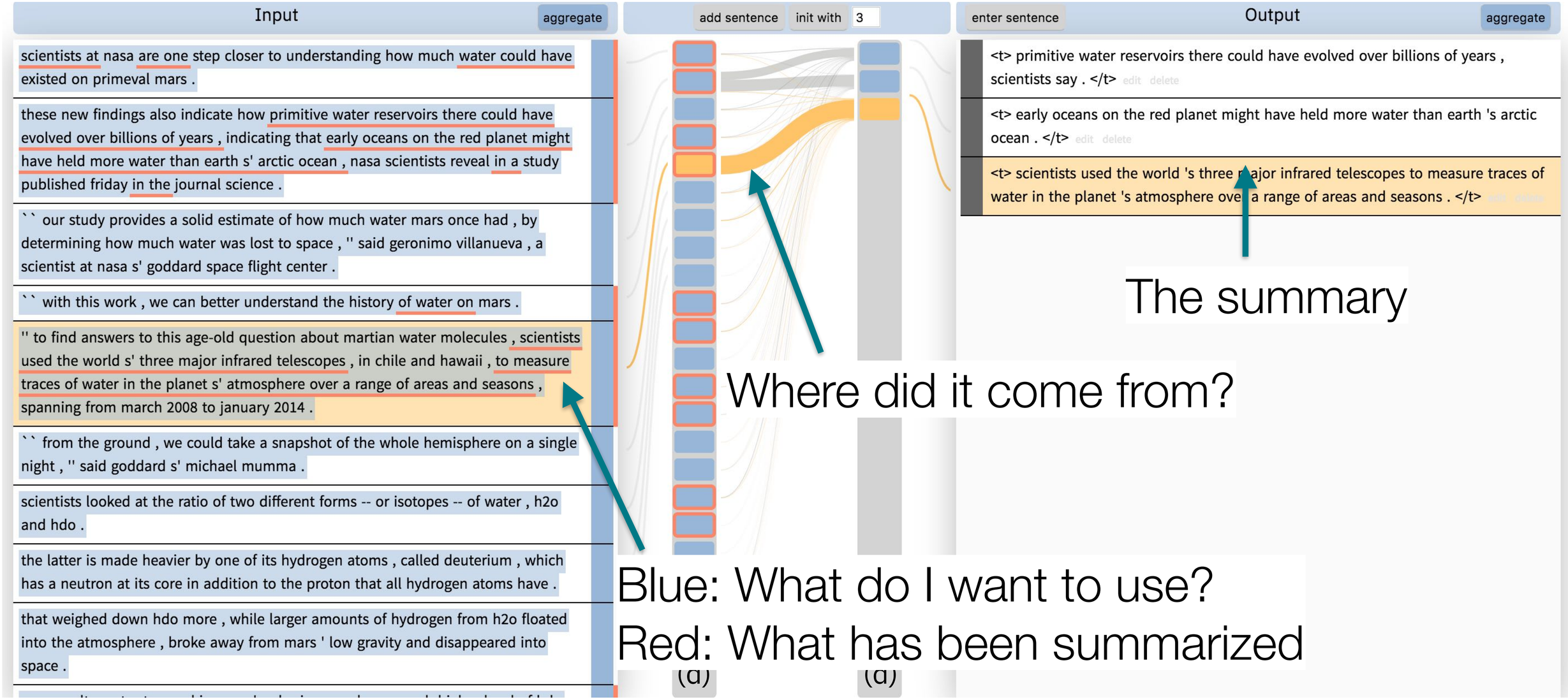
enter sentence Output aggregate

<t> primitive water reservoirs there could have evolved over billions of years , scientists say . </t> edit delete

<t> early oceans on the red planet might have held more water than earth 's arctic ocean . </t> edit delete

<t> scientists used the world 's three major infrared telescopes to measure traces of water in the planet 's atmosphere over a range of areas and seasons . </t> edit delete




Input
aggregate
scientists at nasa are one step closer to understanding how much water could have existed on primeval mars .
these new findings also indicate how primitive water reservoirs there could have evolved over billions of years , indicating that early oceans on the red planet might have held more water than earth s' arctic ocean , nasa scientists reveal in a study published friday in the journal science .
`` our study provides a solid estimate of how much water mars once had , by determining how much water was lost to space , '' said geronimo villanueva , a scientist at nasa s' goddard space flight center .
`` with this work , we can better understand the history of water on mars .
'' to find answers to this age-old question about martian water molecules , scientists used the world s' three major infrared telescopes , in chile and hawaii , to measure traces of water in the planet s' atmosphere over a range of areas and seasons , spanning from march 2008 to january 2014 .
`` from the ground , we could take a snapshot of the whole hemisphere on a single night , '' said goddard s' michael mumma .
scientists looked at the ratio of two different forms -- or isotopes -- of water , h2o and hdo .
the latter is made heavier by one of its hydrogen atoms , called deuterium , which has a neutron at its core in addition to the proton that all hydrogen atoms have .
that weighed down hdo more , while larger amounts of hydrogen from h2o floated into the atmosphere , broke away from mars ' low gravity and disappeared into space .
add sentence
init with
3
enter sentence
Output
aggregate
<t> primitive water reservoirs there could have evolved over billions of years , scientists say . </t>
edit
delete
<t> early oceans on the red planet might have held more water than earth 's arctic ocean . </t>
edit
delete
<t> scientists used the world 's three major infrared telescopes to measure traces of water in the planet 's atmosphere over a range of areas and seasons . </t>
The summary
Where did it come from?
Blue: What do I want to use?
Red: What has been summarized

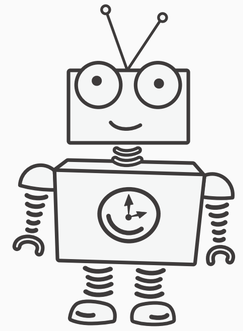
Add a new sentence that starts with “The water is …”

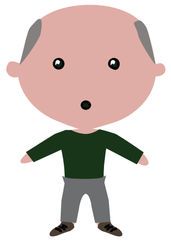
CSI:Summarization

Please choose a .txt file containing the text to summarize: Choose File No file chosen

select: all match none

Input	add sentence init with 3	Output
scientists at nasa are one step closer to understanding how much water could have existed on primeval mars .		<t> primitive water reservoirs there could have evolved over billions of years , scientists say . </t> edit delete
these new findings also indicate how primitive water reservoirs there could have evolved over billions of years , indicating that early oceans on the red planet might have held more water than earth s' arctic ocean , nasa scientists reveal in a study published friday in the journal science .		<t> early oceans on the red planet might have held more water than earth 's arctic ocean . </t> edit delete
`` our study provides a solid estimate of how much water mars once had , by determining how much water was lost to space , '' said geronimo villanueva , a scientist at nasa s' goddard space flight center .		<t> scientists used the world 's three major infrared telescopes to measure traces of water in the planet 's atmosphere over a range of areas and seasons . </t> edit delete
`` with this work , we can better understand the history of water on mars .		<t> water trapped in mars ' polar ice caps has a much higher level of hdo than fluid water on earth . </t> edit delete
'' to find answers to this age-old question about martian water molecules , scientists used the world s' three major infrared telescopes , in chile and hawaii , to measure traces of water in the planet s' atmosphere over a range of areas and seasons , spanning from march 2008 to january 2014 .		
`` from the ground , we could take a snapshot of the whole hemisphere on a single night , '' said goddard s' michael mumma .		
scientists looked at the ratio of two different forms -- or isotopes -- of water , h2o and hdo .		
the latter is made heavier by one of its hydrogen atoms , called deuterium , which has a neutron at its core in addition to the proton that all hydrogen atoms have .		
that weighed down hdo more , while larger amounts of hydrogen from h2o floated into the atmosphere , broke away from mars ' low gravity and disappeared into space .		
as a result , water trapped in mars ' polar ice caps has a much higher level of hdo than fluid water on earth does , the scientists said .		
the scientists compared the ratio of h2o to hdo in mars ' atmosphere today to the ratio of the two molecules trapped inside a mars meteorite , a stone that broke off from mars -- perhaps when an asteroid hit -- and landed on earth some 4.5 billion years ago .		
they were able to determine how much that ratio had changed over time and estimate how much water has disappeared from mars -- about 87 % .		
the findings indicate that the red planet could have had its fair share of blue waters , possibly even yielding an ocean .		
according to nasa , there might have been enough water to cover up to 20 % of mars ' surface .		
that would amount to an ocean proportionally larger than the atlantic on earth .		
`` this ocean had a maximum depth of around 5,000 feet or around one mile deep , '' said villanueva .		
nasa scientists say that much of this water loss happened over billions of years , along with a loss of atmosphere .		
and as the planet s' atmospheric pressure dropped , it was harder for water to stay in liquid form .		
heat also contributed to its evaporation .		
as a result , the remaining primeval ocean water continued to move toward the poles , where it eventually froze .		





I wrote another sentence. What does the summary cover now?

Input | aggregate

from mars -- perhaps when an asteroid hit -- and landed on earth some 4.5 billion years ago .

they were able to determine how much that ratio had changed over time and estimate how much water has disappeared from mars -- about 87 % .

the findings indicate that the red planet could have had its fair share of blue waters , possibly even yielding an ocean .

according to nasa , there might have been enough water to cover up to 20 % of mars ' surface .

that would amount to an ocean proportionally larger than the atlantic on earth .

`` this ocean had a maximum depth of around 5,000 feet or around one mile deep , '' said villanueva .

nasa scientists say that much of this water loss happened over billions of years , along with a loss of atmosphere .

and as the planet s' atmospheric pressure dropped , it was harder for water to stay in liquid form .

heat also contributed to its evaporation .

as a result , the remaining primeval ocean water continued to move toward the poles , where it eventually froze .

`` with mars losing that much water , the planet was very likely wet for a longer period of time than was previously thought , suggesting it might have been habitable for longer , '' said mumma .

cnn s' ben brumfield contributed to this report .

add sentence | init with 3

enter sentence | Output | aggregate

<t> primitive water reservoirs there could have evolved over billions of years , scientists say . </t> edit delete

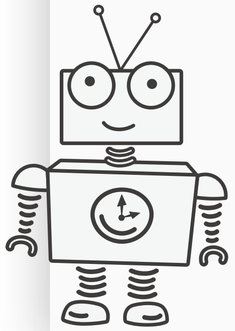
<t> early oceans on the red planet might have held more water than earth 's arctic ocean . </t> edit delete

<t> scientists used the world 's three major infrared telescopes to measure traces of water in the planet 's atmosphere over a range of areas and seasons . </t> edit delete

<t> water trapped in mars ' polar ice caps has a much higher level of hdo than fluid water on earth . </t> edit delete

the water was very likely wet for a longer period of time than was previously thought

<t> `` this ocean had a maximum depth of around 5,000 feet or around one mile deep , '' nasa scientists say . </t> edit delete



Lessons Learned

- Users can have semantically meaningful interactions with the model (CSI: Collaborative Semantic Inference).
- It is easy to over-constrain the text generation; SOTA ML models cannot handle this, yet.
- Need a higher-level interface to specify the constraints. Idea: use a constraint graph to specify model hooks.

GenNI: Human-AI Collaboration for Data-Backed Text Generation

Hendrik Strobelt, Jambay Kinley, Robert Krueger,
Johanna Beyer, Hanspeter Pfister, Alexander M. Rush

[TVCG 2021]

Harvard John A. Paulson
School of Engineering
and Applied Sciences

CORNELL
TECH

IBM Research AI
Visual AI Lab

Table2Text Generation

name	eattype	food	pricerange	area	familyfriendly	near
the mill	pub	english	less than £20	city centre	yes	raja indian cuisine

the mill is a english pub . create example ⇦

name	eattype	customer_rating	near
strada	pub	1 out of 5	all bar one

strada is a pub . create example ⇦

name	eattype	food	pricerange	familyfriendly	near
the wrestlers	restaurant	japanese	more than £30	yes	raja indian cuisine

the wrestlers is a japanese restaurant . create example ⇦

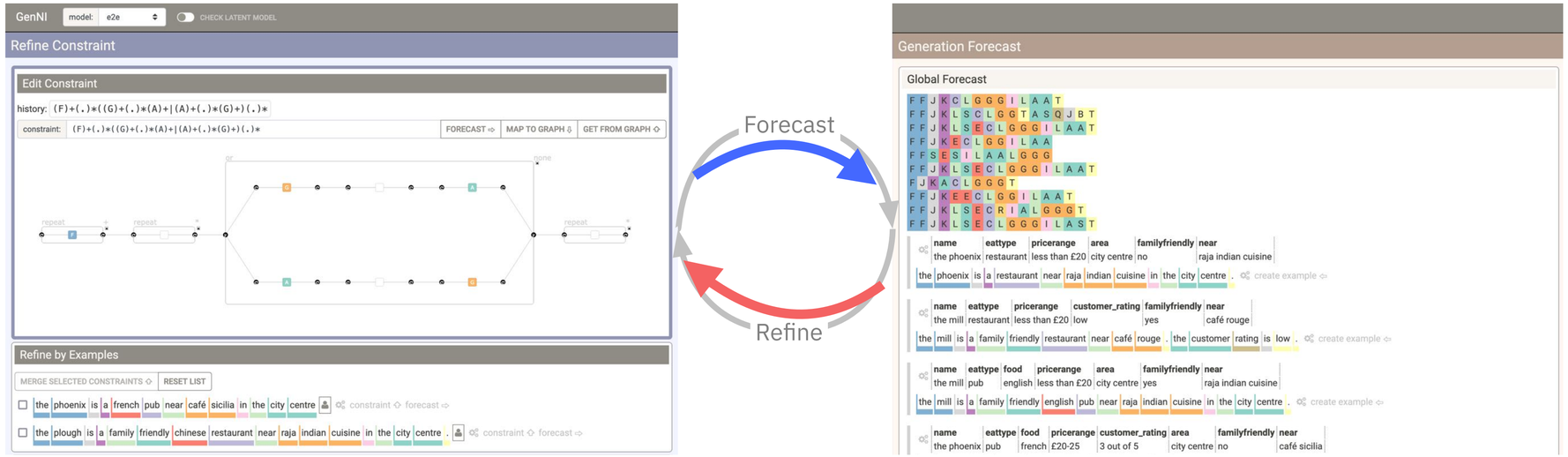
name	eattype	food	pricerange	area	area	familyfriendly	near
the plough	pub	chinese	high	city centre	riverside	yes	raja indian cuisine

the plough is a chinese pub . create example ⇦

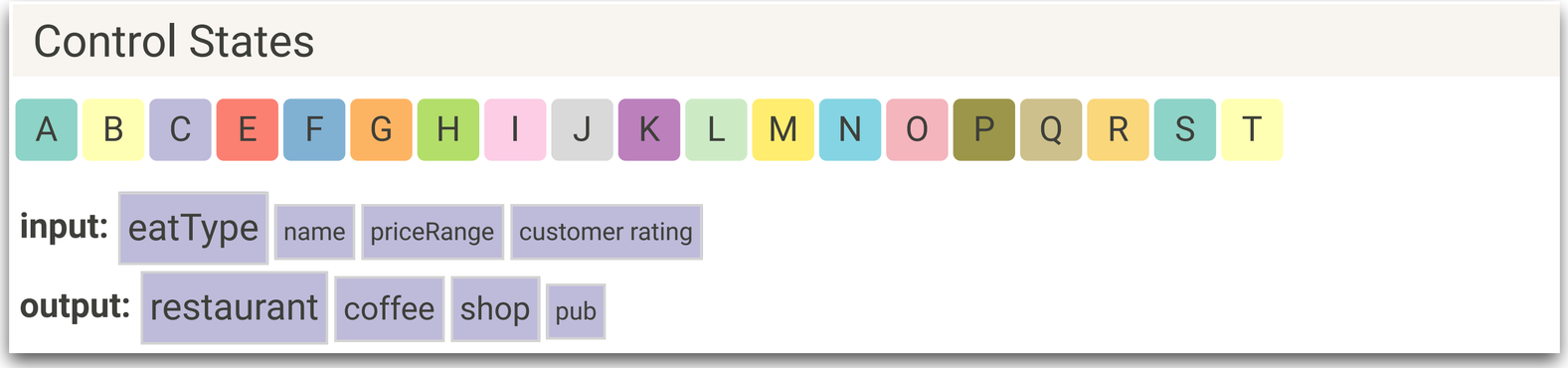
Higher-order Model Hooks

Control states (model hooks)
as constraint graph

Generated output text



Control States (aka Model Hooks)



Model Output

Table

name	eattype	food	pricerange	customer_rating	area	familyfriendly	near
the phoenix	pub	french	£20-25	3 out of 5	city centre	no	café sicilia

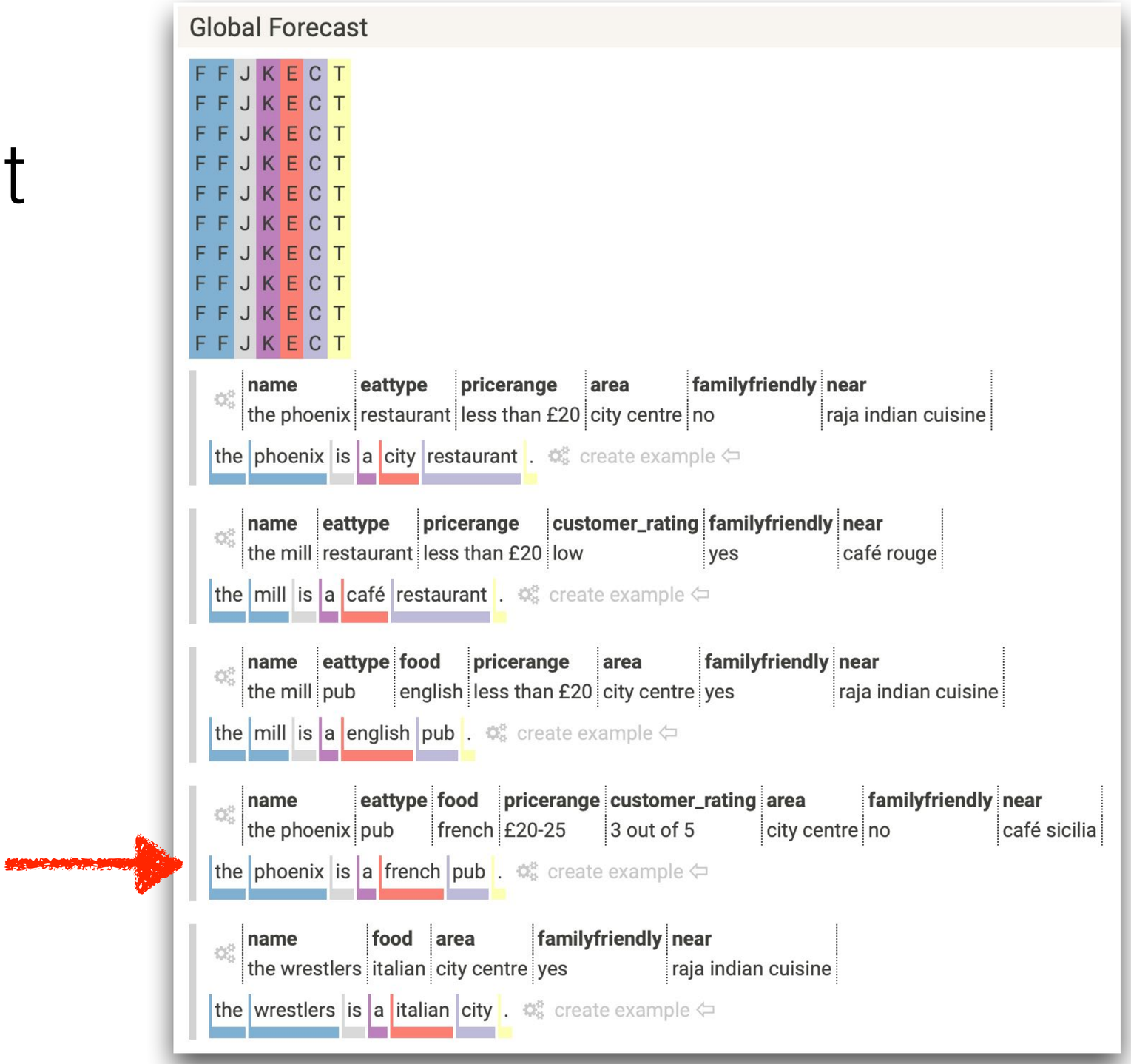
Output

User Correction

name	eattype	food	pricerange	customer_rating	area	familyfriendly	near
the phoenix	pub	french	£20-25	3 out of 5	city centre	no	café sicilia

Alternative Model Outputs (⇄)

Model Output



User Correction

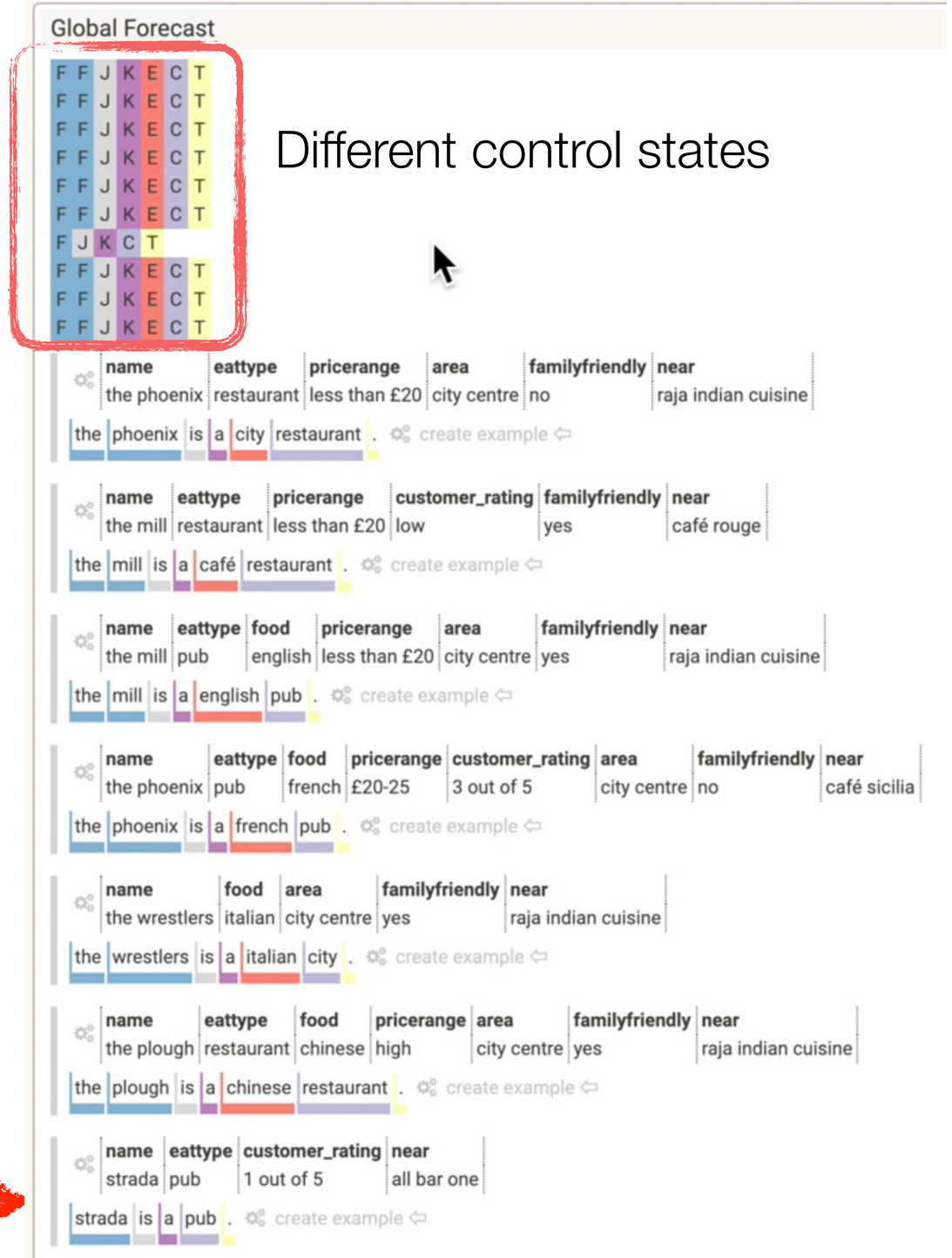
name	eattype	customer_rating	near
strada	pub	1 out of 5	all bar one

strada strada is a all pub . create example ⇦

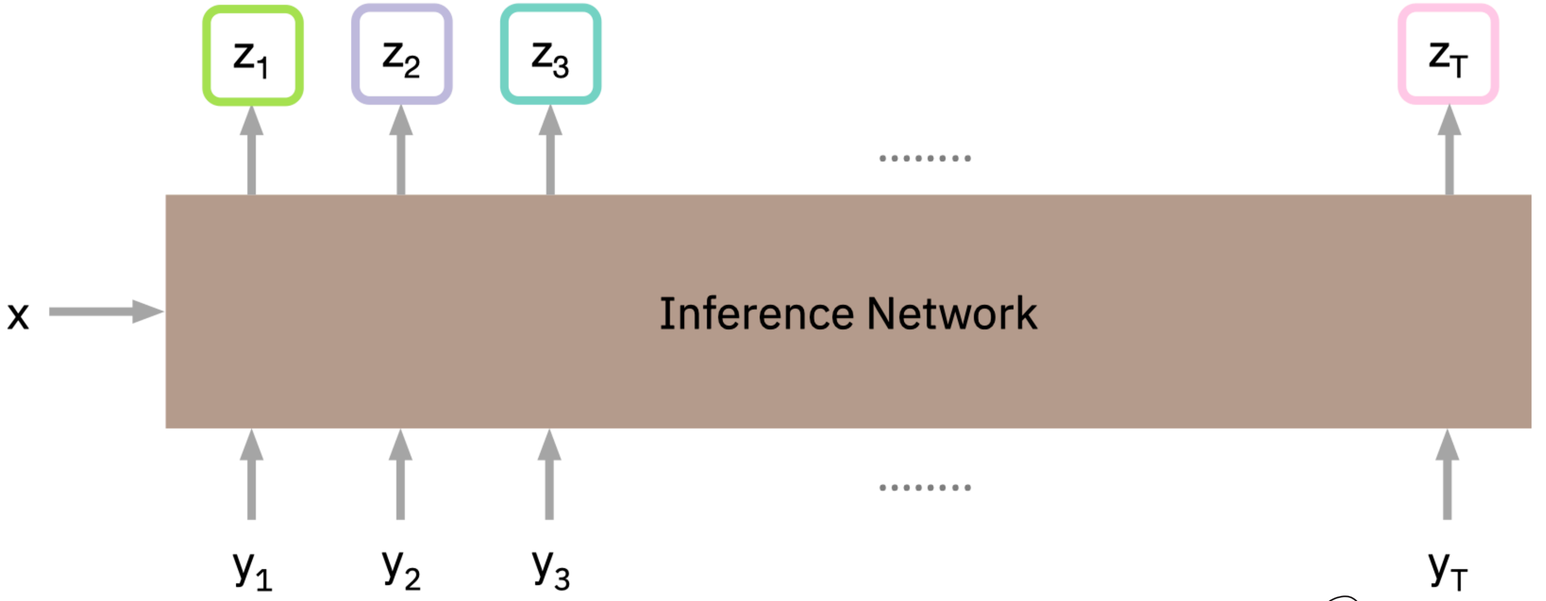
name	eattype	customer_rating	near
strada	pub	1 out of 5	all bar one

strada is a pub . create example ⇦

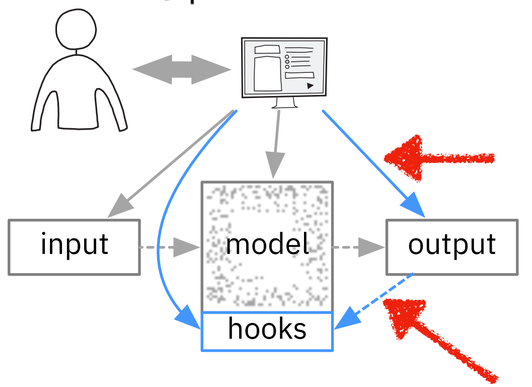
Model Output



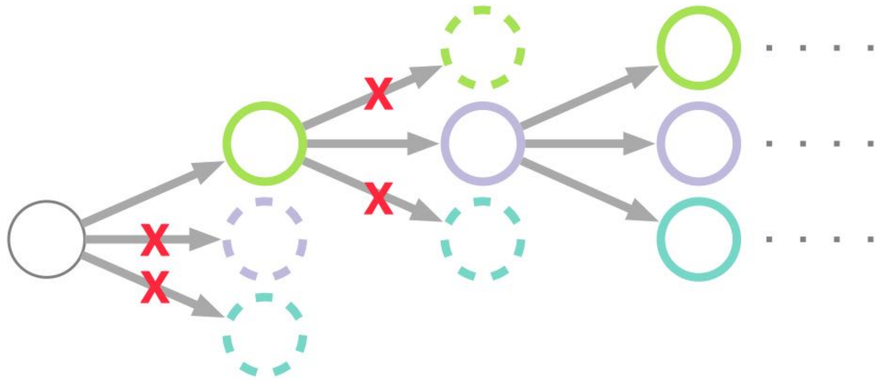
Infer Model Hooks



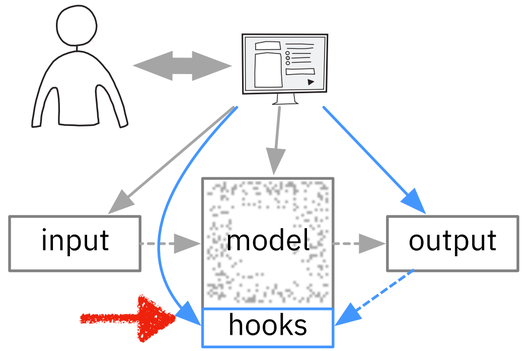
x: table input
y: **user-generated desired** text output
z: **inferred** control states (aka model hooks)



Constrained Generation

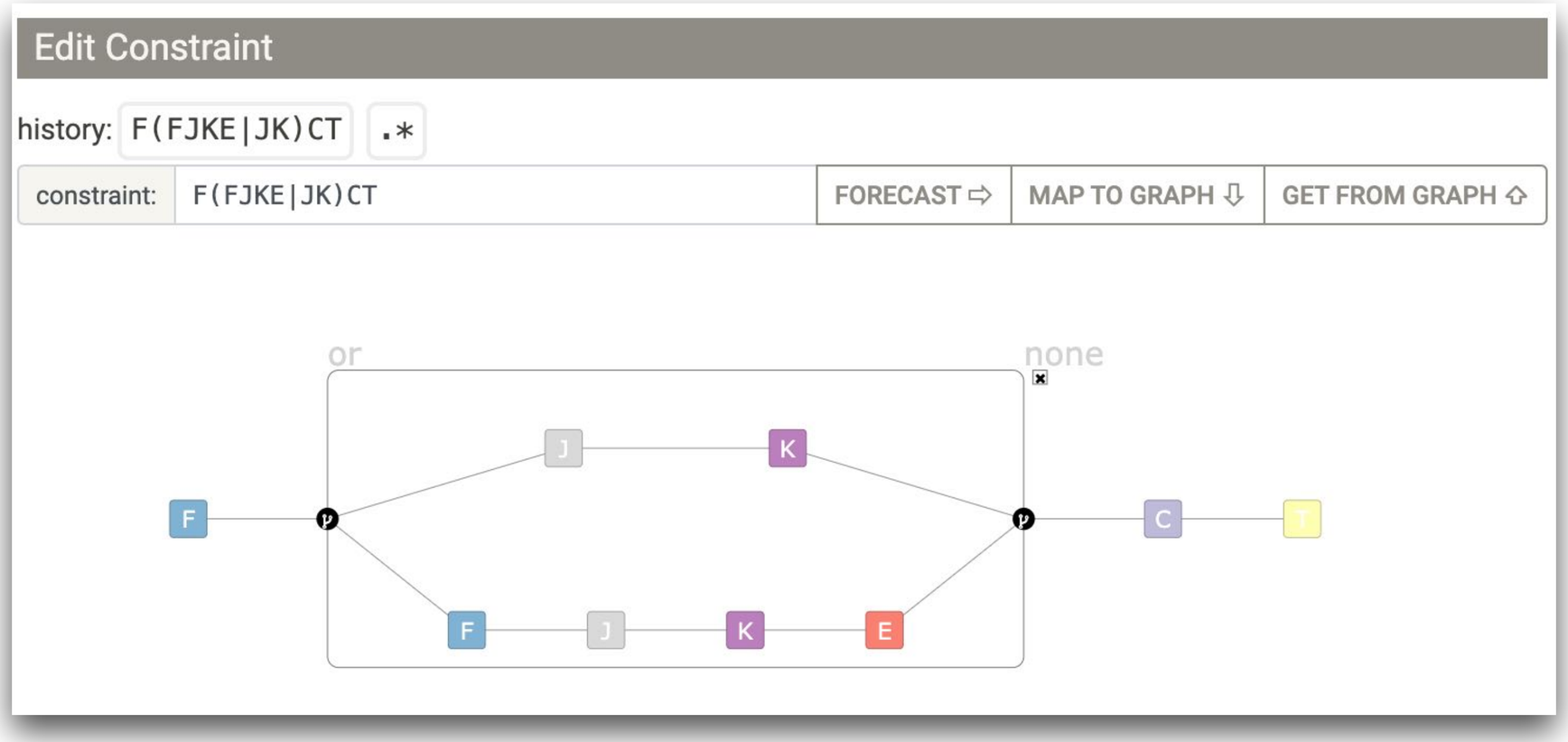


Beam tree with constraints



Constraint Graph Editor

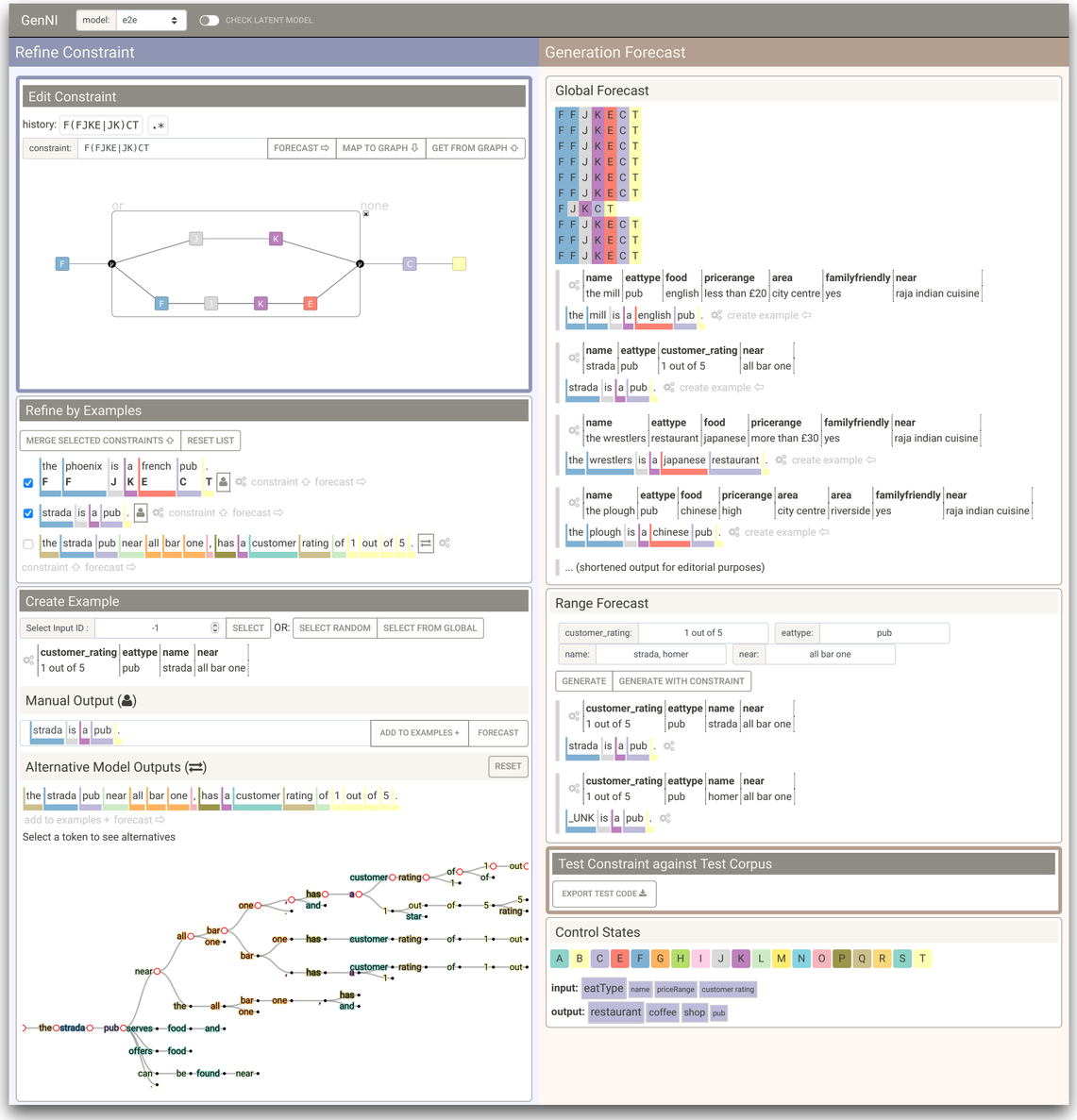
Edit control state constraints in visual RegEx editor



Lessons Learned

- Higher-order constraint graphs combined with direct user input to control the output are intuitive.
- This lesson applies also to other deep learning models, e.g., in computer vision.
- How do we evaluate human-AI collaboration?
- Could evaluate models using both automated and human metrics, e.g., GEM Benchmark (https://gem-benchmark.com)

Soon at https://genni.vizhub.ai



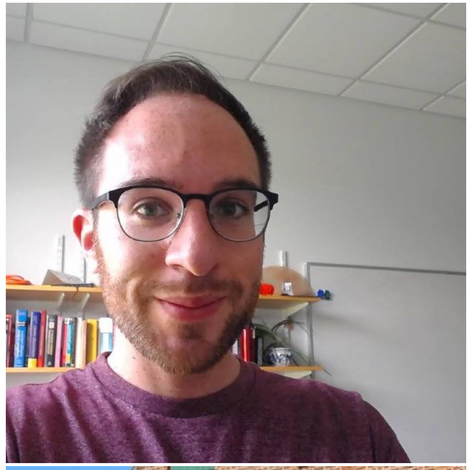
Call for Action

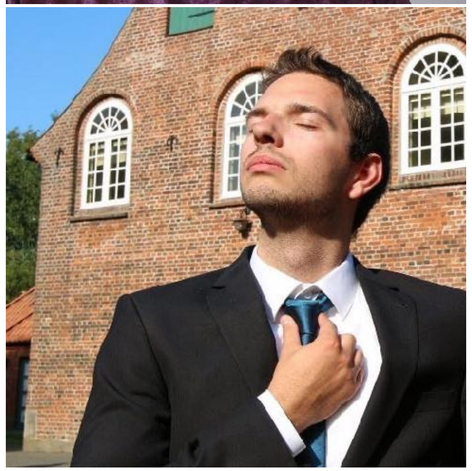
Three Messages

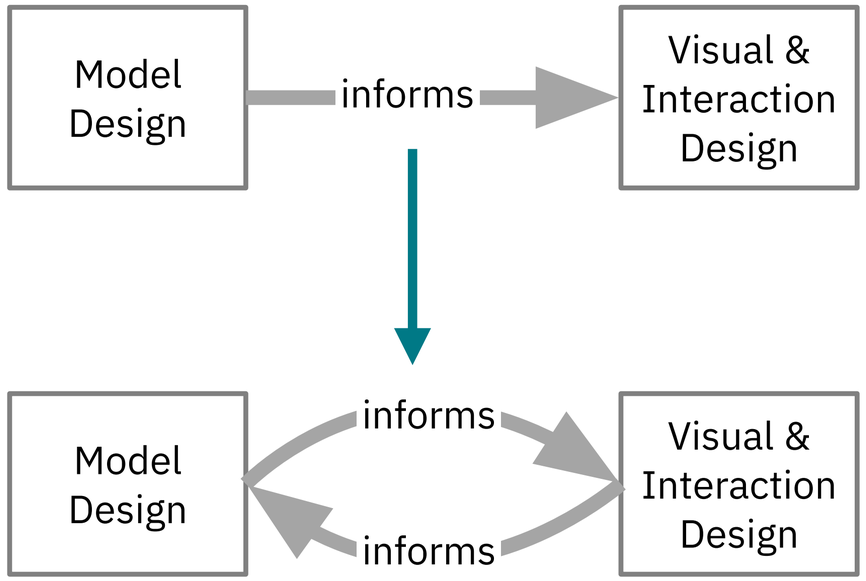
1. Interfaces to deep learning systems need to follow the principles of **human-centered design**.

Co-Design

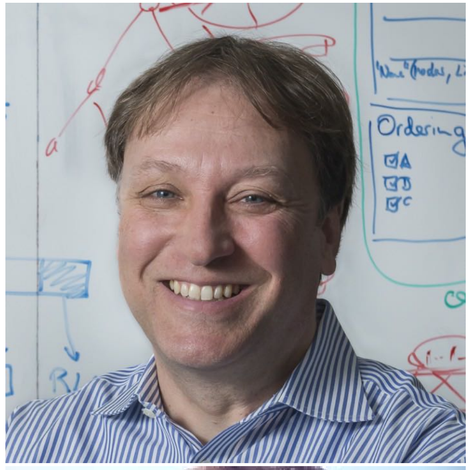
ML Experts







Visualization / HCI Experts



Three Messages

1. Interfaces to deep learning systems need to follow the principles of **human-centered design**.

2. To re-establish the human agency over deep learning systems we introduce the concept of **interactive collaboration**.


Deep Neural Network Model
KUKA

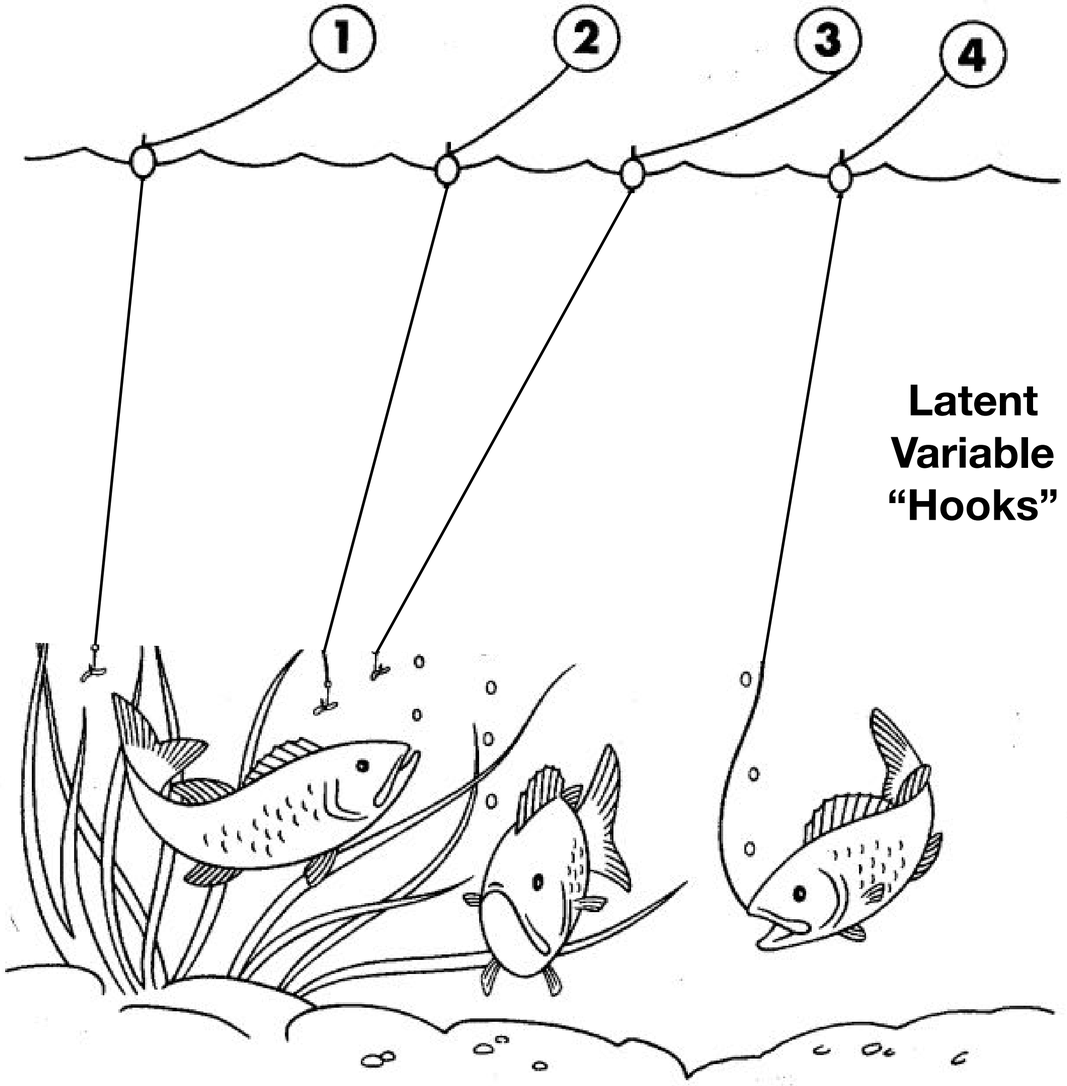
Three Messages

1. Interfaces to deep learning systems need to follow the principles of **human-centered design**.

2. To re-establish the human agency over deep learning systems we introduce the concept of **interactive collaboration**.

3. Neural network models need to be extended to include **intermediate representations (“hooks”)** that can be **understood and acted upon by humans.**

CATCH A FISH



Latent Variable "Hooks"

WHICH LINE WOULD YOU PULL IN TO CATCH THE FISH?

Towards Visually Interactive Neural Probabilistic Models

Hanspeter Pfister
pfister@seas.harvard.edu

Harvard John A. Paulson
School of Engineering
and Applied Sciences

Award III-1901030

Demos, source code, papers are available:

http://lstm.seas.harvard.edu/

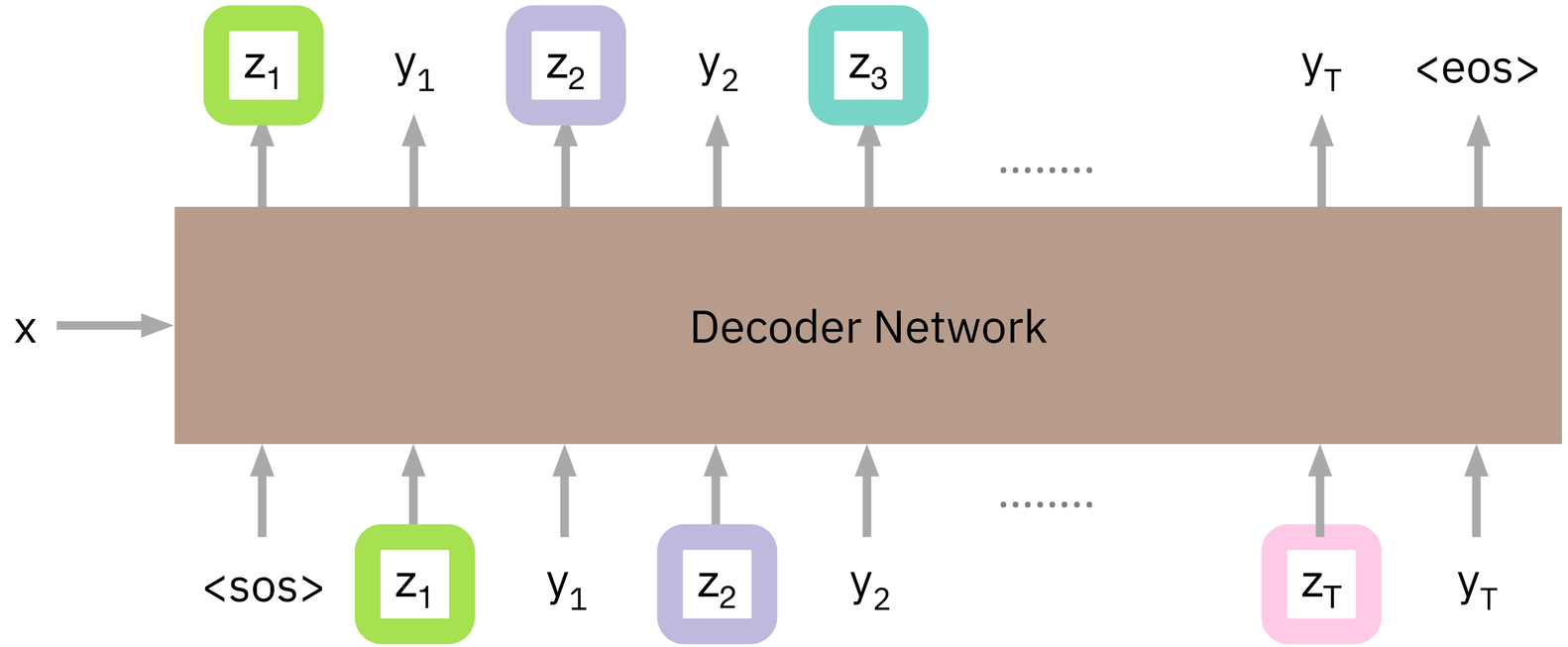
http://seq2seq-vis.io/

http://vcg.seas.harvard.edu

Soon at https://genni.vizhub.ai

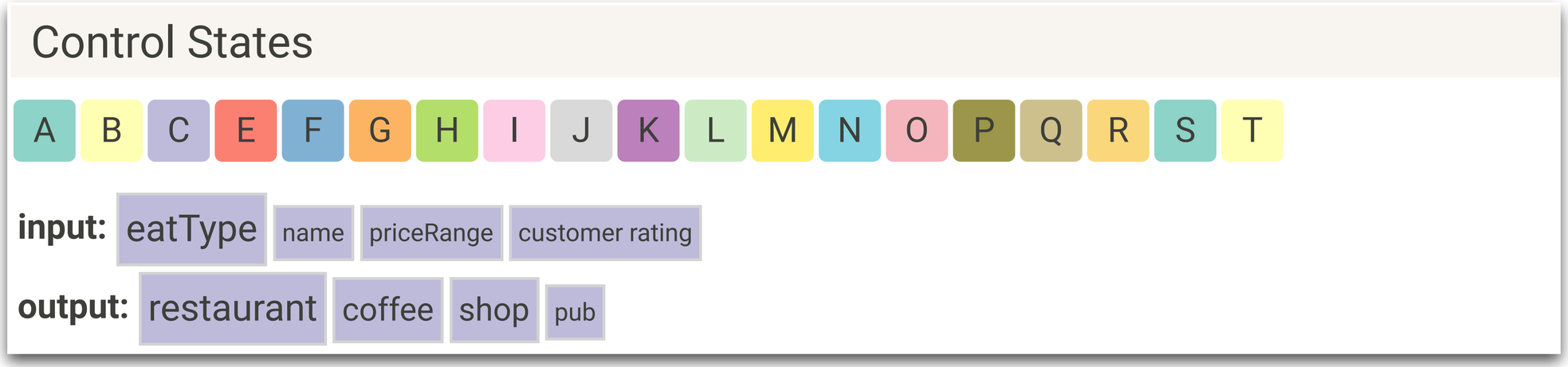
Appendix

Generation



Control States

Each control state z corresponds to a high-level cluster of the corresponding word's semantics learned by the model for the problem.



Open Issues

- For which tasks do we want to involve humans? Could we ever have somebody do interactive corrections of models? Maybe on the data side? Maybe providing confidence values?

- Call to action: let's create tools for this!

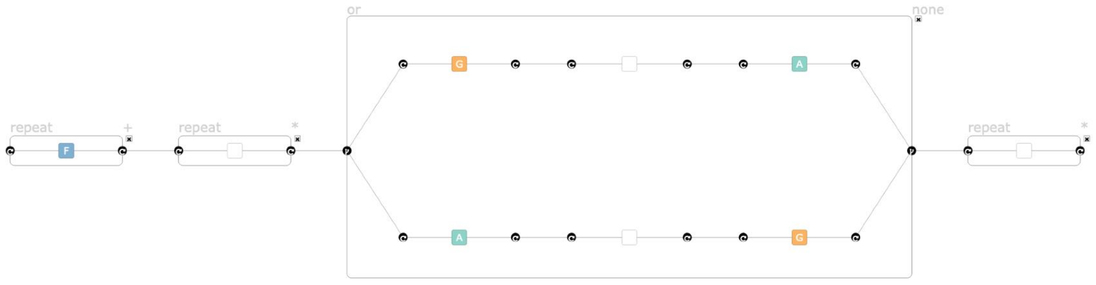
GenNI | model: e2e | CHECK LATENT MODEL

Refine Constraint

Edit Constraint

history: `(F)+(.)*((G)+(.)*(A)+|(A)+(.)*(G)+)(.)*`

constraint: `(F)+(.)*((G)+(.)*(A)+|(A)+(.)*(G)+)(.)*` | FORECAST ⇨ | MAP TO GRAPH ⇩ | GET FROM GRAPH ⇧



Refine by Examples

MERGE SELECTED CONSTRAINTS ⇧ | RESET LIST

☐ the phoenix is a french pub near café sicilia in the city centre constraint ⇧ forecast ⇨

☐ the plough is a family friendly chinese restaurant near raja indian cuisine in the city centre . constraint ⇧ forecast ⇨

Create Example

Select Input ID : -1 | SELECT | OR: SELECT RANDOM | SELECT FROM GLOBAL

name	eattype	food	pricerange	area	familyfriendly	near
the plough	restaurant	chinese	high	city centre	yes	raja indian cuisine

Manual Output (✍)

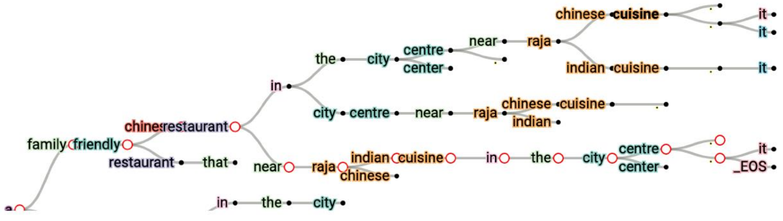
the plough is a family friendly chinese restaurant near raja indian cuisine in the city centre . | ADD TO EXAMPLES + | FORECAST

Alternative Model Outputs (⇄)

RESET

the plough is a family friendly chinese restaurant near raja indian cuisine in the city centre . add to examples + forecast ⇨

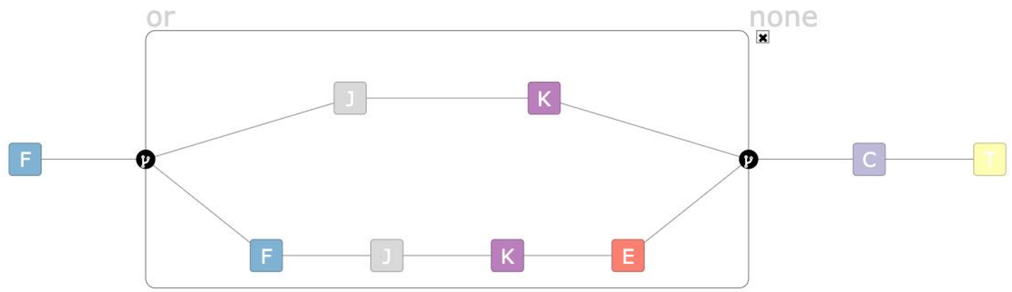
Select a token to see alternatives



Edit Constraint

history: `F(FJKE|JK)CT` `.*`

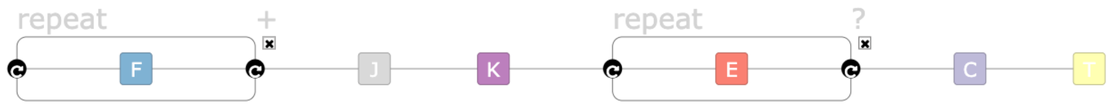
constraint: `F(FJKE|JK)CT` | FORECAST ⇨ | MAP TO GRAPH ⇩ | GET FROM GRAPH ⇧



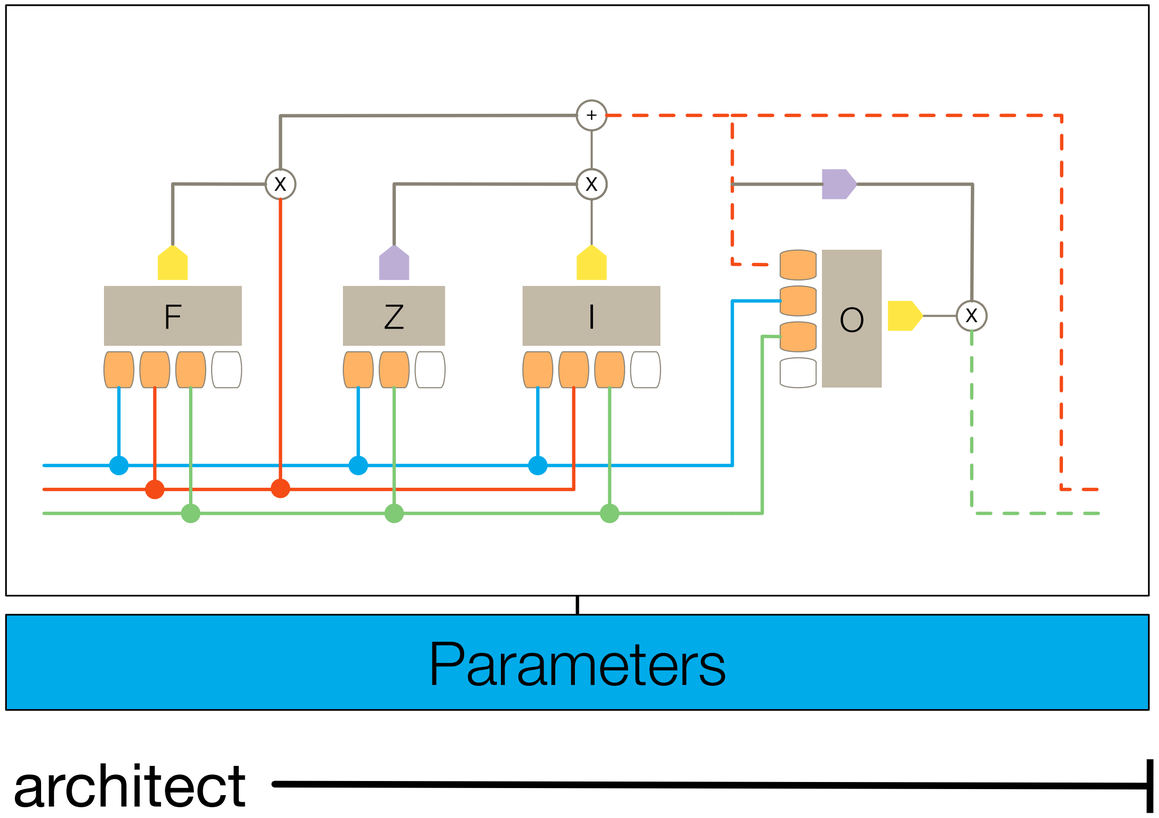
Edit Constraint

history: `(F)+JK(E)?CT` `(F)+(JKE|JK)CT` `F(FJKE|JK)CT` `.*`

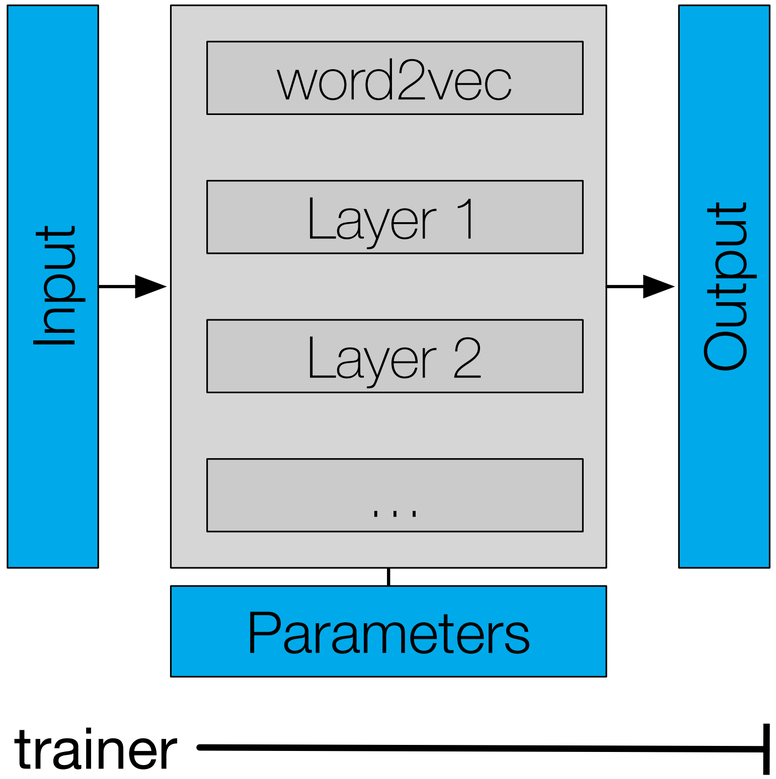
constraint: `(F)+JK(E)?CT` | FORECAST ⇨ | MAP TO GRAPH ⇩ | GET FROM GRAPH ⇧



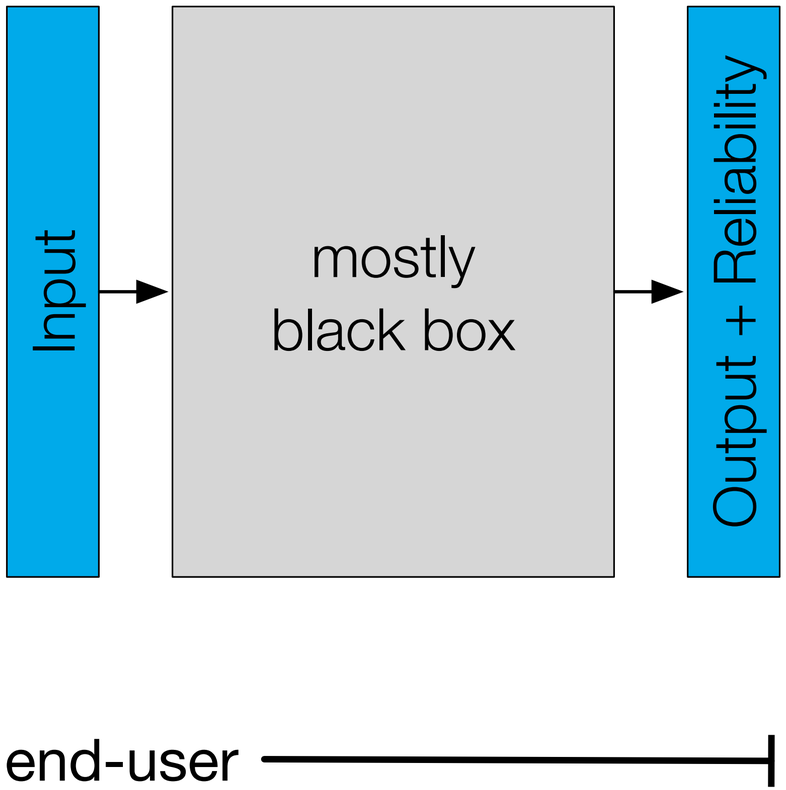
User Role: Architect



User Role: Trainer



User Role: End User

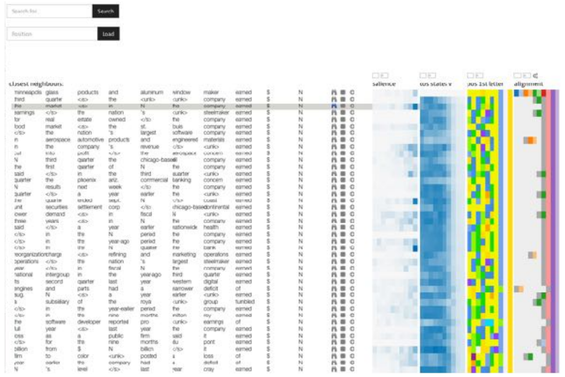


Cell that turns on inside quotes:

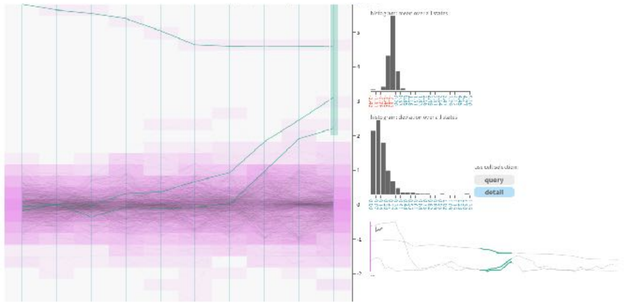
```
"You mean to imply that I have nothing to eat out of.... On the
contrary, I can supply you with everything even if you want to give
dinner parties," warmly replied Chichagov, who tried by every word he
spoke to prove his own rectitude and therefore imagined Kutuzov to be
animated by the same desire.

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

A. Karpathy, J. Johnson, and F. F. Li, “Visualizing and Understanding Recurrent Networks,” ICLR 2016 Workshop

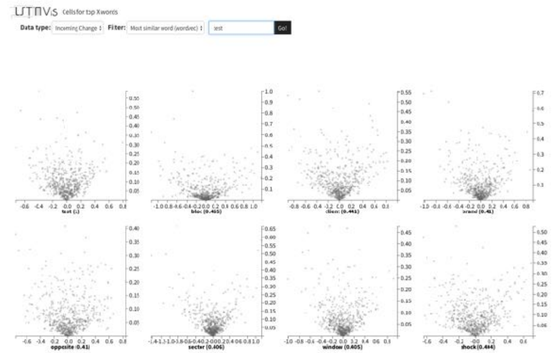
Nine Design Iterations



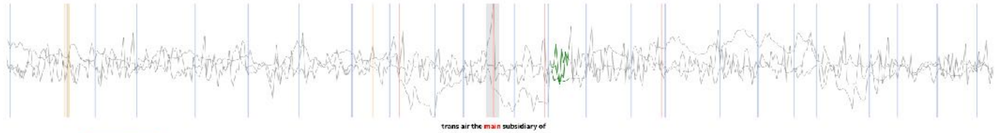
heatmap



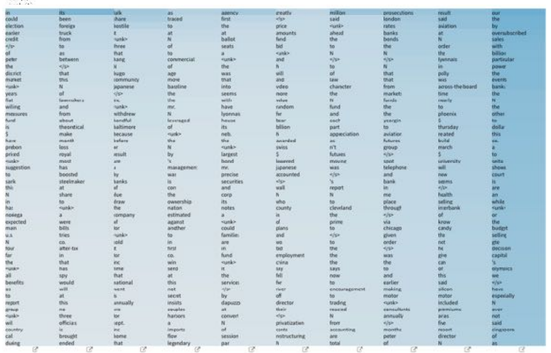
parallel coord. plot



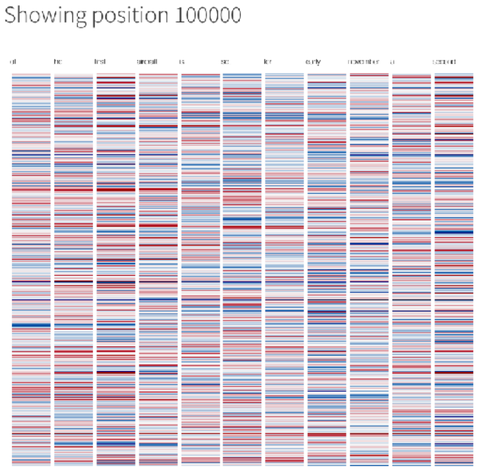
states comparison



linechart



gradient



states

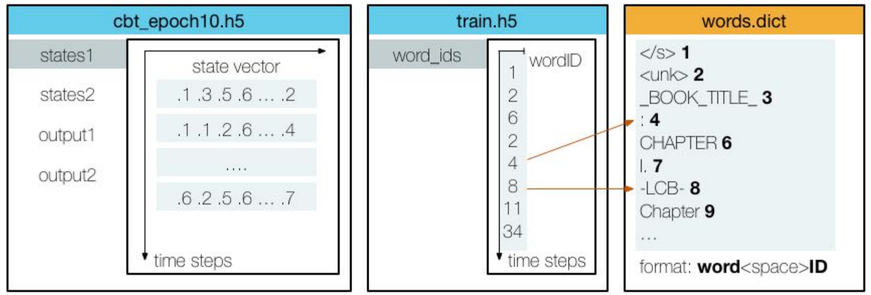
Call for Reproducibility and Public Adoption: open source with **documentation**

Adding Your Own Data

If you want to train your own data first, please read the Training document. If you have your own data at hand, adding it to LSTMVis is very easy. You only need three files:

- HDF5 file containing the state vectors for each time step (e.g. `cbt_epoch10.h5`)
- HDF5 file containing a word ID for each time step (e.g. `train.h5`)*
- Dict file containing the mapping from word ID to word (e.g. `words.dict`)*

A schematic representation of the data:



*If you don't have these files yet, but a space-separated `.txt` file of your training data instead, check out our text conversion tool

Config File

a simple example of an `lstm.yml` is:

```
name: children books  # project name
description: children book texts from the Gutenberg project # little description

files: # assign files to reference name
  states: cbt_epoch10.h5 # HDF5 files have to end with .h5 or .hdf5 !!!
  word_ids: train.h5
  words: words.dict # dict files have to end with .dict !!

word_sequence: # defines the word sequence
  file: train # HDF5 file
  path: word_ids # path to table in HDF5
  dict_file: words # dictionary to map IDs from HDF5 to words

states: # section to define which states of your model you want to look at
  file: states # HDF5 files containing the state for each position
  types: [
    {type: state, layer: 1, path: states1}, # type={state, output}, layer=[1..x], path = HDF5 path
    {type: state, layer: 2, path: states2},
    {type: output, layer: 2, path: output2}
  ]
```

HendrikStrobelt / LSTMVis

Unwatch 42 | Star 524 | Fork 118

Social Media

1.	Twitter	392	(44.95%)
2.	reddit	335	(38.42%)
3.	Facebook	86	(9.86%)
4.	Hacker News	25	(2.87%)
5.	Sina Weibo chinese facebook	18	(2.06%)
6.	VKontakte russian facebook	12	(1.38%)

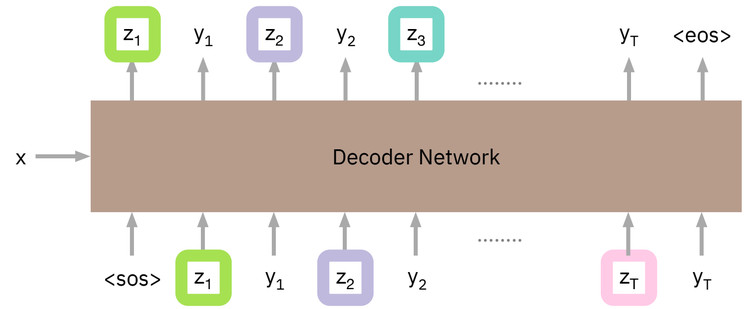
24h

1.	Twitter	2,132	(37.57%)
2.	Google+	1,087	(19.15%)
3.	reddit	920	(16.21%)
4.	Facebook	848	(14.94%)
5.	Sina Weibo	404	(7.12%)
6.	VKontakte	74	(1.30%)

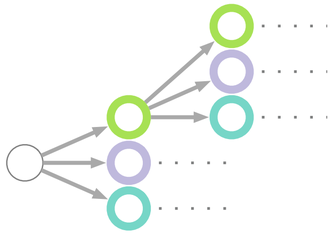
100d

GenNI - Generation Negotiation Interface

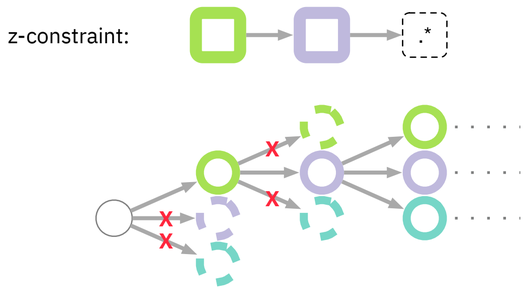
Generation



M1: Free Generation — Input: x —> Output: y, z

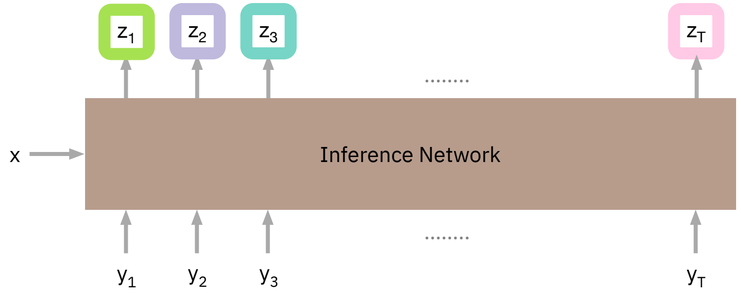


M2: Controlled Generation — Input: x, z-constraint —> Output: y, z



Posterior Inference

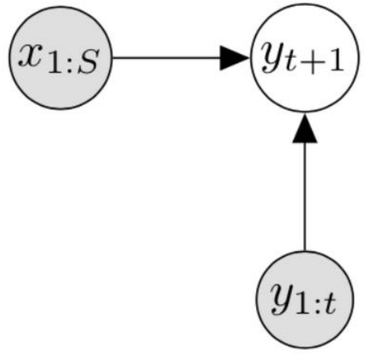
M3: Inferred Control State — Input: x, y —> Output: z



Prediction Network

Predict the next word y_{t+1} based on the previous predictions $y_{1:t}$ and the input x

$$p(y_{t+1} \mid y_{1:t}, x)$$

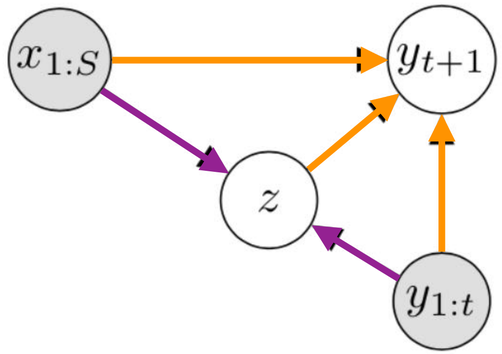


Latent Variable Hooks

Introduce new latent variable z such that:

$$p(y_{t+1} \mid y_{1:t}, x) = \sum_{z} p(y_{t+1} \mid y_{1:t}, x, z) \times p(z \mid y_{1:t}, x)$$

Prediction Network **Hook Network**



Generate Output

Lessons

- Gennie: Correct the model in a generic way, not just correcting one instance.
- Go from one instance to some form of generalization. Cannot correct each example manually.
- Single instance modification does not solve the problem. Hooks are a good way to do it on a more general basis.