



# exploRNN: teaching recurrent neural networks through visual exploration

Alex Bäuerle<sup>1</sup> · Patrick Albus<sup>1</sup> · Raphael Störk<sup>1</sup> · Tina Seufert<sup>1</sup> · Timo Ropinski<sup>1</sup>

Accepted: 10 June 2022 / Published online: 14 July 2022  
© The Author(s) 2022

## Abstract

Due to the success and growing job market of deep learning (DL), students and researchers from many areas are interested in learning about DL technologies. Visualization has been used as a modern medium during this learning process. However, despite the fact that sequential data tasks, such as text and function analysis, are at the forefront of DL research, there does not yet exist an educational visualization that covers recurrent neural networks (RNNs). Additionally, the benefits and trade-offs between using visualization environments and conventional learning material for DL have not yet been evaluated. To address these gaps, we propose *exploRNN*, the first interactively explorable educational visualization for RNNs. *exploRNN* is accessible online and provides an overview of the training process of RNNs at a coarse level, as well as detailed tools for the inspection of data flow within LSTM cells. In an empirical between-subjects study with 37 participants, we investigate the learning outcomes and cognitive load of *exploRNN* compared to a classic text-based learning environment. While learners in the text group are ahead in superficial knowledge acquisition, *exploRNN* is particularly helpful for deeper understanding. Additionally, learning with *exploRNN* is perceived as significantly easier and causes less extraneous load. In conclusion, for difficult learning material, such as neural networks that require deep understanding, interactive visualizations such as *exploRNN* can be helpful.

**Keywords** Neural network education · Recurrent neural networks · Sequential data · Visual education

## 1 Introduction

With its recent advances, DL has gained immense traction in research, industry, and education. As job opportunities related to machine learning are unprecedented, many want to learn about and understand DL technologies.

While initial progress in DL was mainly possible due to the rise of convolutional neural networks (CNNs), large training data sets, and GPU training in the context of image recognition [1–3], other network architectures, such as RNNs, which are able to process sequential data, are becoming increasingly important. At the same time, these more advanced learning architectures are more difficult to comprehend, as they employ concepts that are fundamentally different from classical computer science. Thus, by making the process behind RNNs transparent and easy to understand, research

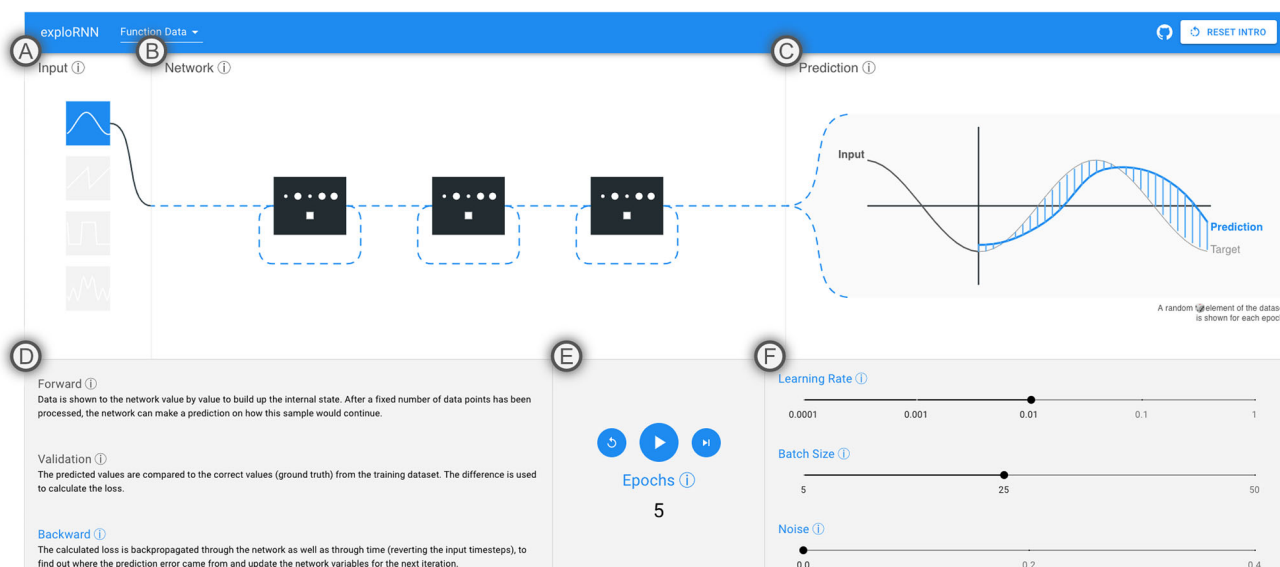
in sequential learning tasks can be accelerated as the field opens up to additional users and contributors.

Along this line, the visualization community has shown how interactive visual explorables can be effective for learning about DL concepts [4–7]. Since different architectures come with their unique challenges, existing educational applications usually focus on one type of architecture. Unfortunately, the set of existing applications still does not cover RNNs. This is despite the fact that RNNs are widely adopted in tasks such as speech processing [8,9], handwriting recognition [10], and machine translation [11], among many others. While RNNs are capable of solving such sequential tasks, they also bring their unique architectures and concepts to capture temporal information. As these concepts differ from other network types, RNN education could be of great benefit. To facilitate RNN education, we propose *exploRNN*, an interactive explorable visualization for RNNs that runs directly in any modern web browser (Fig. 1).

The focus of *exploRNN* is to make learning about these abstract and complex network types easier, more motivating, and more applicable to real problems. By presenting

✉ Alex Bäuerle  
alex@a13x.io

<sup>1</sup> Ulm University, Ulm, Germany



**Fig. 1** **a** Simple input types illustrate the abstract concepts behind RNNs. **b** An animated, modifiable network architecture shows the data flow. **c** The prediction visualization shows the network input, prediction, ground truth, and error bars, all animated to communicate their

temporal nature. **d** Text helps explain the training process. **e** RNNs can be interactively trained. **f** Training parameters can be interactively explored

learning material in a way that is conducive to learning, learners should need fewer unnecessary cognitive resources (CR) [12]. These freed-up resources are then available to be used for a deeper understanding (DU) of the learning material. We also expect that this would result in a more motivating and joyful (MJ) learning experience compared to traditional learning methods, such as text. In turn, learners might be willing to spend more time learning, and more learners could be attracted in general, effectively increasing overall knowledge gain. To assess these hypotheses, we compare *explorNN* with text-based learning in a between-subjects study with 37 participants. Our evaluation provides insights into when, and under which conditions, visual interactive learning environments can outperform conventional learning material.

Along this line, we make the following contributions:

- *Educational Objectives and Design Challenges* for educational RNN visualizations, informing our visualization design.
- *An interactive visualization approach for RNN education*, enabling investigation at different levels of granularity.
- *A quantitative, comparative evaluation*, investigating the effectiveness of our approach and providing hints for other interactive educational visualizations.

*explorNN* can be accessed online at: <https://mi-pages.informatik.uni-ulm.de/explornn>, contributing to a fast-growing corpus of visualization work in the field of DL. To our

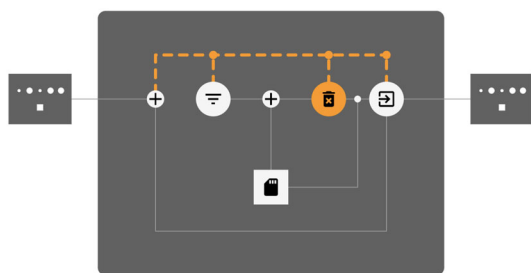
knowledge, *explorNN* is the first educational visualization interface that is targeted at RNNs, an important and growing class of neural networks (NN). Additionally, our study is the first to compare conventional learning material to a visual, interactive learning environment for DL education.

## 2 Background: RNNs

We would like to invite readers who want to refresh their knowledge on RNNs to use *explorNN* at <https://mi-pages.informatik.uni-ulm.de/explornn> as an interactive learning experiment. This chapter contains a brief summary of the knowledge that is communicated in *explorNN*.

CNNs and multi-layer perceptrons (MLPs), which are used for most classical DL tasks, process data in a feed-forward manner. On the contrary, RNNs provide a cyclical architecture in which the output of the previous timestep is used in combination with new inputs to inform the activation of a cell. The main difference in training RNNs is backpropagation through time (BPTT), where the prediction error is propagated not only back through the layers but also within the recurrent connections of the layers.

We visualize the LSTM architecture (cf. Fig. 2). Although this is not the most simple recurrent architecture that exists, it is superior in capturing long-range dependencies, as it mitigates the vanishing gradient problem [13–15], and is thus widely used. The main features of an LSTM cell are the gating mechanisms and the cell state. The three gates within an



**Fig. 2** LSTM cell with all operations visualized. The input is added to the output of the previous time step and then used by the three gates for the gate activation

LSTM cell are computed based on the input at time step  $t$ ,  $x^t$  and the activation of the cell at time step  $t - 1$ ,  $a^{t-1}$  as follows:

- ⊕ Input Gate:  $i^t = sigmoid(W_{ix}x^t + W_{ia}a^{t-1} + b_i)$ , what new information to use to update the cell state  $c^t$ .
- ⊗ Forget Gate:  $f^t = sigmoid(W_{fx}x^t + W_{fa}a^{t-1} + b_f)$ , what information in the cell state  $c^t$  can be forgotten.
- ⊘ Output Gate:  $o^t = sigmoid(W_{ox}x^t + W_{oa}a^{t-1} + b_o)$ , what part of the cell state  $c^t$  is used to compute the activation.

The cell state  $c^t$  at timestep  $t$  is then computed as  $c^t = f^t \circ c^{t-1} + i^t \circ tanh(W_{cx}x^t + W_{ca}a^{t-1} + b_c)$ , where  $\circ$  is the hadamard product.

While there are other architectures that also use the concept of cell state and modular updating, such as gated recurrent units (GRUs) [16], their underlying idea does not greatly differ. However, since LSTMs were the first to introduce the explained concepts, are more general in their application [17], and often outperform GRUs [18], we focus on conveying the LSTM architecture.

### 3 Related work

In this section, we first give a brief overview of the explorable explanation literature before elaborating on the corpus of related work in the area of educational visualizations for DL non-experts and RNN visualizations.

#### 3.1 Explorable explanations

Explorable learning environments were invented long before DL raised awareness in the broader public. Their effectiveness was investigated in the line of work by Hundhausen et al. [19,20]. Schweitzer and Brown then described design characteristics and an evaluation of active learning settings in classrooms by using visualization [21]. There also exists a line of work on the use of visualization for programming education [22–24]. These approaches show how visualization can communicate algorithmic thinking effectively. We combine these ideas with more recent concepts, which have

been proposed under the term *exploration* in the area of science education [25], where explorable explanations provide benefits for learning.

There are numerous helpful visualizations conveying properties of NN architectures, their functionality, or application scenarios [5,6,26]. However, we will focus on educational, explorable visualization approaches that have been proposed for different network types. One of the most prominent interactive educational visualization approaches has been proposed by Smilkov et al. [7]. In their explainable *Tensorflow Playground*, one can select the properties of a NN to be trained. They also allow the customization of certain training parameters and deployed their approach as a web-based application. Similarly, in *Revacnn*, users can explore the activations of a CNN by modifying the network structure and training the network in the browser [27]. While these approaches help teach the most basic concepts of MLPs and CNNs, respectively, more advanced architectures need further, specialized visualizations. Another approach that is closely related to ours, but works on a different type of NNs, namely GANs, is *GanLab* [4]. They focus on how the generator and discriminator are used adversarially to yield synthetic data that resembles the data distribution it was trained on. However, GANs bring their own visualization challenges, which are fundamentally different from those we found for RNNs. Additionally, neither of these systems was systematically and quantitatively evaluated.

As none of these visualization approaches is designed to help non-experts understand how RNNs function, with their unique concepts of memory and temporal dependence, our aim is to fill this gap in the literature with *exploRNN*. Additionally, we shed light on the usefulness of such interactive learning environments through our quantitative user study. We specifically examine the difference in learning outcomes across different learning hierarchies [28], the complexity of the learning experience by means of cognitive load [12], and qualitative assessments such as motivation, perceived quality of content, and joy throughout the learning process. We hope that our findings in this area can be an indication for similar learning environments and motivate others to conduct similar experiments.

#### 3.2 RNN visualizations

Apart from educational visualizations, there is another line of visualization work targeted toward investigating RNNs. These approaches are designed mainly for researchers who want to understand and debug their models. An early approach toward visualizing RNNs was proposed by Karpathy et al., who visualize the activation of RNN cells for expert analysis [29]. Strobelt et al. published *LSTMVis*, in which the hidden state dynamics of RNNs is investigated [30]. They specifically demonstrate how text understanding can

be analyzed through investigating the structure and change of the cell state. They also presented *Seq2Seq-Vis*, in which sequence-to-sequence models can be probed to reveal errors and learned patterns [31]. Along the same line, Ming et al. introduced RNNVis [32]. They analyze the functionality of individual hidden state units by observing their reaction to specific text segments. With RNNbow, Cashman et al. published a visualization, in which the gradients of RNNs can be analyzed [33]. They attribute the gradient to individual letters in a textual input sequence. This way, researchers can inspect how their models learn. In another approach, Shen et al. proposed visualizations for RNNs [34] operating on multi-dimensional sequence data. Here, developers can inspect hidden unit responses to get insight into different networks. Similarly, Garcia and Weiskopf proposed a visualization for the inspection of hidden states of RNNs [35]. However, all approaches described here are expert tools that help during development. Contrary to this, we aim to convey the general idea of RNNs to novices in this area of DL.

Insights on the effects of using exploration and visualization for learning in general, as well as present educational visualizations for NNs, show how interactive exploration can help a broader audience with access to learning experiences. Therefore, we propose *exploRNN*, which provides insight into the function of RNNs for users who know the basics of DL, but are laymen in the area of sequential learning. Our evaluation also provides the first comparative analysis of interactive learning environments and classical learning approaches for NN education.

## 4 Educational objectives

To inform the visualization design of a learning experience, educational objectives are needed, which we defined based on Bloom's taxonomy [28]. Our target users already understand the fundamental concepts of DL and know about feed-forward NNs. Without this background knowledge, the theory behind those techniques would first need to be explained, which would extend the scope of *exploRNN*. As our target audience aims to learn the yet unknown concepts of RNNs, we focused on *recall* (O1&2), *comprehension* (O2&3), and *transfer* (O3&4) of the learned information. Later, this learning can be applied in the wild to access levels four to six (*analyze, evaluate, create*) of Bloom's taxonomy. Formulated on this basis, our educational objectives are:

**O1** *Justification*. Users should know that RNNs, in contrast to other network types, can be used for sequential data. This also includes BPTT, through which RNNs can learn temporal dependencies, which classical feed-forward networks cannot.

**O2** *Cell Structure*. Users should then understand how LSTM cells are built and what functionality their individ-

ual components have. Here, the cell gates, as well as the memory element, are of special importance, as they enable the processing of sequential data.

**O3** *Training Setup*. To understand the training process of such networks, users should know important parameters for the setup of RNNs. This includes the network structure, training parameters, and how data are fed to the network.

**O4** *Task*. Finally, to transfer this theoretical knowledge about RNNs to real applications, users should learn about different application areas and data types that can be used with RNNs. In the end, they should be able to describe how RNNs could be used for their own application scenarios.

Similarly to a lecture at a university or a textbook, our learning environment is designed to provide an introduction to RNNs from which interested users can start experimenting with the techniques. Accordingly, our educational objectives not only motivate the importance of RNNs, but are also aimed at providing insights about the input data and related tasks, as well as how the training process and LSTM cells work.

## 5 Design challenges

Since RNN cells are a special form of NN layers, they open up unique challenges for visualization-based education. We observed both visualization design challenges and technical challenges, which we describe in this section.

### 5.1 Visualization design challenges

We first discuss the following visualization design challenges that we identified in the context of an interactive learning environment aimed at RNNs and illustrate how they relate to our educational objectives:

**V1** *Complexity*. As mentioned in Sect. 1, one of our central goals is to simplify learning by reducing cognitive load [12]. However, RNNs are typically trained on a large amount of complex data that can be difficult to grasp (O1) (O4) [36,37]. The same holds for network architectures, which are also often too complex to fully comprehend in their entirety (O3) [38,39]. Consequently, all visualizations must be interpretable and intuitive, but realistic enough to form a compelling use case [40,41].

**V2** *Dynamics*. An educational system to teach RNN concepts should clarify the dynamics of the sequential data on which these networks operate (O1), as well as the dynamics of the training process (O3). These dynamic processes must be visually communicated, including data type and data processing, both forward (inference) and backward (back-propagation), within the network [42].

**V3** *Multiscale*. RNN structures need to be communicated at different granularities, i.e., network, cell, and cell components (O2). These multiple scales need to be fluidly

inspectable, while at the same time, the granularity at which the user currently operates must be communicated [41,43].

**V4** *Supervision*. In classical learning settings, teacher supervision or other opportunities to seek further information is provided. Contrary to this, *exploRNN* is designed as a standalone learning environment that does not require external guidance **O1-4**. Thus, supervision has to be substituted by visual guidance [44,45].

## 5.2 Technical challenges

Whereas the visualization design challenges are based directly on our educational objectives, the following technical challenges relate to the development of such an interactive, explorable learning environment:

**T1** *Training Time*. Typically, training processes can take up to several days to convergence [46,47]. However, for an interactive learning experience, waiting days for convergence is not feasible. To provide direct feedback to the user, our networks thus have to converge in minutes instead of hours or days.

**T2** *Training Steps*. Normally, computation is done as fast as possible to minimize the time it takes for the network to converge. However, we want the user to be able to follow the training process and observe individual training steps [44]. Thus, training steps should be separated temporally from the visualization.

**T3** *Deployment*. Modern-day learning is often conducted via online courses, blog posts, or explainable web pages [48]. Although this makes such learning environments accessible to a broad audience, it also limits the technical freedom of such applications [49]. Therefore, educational environments should be deployed to a broad audience, while also providing diverse functionality.

## 6 Visualization design

In the following, we discuss the visualization design of *exploRNN*.

We explain how we tackle the aforementioned visualization challenges **Vx** and learning psychology goals **CR/DU/MJ** while targeting the educational objectives **Ox** defined in Sect. 4. We first describe the overall visualization concepts we implemented for *exploRNN*. Then, we elaborate on the different views of our environment in the upcoming subsections.

**Scales** To show both an overview of the training process **O3** and give detailed insight into the computation that is performed within one recurrent cell **O2** **DU**, we employ an overview first, zoom and filter, then details on demand visualization design, following Shneiderman's mantra **V3** [43]. Therefore, *exploRNN* consists of two main views, the network overview (Fig. 1), which displays the training progress

on the network scale, and the LSTM cell view (Fig. 6), which allows for a detailed inspection of an LSTM cell. This is in line with our goal of reducing complexity **CR** by focusing on individual steps of the learning process rather than presenting everything at once.

**Animation** Animation has shown to be effective in visualizing data relationships and algorithms **O1-3** [42,50,51]. Furthermore, animation has shown to be associated with fun and excitement [52], which is in line with our goal of making learning more enjoyable **MJ**. Thus, to visually communicate how the network operates on sequential data, we use animation throughout our visualizations **V2**.

**Onboarding** Novel visualizations and interactive systems can be hard to understand [53]. We designed *exploRNN* in a way that allows exploration without running the risk of making irreversible errors or needing teacher supervision **V4**. However, instructional aids may be important to understand such complex content **DU** [54]. Therefore, we use an onboarding process for our educational environment [55] (cf. Fig. 3). With this process, we aim to further reduce the cognitive load during learning compared to classical learning environments **CR** [56]. For example, the sequential nature of RNNs **O1** **V2** and the data and tasks that RNNs can be used for **O4** are communicated in *exploRNN*.

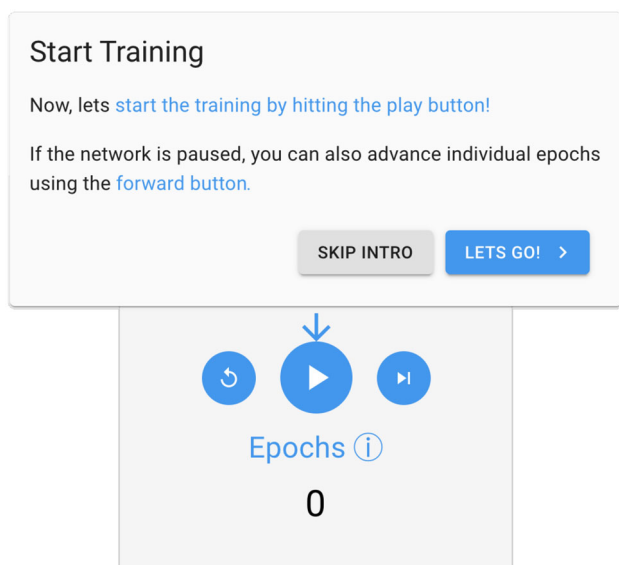
**Textual explanations** In contrast to other learning environments, which show static textual explanations below the main visualization [4,7], we instead provide such additional information as *details on demand* **V3** [43]. This way, users can access more information for exactly the components they want to learn more about **DU**, while not having to read a lot of text **CR**. Our interactively explorable dialog boxes, as shown in Fig. 4, provide information about all important elements of the learning environment. Such dialogs exist for all headings and are anchored through an **i** icon, and for all components of an LSTM cell, which is referred to in our onboarding process.

### 6.1 Network overview

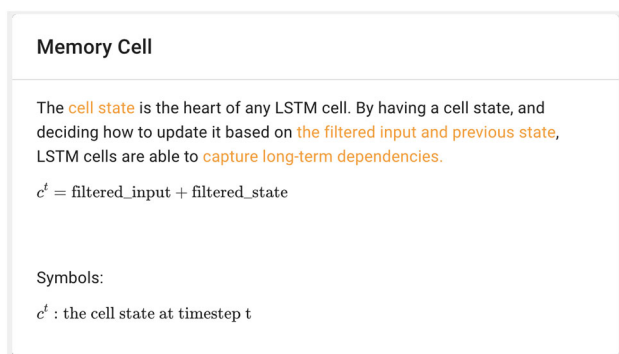
In the network overview, following the natural reading direction of western cultures, as well as related work on NN architecture visualization [57–59], we arrange the network from left to right. On the left, one can see the input type that is currently used to train the network **A**. Centered, we present an abstracted visualization of the network, where users can see how many layers the network contains **B**. On the right, a visualization of the prediction along with the prediction error shows how training progresses **C**. Below these visualizations, information about the training process, controls for the training process, and means to change training parameters are shown **D-F**.

**A** *Input*. To experiment with the network, users can select the input data that is used to train the network from





**Fig. 3** Onboarding dialogs guide the user through our visualizations, so that no manual introduction is needed, and the user can explore *exploRNN* on their own. Textual descriptions with highlights provide detailed explanations for individual components. Positioning and arrows reveal associations between dialogs and components



**Fig. 4** Users can access more detailed explanations for many elements of our visualizations such as training steps, hyper-parameters, and operations in a cell

a set of explanatory input types. Data for an interactive and explainable visualization of NNs needs to both explain the network functionality (O1) and be easy to understand (V1). Therefore, current educational visualizations use an abstract, two-dimensional distribution of points to train their networks on [4, 7]. With *exploRNN*, we follow this approach of employing data that is as simple as possible (CR). As RNNs are focused on sequential data, we decided to use periodical mathematical functions and simple text snippets, which map nicely to the sequential nature of RNNs (V2). The functions that can be used as training data in *exploRNN* are a sinusoidal function, a sawtooth function, an oscillating function, and a composite sinusoidal function and vary in their periodicity. To demonstrate the sequential and dynamic nature of these

input functions, we animated those that are in use so that they seem to flow while being input to the network (MJ).

In addition to abstract function continuation, we also provide text-based data to train the network on (DU). To allow for interactive training, we employ rather simplistic text samples. These include a recurring character sequence (*ababab...*) and the well-known text *lorem ipsum*. Here, we employ a similar design language as with function data, to show that most ideas can be transferred across tasks. By incorporating this text learning scenario, users of *exploRNN* get to learn and inspect not only abstract problems, but can also experience more realistic scenarios (O4) (MJ).

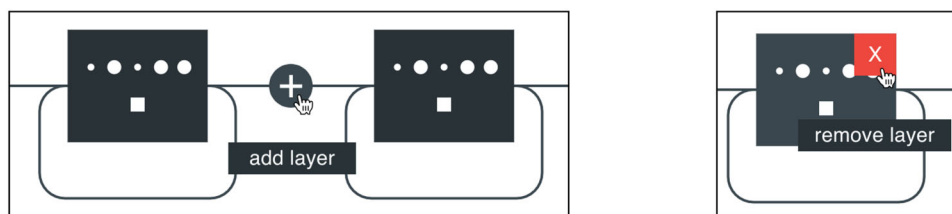
**B) Network.** In the network visualization, we want to communicate the recurrent nature of our network (O1), but at the same time, show all layers. Thus, instead of the more frequently used unrolling of RNN layers [60], we add a loop to the layer glyph to symbolize this recurrence. This symbolizes the feedback loop of information output at  $t$  back to the input of a cell at  $t + 1$  (V2), which enables BPTT.

Our network visualization is animated as data flows through its layers (O3) (MJ). For the prediction step, dashed lines flow in the forward direction to symbolize forward data processing. For the backpropagation step, they flow backward to resemble the backpropagation of the error (V2). Dashed lines are moving from input to output during the prediction phase, and from output to the first network layer during training, because backpropagation is not applied to the input domain.

Users can also investigate how the training progresses differently depending on the number of recurrent layers in the network (O3). Therefore, layers can be added or removed from the network to be trained, as shown in Fig. 5 (DU). As with most explorable components, we explain the implications of this in our introduction, and users can click the (i) next to the network heading.

**C) Predictions.** Commencing the top row of visualizations is the data plot, where we visualize an input sample and the prediction of the network along with its ground truth (O3). Here, multiple data points that are processed by the network one after the other are used to inform a prediction, which is visualized by sliding a gray box over the input data that is currently processed (V2). Additionally, the prediction values slowly build up with animations to clarify that this prediction is building up sequentially (MJ). We then use vertical lines in the function plots, which slowly emerge between the prediction and the target value. This vertical line encoding is in analogy with the way we calculate errors, namely, by looking at the prediction values and calculating the difference to the ground truth (DU). The error calculation is embedded temporally between the inference (forward network animation) and backpropagation (backward network animation) phases of the training process (V2). Altogether, through this animated component, while not being interactive itself, users

**Fig. 5** left: Adding a layer between two existing layers. right: Removing a layer from the network



can inspect the results of modifications that have been made in other places (O3).

**D Process.** According to the typical NN training setup, we divide the training process into three distinct steps: inference (forward), validation (error calculation), and back-propagation (backward). The explanation pane in the lower left of the network overview (see Fig. 1) displays which step is currently executed and provides an explanation of what happens in each of these steps (DU). Through this, the user can learn more about the training dynamics of the network (O3). As described previously, animations in other components complement this dynamic nature of the training process (V2).

**E Controls.** In the network overview, the network is trained by means of epochs *first* provide an *overview* [43] of the training process (V3 CR). To experiment (O3), users can interact with the control area in the bottom center of our environment (MJ). In addition to automatically advancing epochs, which can be controlled with the ► and || buttons, users can also trigger network training for a single epoch, by pressing the ► button (DU). The training process can always be reset using the ↺ button. A back button to go to a previous epoch is not included in *exploRNN* as this would require saving multiple previous states of the network parameters, which would require significant browser memory. Therefore, and as individual epochs normally do not change the network behavior completely, going back one training epoch during training is not a common operation during neural network training, so we think users will not miss such functionality.

**F Training Parameters.** To communicate the training setup of an RNN (O3), a trade-off between completeness and simplicity must be made (V1). Thus, we let the user freely choose some training parameters, but employ restrictions for others (CR). As mentioned, users can add or remove individual network layers and use different preset training inputs. In addition, they can change the learning rate, batch size, and noise (DU). The learning rate and batch size allow for exploration of different training settings (O3). Noise can be added to make the training data more realistic, resembling real-world scenarios of imperfect measurements (O4). Parameter changes can be made through sliders, which are positioned on the bottom right. To provide an intuition about the influence of these parameters, we include pretrained models that are loaded during the onboarding steps which explain each indi-

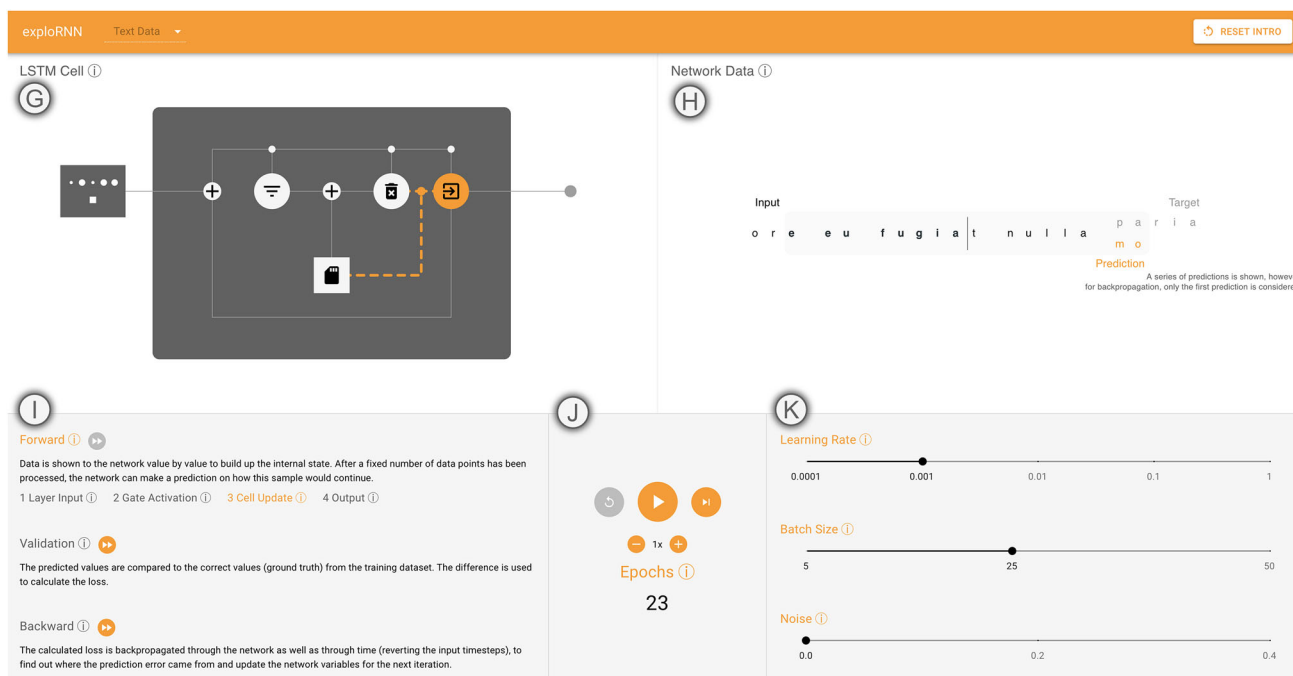
vidual parameter (V4). Other parameters, such as units per layer or optimization strategies, cannot be changed in our implementation. This trade-off between freedom of exploration and simplicity proved to be effective in educating users about the influence of different training parameters and keeping their cognitive load low (V1).

Hierarchical aggregation can help simplify visualization designs (CR [61]). Thus, after getting an overview of the network, the user can inspect another hierarchy level in detail, namely individual LSTM cells (O2 [43]). When selecting one of the layers in the network overview a zooming transition onto one of the network layers gradually reveals the structure of an LSTM cell to support the user's mental image of looking into one of the layers (V3 CR). With this multiscale approach, where users can navigate between views, orientation is important (CR). Therefore, a color coding indicates the current level of detail. This highlight color is blue for the network overview, whereas orange is used for the LSTM cell view. Orange and blue are complementary colors, which makes them easily distinguishable, and they can be differentiated by vision-impaired users [62].

## 6.2 LSTM cell view

In the LSTM cell view, we show a detailed visualization of the selected cell on the left, embedded in small pictograms of neighboring cells (G). On the right of this cell visualization, one can see the input, target, and prediction values of the network, where new points are added as they flow through the cell (H). Below these visualizations, we show information about the training process, controls for the training process, and means to change training parameters, similar to the network overview (I-K).

**G Cell Architecture.** To convey the functionality of one recurrent unit (O2), we show all computational elements within a cell (DU). Wherever information is combined, we show a + icon. Icons for the input (≡), forget (⊖), and output gates (⊗) visualize the gating functionality of an LSTM cell. While all gates that transform the data are depicted with circular icons, the cell state, which represents the saved state of the cell, is represented by a squared ■ icon, illustrating the semantic difference between these components. Each of these cell components can be selected to get a detailed expla-



**Fig. 6** LSTM cell view. (G) Visualization of data flow through the cell. (H) Input to the network and its prediction. Visualization of the training error computation. A gray sliding window indicates which data points are needed to initialize the cell state. (I) Explanations with more detailed

steps for the forward direction of data flow. (J) In addition to interactively training the network, users can change the speed at which the visualization for cell steps advances. (K) Just as in the network overview, users can modify training parameters

nation of its functionality, as shown in Fig. 4, marking another level of detail in this visualization (V3) (CR).

Data flow is visualized through connecting lines (O2) and step-by-step animations of the cell components (MJ). Here, elements that process data in the currently visualized computation step are highlighted. As in the network overview, connections moving data are symbolized with dashed lines. Those lines flow forward during inference and backward during backpropagation. This way, we communicate how the hidden state and output of these cells is computed and visualize how the data flows from one to the next operation or gate (V2) (DU).

The reverse data flow of BPTT occurs not only once within a cell to backpropagate to the previous layer, but multiple times, for all input time steps (O1). The connections within the cell also clarify that there are two recurrent cycles, one from the output of the cell back to the input, and one within the cell to update the cell state based on its state in the previous iteration (DU). As a result, while other visualizations require *unrolling*, where time steps are visualized by displaying multiple cells in concatenation [60], we communicate recurrence through step-by-step animation. This removes the ambiguity of stacked layers vs. unrolled cells, which was shown to hinder learners in our first experiments (V1) (CR).

(H) *Data Plot*. Right of the cell visualization, we show the input data, network prediction, and ground truth all in

one graph. In contrast to the network overview, where the network is directly connected to this output graph, the cell is disentangled from this visualization. As the depicted cell typically receives data from previous cells and outputs data to subsequent cells, this visualization, where animation steps are synchronized but not visually connected on both the input and output side, better reflects the network architecture of RNNs (V3) (DU). Users can inspect this view during interactively controlled training to see how the network processes input data to make predictions sequentially and how it calculates the training loss in relation to the processing steps within a cell (O1&2) (V2).

(I) *Training Process*. The three steps of inference, validation, and backpropagation are just as relevant in the LSTM cell view as they are in the network overview (O3). As the training speed is lower in the LSTM cell view, users can skip part of the data processing and go directly to the processing step of interest (V1&2). For the forward pass, we add additional explanations for the different processing steps of *receiving the layer input*, *calculating the gate activations*, *updating the cell state*, and *outputting the activation value* (DU). These explanations are highlighted in synchronization with the processing steps during the forward pass to the data flow in the cell visualization above (MJ), allowing users to draw links between the processing steps and the explanations they are interested in (V2&4) (CR).



**J Controls.** In the LSTM cell view, processing is done by means of compute steps, showing a much more detailed processing pipeline than in the network overview **V3**. As in the network overview, the control area can be used to experiment with the training process **O3**. The more fine-grained advancement of the visualization is also adopted by the degree to which the animation advances with the forward button, since it only executes the next compute step within a cell **DU**. In addition to what can be done in the network overview, the speed of the animations for data processing within this cell can be adjusted, so that users can explore the processing steps at their own pace **V2**.

We want to emphasize the buildup of state within a cell based on multiple input time steps. Thus, we show how the network processes these inputs in great detail, whereas we made the animation of the backpropagation take less time than forward processing. As *exploRNN* is not designed to represent accurate timings anyway, this is our way of visualizing cell processes in detail, while also preserving the ability to observe multiple epochs.

**K Training Parameters.** Training parameters can be adjusted in the LSTM cell view just as in the network overview, giving the user even more control over the training process and room for experimentation **O3**.

To get back to the network overview, one can click anywhere outside of the LSTM cell in Fig. 6 **G V3**.

## 7 Technical realization

While the visualization design described above has been carefully crafted to meet the educational objectives described in Sect. 4 and the visualization design challenges outlined in Sect. 5.1, its technical realization needs to take into account the technical challenges identified in Sect. 5.2. In this section, we detail how we tackled these technical challenges.

**T1 Training Time.** While an RNN for a complex application cannot be trained live in the browser, we simplify the problem in multiple ways. By employing simplistic data sets, the model can converge after relatively few epochs. Additionally, we limit the number of data points that are fed to the network per epoch. Therefore, epochs are processed sufficiently fast for our interactive visualization approach. We also limit the network size to at most seven layers, so that memory consumption and processing time are reduced. In turn, users can see the training progress and get visible prediction improvements after only a few epochs, while one such epoch takes seconds to compute.

**T2 Training Steps.** A key aspect of our approach is the decoupling of computation and visualization. Through this decoupling, we are able to show the training steps in an observable manner and enable exploration at the user's own

pace. This helps users understand how the model processes input data and predicts new data points.

**T3 Deployment.** To be able to make *exploRNN* publicly available for a large audience, we implemented it as an interactive browser application using HTML and JavaScript. To train the RNN, we use TensorflowJS [49], for animated visualizations of the trained network, we use P5.js [63]. This way, we are able to provide an interactive, web application that visualizes the training dynamics of RNNs through animation, which is accessible at: <https://mi-pages.informatik.uni-ulm.de/explorinn/>.

## 8 Limitations

While *exploRNN* provides a novel environment for learning about RNNs, there is still room for more advanced visualization designs that could be explored in the future. Some of these limitations are explained in the following.

*Explanations* *exploRNN* offers a lot of experimentation that is complemented by textual explanations. However, the number of textual explanations that fit into the context of such an educational system, which is designed to provide an overview of this complex topic, is insufficient to fully explain RNNs. For specific questions that are not addressed by our interactive system, we refer to developer documentation and scientific papers.

*Drill-Down* *exploRNN* explains RNNs on both a network and a cell level. Apart from seeing the data flow on these granularities and textually describing the components of a cell, visualizing the workings of these components could further benefit the learning experience. However, these components are just mathematical functions to which neither the input nor output have a directly discernable meaning. If we were to, e.g., visualize the internals of a memory component, users could only see matrices of seemingly meaningless numbers flowing through these cells. This would not add any benefit and might even result in confusion about such a visualization. To explain these internal components, novel interpretability techniques might help. Inventing and implementing those is beyond the scope of this work.

*Component change* To see the influence of individual components in a cell, changing or removing them could be an interesting addition to the workflow. We did not implement this capability for two reasons. First, adding such functionality goes deep into the working of individual cells, which would exceed the learning objective of getting an overview of RNNs and LSTM cells. In turn, we assume that changing single components in individual cells is unlikely to have a measurable and interpretable effect on the overall learning outcome. Second, we would have needed to implement our own DL library for this to be possible, as TensorflowJS has predefined LSTM layer implementations.

*Degrees of freedom* While users can change some hyperparameters and network settings in our environment, we deliberately do not expose all possible settings to our users. The goal of this limited exploration setting is that users can get an overview of important manipulations to be made, while at the same time not overwhelming our target audience. As for limited explanations, we refer to developer documentation and scientific papers for users that want to explore these details.

*Layer types* In our implementation, we focused on conveying LSTM cells. However, there are numerous other cell architectures for RNNs. Although we do not think this limited focus hinders learners with understanding RNNs on a high level, it would, nonetheless, be helpful for users specifically interested in certain cell types to include these in *exploRNN*.

## 9 User study

To evaluate the effectiveness of our approach, we conducted a user study with 37 participants (30 male, seven female) aged between 21 and 32. Participants were recruited from a DL course at our local university. Our study was a lecture at the end of the course, after students had already learned about feed-forward NNs. Participants were randomly assigned to one of two groups. The *exploRNN* group received the interactive application, and the text group was presented a text-based learning environment.

To look at learning outcome in detail, our evaluation was divided into the first three distinct, hierarchical cognitive learning goals according to Bloom's taxonomy [28], namely recall, comprehension, and transfer. We expect higher learning outcomes for the *exploRNN* group compared to the text group at all three levels. For a closer look at the cognitive processes involved in learning (CR), we also collected data for the three types of cognitive load [12]. Intrinsic cognitive load (ICL) results from the natural complexity that underlies the learning content. Since the difficulty does not differ, there should be no difference between the two groups. Extraneous cognitive load (ECL) is caused by inadequate instruction or presentation of information. Due to the step-by-step presentation of information and the direct connection of textual information and explanatory figures in the *exploRNN* group, we expect lower ECL for the *exploRNN* group compared to the text group. Lastly, germane cognitive load (GCL) represents the invested learning-related load. GCL is connected to the processes that are needed to construct and automate mental representations [12]. Following the reduced ECL in the *exploRNN* group, learners should have more free cognitive capacity in working memory to invest in learning-related GCL.

## 9.1 Hypotheses

Based on the described theory, we hypothesize the following. We expect a higher learning outcome, differentiated by recall (H1), comprehension (H2) and transfer (H3) in the *exploRNN* group than in the text group. Furthermore, we expect no differences between the groups for ICL (H4). We expect a lower ECL in the *exploRNN* group than in the text group (H5). For the GCL, we expect it to be higher in the *exploRNN* group compared to the text group (H6).

## 9.2 Method

Our study was split into different steps, which we explain in the order they were presented to the participants.

*Prior knowledge.* Prior knowledge was measured with seven open-ended questions on NNs and DL techniques (e.g., *Name two activation functions used in deep learning.*). The questions were developed by a domain expert. All answers were rated by a domain expert, following a predefined solution to ensure objectivity. A total of one point could be scored for each question, with partial points of .5. The maximum score for the prior knowledge test was seven.

*Motivation (MSLQ).* To assess motivation, the MSLQ [64] subscale for motivation was used. The MSLQ is a self-report questionnaire designed for an academic setting. Motivation was measured with twelve items (e.g., *I'm confident I can do an excellent job on the tests in this study.*). Learners were instructed to respond as accurately as possible, reflecting their attitudes and behaviors toward the learning module. Responses were given on a 7-point Likert scale ranging from 1 *strongly disagree* to 7 *strongly agree*. Cronbach's Alpha was computed for the internal consistency of the measures [65], and the reliability was  $\alpha = .95$ .

*Learning material.* The learning material was presented either as a text with illustrating figures, formulas, and graphs (see our supplementary material) or through *exploRNN* (see website). For both conditions, the information was the same. The only difference was the presentation medium and the lack of interactivity in text-based learning.

*Learning outcome.* To assess learning outcome, a domain expert developed a posttest with 11 open questions on the content of the learning session. To better understand cognitive processes, the posttest was differentiated by the first three levels of Bloom's taxonomy (DU) [28]. Recall was measured with four questions (e.g., *Name the backpropagation algorithm that is used for RNNs.*). Comprehension was also measured with four questions (e.g., *Explain the meaning of this formula:  $c_t = \text{filtered\_input} + \text{filtered\_state}$ .*). The main purpose of these questions was to test how well people could explain and discuss the learning content. Transfer was measured with three questions (e.g., *Assuming you have a poem continuation network and training data with poems from the*

internet. If your network now makes a prediction, how do you determine if it is correct, to calculate the loss?). These questions were designed to test the ability of learners to draw inferences from the learning content and apply it to new contexts. Similarly to the prior knowledge test, each question was rated by a domain expert, following a predefined solution to ensure objectivity. A total of one point could be scored for each question, with partial points of .5. The maximum score for recall and comprehension was four each, and for transfer, it was three, so the total maximum score for the posttest was eleven. We did an ANOVA on the learning outcome to test for statistical significance.

**Cognitive load.** To measure cognitive load (CR), the differentiated cognitive load questionnaire was used [66]. It contains two items for ICL, three items for ECL, and three items for GCL, all measured as self-reports on a 7-point Likert scale from 1 *strongly disagree* to 7 *strongly agree*. To measure internal consistency, the Cronbach Alpha was calculated [65]. Reliability was  $\alpha = .66$  for ICL,  $\alpha = .81$  for ECL, and  $\alpha = .77$  for GCL. As for learning outcome, we tested for significance with an ANOVA.

**System usability** To quantitatively measure the system usability, the System Usability Scale (SUS) was used [67]. This scale is a self-report measurement consisting of 10 items related to the usability of *exploRNN* (e.g., *I found the system very cumbersome to use*). Responses to the items were given on a 7-point Likert scale ranging from 1 *strongly disagree* to 7 *strongly agree*. The internal consistency of this scale was  $\alpha = .74$ .

**Qualitative questions** For an impression of the quality of the learning material, further questions were implemented. Three open-ended questions were related to likeability (*What about the learning experience did you like especially, what did you not like?*), missing functionality (*Was there something you would have liked to do but could not?*), and additional comments (*Other remarks*). (MJ). For liking (*I would like to use this learning material*) and recommendation (*I would recommend this learning material to my friends*) of the material, two items could be rated on a 5-point Likert scale from *very unlikely* to *very likely*. Content (*How was the quality of the content?*) and design (*How was the design of the learning experience?*) could be rated with 0–5 stars.

### 9.3 Results

In the following, we present the results of the user study.

**Descriptive data.** The analysis of the descriptive statistics showed that subjects of the text group and the *exploRNN* group did not differ in most of the variables. T tests (variances were equal for all variables) with respect to age ( $p = .33$ ), gender (*exploRNN* group 21% females, text group 16.67% females) ( $p = .74$ ), MSLQ ( $p = .11$ ), self-efficacy (MSLQ) ( $p = .16$ ) and duration ( $p = .79$ ) revealed no significant

differences. Motivation (MSLQ) showed a significant *t* test ( $p = .02$ ), indicating that learners in the text group had a significantly higher score. Descriptive data for all variables per condition are given in Table 1.

To analyze whether prior knowledge and MSLQ should be used as covariates, we conducted a correlation analysis with learning outcomes and cognitive load. Significant correlations could be found for GCL with the MSLQ ( $r = .37$ ,  $p = .024$ ) and for the recall of the posttest with the MSLQ ( $r = .44$ ,  $p = .006$ ). Therefore, they were included as a covariate in the following calculations concerning GCL and recall. No other significant correlations for the potential covariates could be found.

**Learning outcome.** Against our hypotheses, we found a significant difference regarding recall ( $F(1, 34) = 3.91$ ,  $p = .028$ ,  $\eta_p^2 = .103$ ) in favor of the text group but not for comprehension ( $F < 1$ , n.s.) or transfer ( $F < 1$ , n.s.).

**Cognitive load.** Contrary to our expectations, we found a significant difference between text and *exploRNN* group for ICL ( $F(1, 34) = 3.85$ ,  $p = .029$ ,  $\eta_p^2 = .099$ ). ECL showed the hypothesized effect: The *exploRNN* group showed a significant lower score than the text group ( $F(1, 34) = 4.33$ ,  $p = .023$ ,  $\eta_p^2 = .113$ ). Against our hypothesis, GCL was not significantly higher in the *exploRNN* group than in the text group ( $F < 1$ , n.s.).

**System usability.** The SUS questionnaire indicates an *excellent* usability ( $M = 84.47$ ,  $SD = 9.45$ ) [68]. Participants also rated our approach as significantly more likable ( $F(1, 30) = 10.52$ ,  $p = .003$ ,  $\eta_p^2 = .260$ ), more recommendable ( $F(1, 30) = 11.75$ ,  $p = .002$ ,  $\eta_p^2 = .281$ ), and better designed ( $F(1, 30) = 20.711$ ,  $p < .001$ ,  $\eta_p^2 = .408$ ) compared to the learning text.

**Qualitative questions.** We also got some qualitative feedback in our free-form fields. Participants liked our introduction, which apparently made it easy for them to get started with *exploRNN* *the tutorial was nice and the platform was easy to use*. They also mentioned that the graphical support of these textual explanations was helpful for them to form a mental image of the setting: *the mental bridge the graphical presentation helped build was helpful in memorizing and understanding*. Some participants said that they did not remember specific names, as it was not important during the usage of *exploRNN*: *I later did not remember the name of the algorithm that was used, since it was not important during the usage of the tool*. Some participants asked for something similar for other types of networks, e.g., *I would like to have similar resources to cover other topics from the basics such as MLPs up to advanced topics and more complicated kinds of networks*. As described in Sect. 8, we only support a limited set of interactions, which some participants commented on, e.g., *[I missed] changing the activation function of the LSTM gates*.

**Table 1** Means and standard deviations for our study results separated by groups for all variables

Variable	Text		<i>exploRNN</i>	
	M	SD	M	SD
Duratiton (min):	43.6	26.78	40.87	35.67
Age (years):	24.94	2.96	24.05	2.51
Pre-T: (%)	83.14	13.86	80.0	14.57
Post-T: (%)	68.48	20.46	65.47	14.41
Post-T recall*: (%)	79.75	26.30	63.25	29.64
Post-T comp.: (%)	65.00	51.54	65.50	37.26
Post-T trans.: (%)	58.00	69.54	68.33	41.46
ICL*: (%)	61.51	18.51	48.82	20.51
ECL*: (%)	48.14	23.32	37.09	15.52
GCL: (%)	77.51	13.94	75.94	17.13
MSLQ*: (Max: 7)	5.17	0.95	4.56	1.30
SUS: (Max: 100)	–	–	84.47	9.45
Qualit. Like*: (Max: 5)	3.27	1.03	4.35	0.86
Qualit. Recomm.*: (Max: 5)	3.00	1.20	4.24	0.83
Qualit. Content: (Max: 5)	4.00	0.76	4.35	0.61
Qualit. Design*: (Max: 5)	3.20	0.94	4.47	0.62

Numbers annotated with \* indicate a significant difference between the two conditions

## 9.4 Discussion

In the following, we will refer back to the hypotheses we had before conducting the study and discuss the study outcome. *Learning outcome* We looked at both recall and understanding regarding the learning outcome. In contrast to (H1), we found that learners in the text group showed significantly better results for recall. While we found no significant differences between the groups regarding comprehension (H2) and transfer (H3) the results are interesting nonetheless. Although not significant, the descriptive statistics indicate that the score for transfer is about 10% higher for *exploRNN* compared to text (DU). This could be a first indication that learning environments such as *exploRNN* can help learners build a deeper understanding of the subject compared to learning with classic text. However, significant results and further research are needed to support this statement. Compared to recall, these results may indicate that while learners are better at learning terms by heart (surface learning) when they learn with text than with *exploRNN*.

A possible explanation for the better recall performance in the text group could be that learners have more experience with text-reading strategies [69]. This might help with the complex terms explained here, as learners might find it easier to find information that was previously presented [70]. Thus, designing ways to easily retrieve previously presented information could be an interesting direction of future research for such interactive explorables. Another possibility is that learn-

ing with an interactive environment, which is affirmative and provides information step by step, might infuse the illusion of knowing [71]. Learners may think that after a few experiments in *exploRNN*, they have acquired enough knowledge, while there is still much more to explore and learn. In the text group, it is immediately clear to the learner when the text is finished. On the contrary, *exploRNN* can require user initiative for information acquisition. As the learning experience was self-controlled, participants could decide for themselves when to go from the learning content to the posttest. Referring to the illusion of knowing, learners might have felt too competent as they experienced this more guided experience. However, even though learners may be able to transfer what they have learned to other application areas, they may be missing important basic terminology that was presented in the learning material to reflect their knowledge gain in a classical learning test.

*Cognitive load* Against (H4), the *exploRNN* group showed significantly lower ICL than the text group. The perceived difficulty of the learning material is 12.71% lower for *exploRNN* even though the text content was identical in both conditions. This suggests that *exploRNN* makes the learning material appear easier (CR). The reason for this could be that the content is presented step by step in the tutorial of *exploRNN*, instead of all at once as in the text condition [56].

The results regarding (H5) are consistent with our assumptions. With a large effect size, the *exploRNN* group showed lower ECL than the text group (CR). Therefore, *exploRNN* reduced the extraneous cognitive load compared to the text content, although the content was the same in both conditions and there were no unnecessary figures or information in the text. Combined with lower ICL, more cognitive capacity remains for GCL, which is important for learning.

For GCL, we did not find the significant difference we hypothesized in (H6). Although ICL and ECL indicate that more cognitive capacity should be free in the *exploRNN* condition, participants did not invest that cognitive capacity into GCL. This could be because there might already be high investment in GCL in both groups. Another explanation could be that since participants in the text group perceived the learning content as more difficult, they might have invested more GCL to compensate for said difficulty.

*System usability* The results of the SUS questionnaire indicate that our system is easy to use. This supports our proposed visualization and interaction design and shows that our design choice of creating an interactive environment as a learning experience on RNNs matches our target audience well. Additionally, as participants rated our approach as significantly more likeable, recommendable, and better designed, users are likely to experience more joy, and be more motivated when learning (MJ). In combination with the reduced cognitive load *exploRNN* inflicts on users, we hope that this could result in a larger number of users willing to



spend their time learning and more time spent learning per user. In turn, we think that this might outweigh the advantage in some areas of learning outcome with the more familiar text-based learning environment. Further longitudinal studies on NN learning systems need to be done to investigate this in more detail.

*Qualitative feedback.* The open feedback forms also provided interesting insights. In general, participants seemed to like *exploRNN* as a learning experience and even asked for similar interfaces in other contexts. Furthermore, the amount of information and our onboarding process seemed to make *exploRNN* easily usable. Most of the criticism was related to limitations regarding the freedom of interaction, which we deliberately implemented to provide an overview rather than in-depth technical details. Future work might reveal how both an overview and full depth could be combined in NN learning environments. We also learned why recall might be better in the text condition. As participants mentioned, they did not feel they needed to memorize specific terms to be able to use RNNs. This seems natural, as when programming or using RNNs in the wild, remembering specific terms is also not essential as they can be searched for. On the other hand, transfer tends to be much more important when tasked with solving real problems.

For learning material where transfer is important (DU), as in recurrent networks, our descriptive results suggest that interactive visualizations such as *exploRNN* might be helpful. Additionally, the lower cognitive load and higher perceived likeability of our interactive environment might result in more learners spending more time with *exploRNN*. Although we extensively evaluated *exploRNN* in this study, it remains to be seen whether our insights are transferable to other learning environments. If so, the development of future explorables could be much better informed, indicating what is important, what could be discarded, and what needs to be improved on. While this study provides first insights into the effectiveness of such educational NN exploration environments, we hope that similar evaluations of other applications can broaden our insights.

## 10 Conclusions and future work

This paper presents the first interactive learning environment specifically designed for RNNs. We propose a new visualization approach for inspecting RNNs where different levels of granularity are employed. To inform our visualization design, we first introduced educational objectives for this setting. Based on these objectives, we identified design challenges, which we tackle in the proposed learning environment. We hope that this process can be helpful for the development of future interactive learning environments.

Subsequently, we tested the learning outcome, cognitive load, and usability of this learning environment in an empirical study. Our study is the first quantitative evaluation of an interactive NN learning interface and, as such, provides helpful insights and directions for future work. The results of the user study indicate that while the raw learning outcome was not improved compared to conventional methods, *exploRNN* makes learning easier and more fun since cognitive load was significantly reduced by *exploRNN*, while subjective likeability was significantly improved. Based on these findings, we propose to specifically design interactive NN learning environments so that cognitive load is reduced. For broadly accessible education, *exploRNN* can be used in any modern browser at <https://mi-pages.informatik.uni-ulm.de/explornn/>.

As mentioned, more user studies for similar educational explorables could further advance the field and better inform future visualization designs. One such possible research direction is the suspected advantage of the text condition for going back to previously presented information. Here, eye-tracking studies and novel interaction designs might provide new insights. Additionally, it would be interesting to investigate whether such systems indeed lead to more voluntary learning and how that affects learning outcome.

**Acknowledgements** This work was funded by the Carl-Zeiss Scholarship for Ph.D. students. The datasets generated during and analyzed during the presented study are available from the corresponding author on reasonable request.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
2. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
3. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with con-



- volution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
4. Kahng, M., Thorat, N., Chau, D.H.P., Viégas, F.B., Wattenberg, M.: Gan lab: understanding complex deep generative models using interactive visual experimentation. *IEEE Trans. Vis. Comput. Graph.* **25**(1), 1–11 (2018)
  5. Norton, A.P., Qi, Y.: Adversarial-playground: A visualization suite showing how adversarial examples fool deep learning. In: 2017 IEEE Symposium on Visualization for Cyber Security (VizSec), pp. 1–4. IEEE (2017)
  6. Karpathy, A.: ConvnetJS mnist demo. <https://cs.stanford.edu/people/karpathy/convnetjs/demo/mnist.html> (2020)
  7. Smilkov, D., Carter, S., Sculley, D., Viégas, F. B., Wattenberg, M.: Direct-manipulation visualization of deep networks. [arXiv:1708.03788](https://arxiv.org/abs/1708.03788) (2017)
  8. Graves, A., Mohamed, A.-R., Hinton, G.: Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6645–6649. IEEE (2013)
  9. Sak, H., Senior, A.W., Beaufays, F.: Long short-term memory recurrent neural network architectures for large scale acoustic modeling (2014)
  10. Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., Schmidhuber, J.: A novel connectionist system for unconstrained handwriting recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(5), 855–868 (2008)
  11. Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K. et al.: Google’s neural machine translation system: bridging the gap between human and machine translation. [arXiv:1609.08144](https://arxiv.org/abs/1609.08144) (2016)
  12. Sweller, J.: Cognitive load theory. In: *Psychology of Learning and Motivation*, vol. 55, pp. 37–76. Elsevier (2011)
  13. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
  14. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **5**(2), 157–166 (1994)
  15. Pascanu, R., Mikolov, T., Bengio, Y.: On the difficulty of training recurrent neural networks. In: International Conference on Machine Learning, pp. 1310–1318 (2013)
  16. Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734 (2014)
  17. Weiss, G., Goldberg, Y., Yahav, E.: On the practical computational power of finite precision RNNs for language recognition. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 740–745 (2018)
  18. Britz, D., Goldie, A., Luong, M.-T., Le, Q.: Massive exploration of neural machine translation architectures. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1442–1451 (2017)
  19. Hundhausen, C.D., Brown, J.L.: What you see is what you code: a “live” algorithm development and visualization environment for novice learners. *J. Vis. Lang. Comput.* **18**(1), 22–47 (2007)
  20. Hundhausen, C.D., Douglas, S.A., Stasko, J.T.: A meta-study of algorithm visualization effectiveness. *J. Vis. Lang. Comput.* **13**(3), 259–290 (2002)
  21. Schweitzer, D., Brown, W.: Interactive visualization for the active learning classroom. In: Proceedings of the 38th SIGCSE Technical Symposium on Computer Science Education, pp. 208–212 (2007)
  22. Guo, P.J.: Online python tutor: embeddable web-based program visualization for CS education. In: Proceeding of the 44th ACM Technical Symposium on Computer Science Education, pp. 579–584 (2013)
  23. Guo, P.J., White, J., Zanelatto, R.: “Codechella: multi-user program visualizations for real-time tutoring and collaborative learning. In: 2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), pp. 79–87. IEEE (2015)
  24. Drosos, I., Barik, T., Guo, P. J., DeLine, R., Gulwani, S.: Wrex: A unified programming-by-example interaction for synthesizing readable code for data scientists. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2020)
  25. Ynnerman, A., Löwgren, J., Tibell, L.: Explorantion: a new science communication paradigm. *IEEE Comput. Graph. Appl.* **38**(3), 13–20 (2018)
  26. Harley, A.W.: An interactive node-link visualization of convolutional neural networks. In: International Symposium on Visual Computing, pp. 867–877. Springer (2015)
  27. Chung, S., Suh, S., Park, C., Kang, K., Choo, J., Kwon, B.C.: Revacnn: real-time visual analytics for convolutional neural network. In: KDD 16 Workshop on Interactive Data Exploration and Analytics (2016)
  28. Bloom, B.S., et al.: Taxonomy of educational objectives. vol. 1: Cognitive domain, vol. 20, p. 24. McKay, New York (1956)
  29. Karpathy, A., Johnson, J., Fei-Fei, L.: Visualizing and understanding recurrent networks. [arXiv:1506.02078](https://arxiv.org/abs/1506.02078) (2015)
  30. Strobel, H., Gehrmann, S., Pfister, H., Rush, A.M.: Lstmvis: a tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE Trans. Visual Comput. Graph.* **24**(1), 667–676 (2017)
  31. Strobel, H., Gehrmann, S., Behrisch, M., Perer, A., Pfister, H., Rush, A.M.: Seq2seq-vis: a visual debugging tool for sequence-to-sequence models. *IEEE Trans. Visual Comput. Graph.* **25**(1), 353–363 (2018)
  32. Ming, Y., Cao, S., Zhang, R., Li, Z., Chen, Y., Song, Y., Qu, H.: “Understanding hidden memories of recurrent neural networks,” in *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, pp. 13–24 (2017)
  33. Cashman, D., Patterson, G., Mosca, A., Watts, N., Robinson, S., Chang, R.: Rnnbow: visualizing learning via backpropagation gradients in rnns. *IEEE Comput. Graph. Appl.* **38**(6), 39–50 (2018)
  34. Shen, Q., Wu, Y., Jiang, Y., Zeng, W., LAU, A. K., Vianova, A., Qu, H.: Visual interpretation of recurrent neural network on multi-dimensional time-series forecast. *IEEE Transactions on Visualization and Computer Graphics* (2020)
  35. Garcia, R., Weiskopf, D.: Inner-process visualization of hidden states in recurrent neural networks. In: Proceedings of the 13th International Symposium on Visual Information Communication and Interaction, pp. 1–5 (2020)
  36. Soltau, H., Liao, H., Sak, H.: Neural speech recognizer: acoustic-to-word LSTM model for large vocabulary speech recognition. [arXiv:1610.09975](https://arxiv.org/abs/1610.09975) (2016)
  37. Manh, H., Alaghand, G.: Scene-LSTM: A model for human trajectory prediction. [arXiv:1808.04018](https://arxiv.org/abs/1808.04018) (2018)
  38. Graves, A., Jaitly, N., Mohamed, A.-R.: Hybrid speech recognition with deep bidirectional LSTM. In: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 273–278. IEEE (2013)
  39. Park, S.H., Kim, B., Kang, C.M., Chung, C.C., Choi, J.W.: Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder–decoder architecture. In: 2018 IEEE Intelligent Vehicles Symposium (IV), pp. 1672–1678. IEEE (2018)
  40. Gleicher, M.: Explainers: expert explorations with crafted projections. *IEEE Trans. Visual Comput. Graph.* **19**(12), 2042–2051 (2013)
  41. Munzner, T.: Visualization Analysis and Design. CRC Press, Boca Radon (2014)

42. Beck, F., Burch, M., Diehl, S., Weiskopf, D.: The state of the art in visualizing dynamic graphs. In: EuroVis (STARs). Citeseer (2014)
43. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings 1996 IEEE Symposium on Visual Languages, pp. 336–343. IEEE (1996)
44. Amadiou, F., Mariné, C., Laimay, C.: The attention-guiding effect and cognitive load in the comprehension of animations. *Comput. Hum. Behav.* **27**(1), 36–40 (2011)
45. De Koning, B.B., Tabbers, H.K., Rikers, R.M., Paas, F.: Attention guidance in learning from a complex animation: Seeing is understanding? *Learn. Instr.* **20**(2), 111–122 (2010)
46. Khomenko, V., Shyshkov, O., Radyvonenko, O., Bokhan, K.: Accelerating recurrent neural network training using sequence bucketing and multi-gpu data parallelization. In: 2016 IEEE First International Conference on Data Stream Mining and Processing (DSMP). IEEE, pp. 100–103 (2016)
47. Zhu, H., Akrou, M., Zheng, B., Pelegris, A., Jayarajan, A., Phanshayee, A., Schroeder, B., Pekhimenko, G.: Benchmarking and analyzing deep neural network training. In: 2018 IEEE International Symposium on Workload Characterization (IISWC). IEEE, pp. 88–100 (2018)
48. Peters, O.: Digital learning environments: new possibilities and opportunities. *Int. Rev. Res. Open Distrib. Learning* **1**(1), 1–19 (2000)
49. Smilkov, D., Thorat, N., Assogba, Y., Yuan, A., Kreeger, N., Yu, P., Zhang, K., Cai, S., Nielsen, E., Soergel, D. et al.: Tensorflow.js: Machine learning for the web and beyond. [arXiv:1901.05350](https://arxiv.org/abs/1901.05350) (2019)
50. Bartram, L.: Perceptual and interpretative properties of motion for information visualization. In: Proceedings of the 1997 Workshop on New Paradigms in Information Visualization and Manipulation, pp. 3–7 (1997)
51. Chevalier, F., Riche, N. H., Plaisant, C., Chalbi, A., Hurter, C.: Animations 25 years later: new roles and opportunities. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 280–287 (2016)
52. Robertson, G., Fernandez, R., Fisher, D., Lee, B., Stasko, J.: Effectiveness of animation in trend visualization. *IEEE Trans. Visual Comput. Graph.* **14**(6), 1325–1332 (2008)
53. Lee, S., Kim, S.-H., Hung, Y.-H., Lam, H., Kang, Y.-A., Yi, J.S.: How do people make sense of unfamiliar visualizations?: A grounded model of novice’s information visualization sensemaking. *IEEE Trans. Visual Comput. Graph.* **22**(1), 499–508 (2015)
54. Berthold, K., Renkl, A.: Instructional aids to support a conceptual understanding of multiple representations. *J. Educ. Psychol.* **101**(1), 70 (2009)
55. Kang, H., Plaisant, C., Shneiderman, B.: New approaches to help users get started with visual interfaces: multi-layered interfaces and integrated initial guidance. In: Proceedings of the 2003 Annual National Conference on Digital Government Research, pp. 1–6 (2003)
56. Brachten, F., Brünker, F., Frick, N.R., Ross, B., Stieglitz, S.: On the ability of virtual agents to decrease cognitive load: an experimental study. *Inf. Syst. e-Bus. Manag.* **18**(2), 187–207 (2020)
57. Ding, W.: Draw convnet. [https://github.com/gwding/draw\\_convnet](https://github.com/gwding/draw_convnet) (2018)
58. Uchida, Y.: Convnet drawer. <https://github.com/yu4u/convnet-drawer> (2019)
59. Bauerle, A., Van Onzenoodt, C., Ropinski, T.: Net2vis-a visual grammar for automatically generating publication-ready CNN architecture visualizations. *IEEE Trans. Visual. Comput. Graph.* (2021)
60. Olah, C.: Understanding LSTM networks. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (2015)
61. Elmqvist, N., Fekete, J.-D.: Hierarchical aggregation for information visualization: overview, techniques, and design guidelines. *IEEE Trans. Visual Comput. Graph.* **16**(3), 439–454 (2010)
62. Jenny, B., Kelso, N.V.: Color design for the color vision impaired. *Cartograph. Perspect.* **58**, 61–67 (2007)
63. McCarthy, L.: P5.js. <https://p5js.org/> (2020)
64. Pintrich, P.R. et al.: A manual for the use of the motivated strategies for learning questionnaire (MSLQ) (1991)
65. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. *Psychometrika* **16**(3), 297–334 (1951)
66. Klepsch, M., Schmitz, F., Seufert, T.: Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Front. Psychol.* **8**, 1997 (2017)
67. Brooke, J., et al.: Sus-a quick and dirty usability scale. *Usability Eval. Industry* **189**(194), 4–7 (1996)
68. Bangor, A., Kortum, P., Miller, J.: Determining what individual SUS scores mean: adding an adjective rating scale. *J. Usability Stud.* **4**(3), 114–123 (2009)
69. Kolić-Vrhovec, S., Bajšanski, I., Rončević Zubković, B.: The role of reading strategies in scientific text comprehension and academic achievement of university students. *Rev. Psychol.* **18**(2), 81–90 (2011)
70. Kürschner, C., Seufert, T., Hauck, G., Schnotz, W., Eid, M.: Konstruktion visuell-räumlicher repräsentationen beim hör-und leseverstehen. *Z. Psychol./J. Psychol.* **214**(3), 117–132 (2006)
71. Glenberg, A.M., Wilkinson, A.C., Epstein, W.: The illusion of knowing: failure in the self-assessment of comprehension. *Mem. Cognit.* **10**(6), 597–602 (1982)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Alex Bäuerle** received the master’s degree in media informatics from Ulm University in 2017 and is now working at a research associate at the Visual Computing Group at Ulm University. His current research interests are the visualization of neural networks to generate better understanding around these techniques.



**Patrick Albus** completed his master's degree in psychology at Ulm University with a focus on Work and Organizational Psychology and Clinical Psychology. Since September 2018, Patrick Albus has been working as a research assistant at the Institute of Psychology and Education, Department Learning and Instruction, conducting research in the field of instructional design and virtual reality.



**Raphael Störk** studied media informatics at the Ulm University and received his bachelor's degree in 2020. His main interests involve the development of learning applications and the concepts of neural networks and other forms of artificial intelligence.



**Tina Seufert** is Professor and head of the Department for Learning and Instruction since 2008. Her research interests deal with theoretical and practical questions on how to foster learning processes by instructional means. These include the areas of multiple representations, virtual reality, aptitude-treatment interactions and cognitive load. She also develops and tests pedagogical concepts for various learning scenarios such as e-learning, HCI, and continuing education.



**Timo Ropinski** is a professor at Ulm University, where he is heading the Visual Computing Group. Before moving to Ulm, he was Professor in Interactive Visualization at Linköping University in Sweden, where he was heading the Scientific Visualization Group. He has received his PhD in computer science in 2004 from the University of Münster, where he has also completed his Habilitation in 2009.