Window	Start timestamp	Duration	# Events	Event	Frequency	Contribution
1	Day 1 22:58	115 min	11	Read response time Read transfer size Write transfer size	1 5 5	0.00 0.00 0.00
2	Day 2 6:04	105 min	1	Read response time	1	0.00
3	Day 2 20:19	390 min	8	Read response time Read transfer size Write transfer size	1 3 4	<0.01 <0.01 0.04
4	Day 3 19:59	195 min	6	Read transfer size Write transfer size	3 3	<0.01 0.06
5	Day 4 23:00	70 min	3	Read transfer size Write transfer size	1 2	<0.01 0.06
6	Day 5 6:15	120 min	2	Read response time	2	0.015
7	Day 5 22:55	20min	2	Read response time	2	0.02
8	Day 6 22:56	20 min	1	Read transfer size	1	<0.01
9	Day 7 23:01	15 min	2	Read transfer size	2	0.01
10	Day 8 6:02	125 min	3	Disk utilization	3	0.00
11	Day 8 22:57	20 min	9	Read transfer size Write transfer size	5 4	0.05 0.16
12	Day 9 23:12	65 min	3	Read response time	3	0.06
13	Day 11 20:28	205 min	4	Write response time	4	0.18
14	Day 13 4:08	35 min	6	Read response time Write response time	4 2	0.1 0.34
15	Day 14 22:59	15 min	8	Read response time Peak backend write response time Write response time	3 2 3	0.12 0.8 0.63

Table 3: Weights associated with 69 anomalous events clustered in 15 windows for a storage device predicted to fail.

Table 5 we highlight only the occurence of peak backend write response time events for a device that does not fail. Even though these events occur multiple times, they are too distant from the start of the prediction window, thus their weights are low (i.e., 0.025 and 0.04). Additionally, no write response time events occur in the same windows. Therefore, our model is sensitive to events frequency, their exact timestamps and their co-occurence with correlated event types.

Limitations

As this work is only in its initial phase, it is by no means complete. We identify a few limitations to be addressed next.

Data set size – Given the low failure rate in our use case, ideally we should use series of KPIs collected over months or even years to increase the size of the minority class and improve prediction accuracy. For this paper, we wanted to show a proof of concept of how we can provide explanations with LSTMs, while we are continuing to collect data.

Multivariate anomalous events – So far, we focused on single KPI thresholds to identify anomalous events. We plan to use anomaly detection algorithms to explore multivariate time series and identify complex anomalous patterns instead. First, we expect to uncover seasonal patterns of behavior that are not obvious to an expert. Second, we believe model performance will increase and the sizes of clusters will reduce, making it easier for a domain expert to use the explanations provided. However, our primary goal was to show how we can marry LSTMs with built-in explainability to provide interpretable predictions.

Advanced DNN models – We are aware that prediction accuracy can be improved by using more advanced DNN models, such as bidirectional LSTMs, RNNs with gated recurrent units (**?**) or even combinations of LSTMs and CNNs (**?**). In the latter, fully convolutional blocks and LSTM units are run in parallel and their outputs are passed to a softmax classification layer. We plan to try out more advanced models going forward, while keeping the attention layer in place.

Related Work

While there exists a lot of work around explainable models for images and text, little attention has been been given to explaining models based on temporal data, namely time or event series. On the one hand, post-hoc approaches aim at explaining a model's prediction after the event, which means they should be agnostic and applicable on any type of data. On the other hand, ante-hoc methods incorporate explainability directly into the black-box model, which implies they are tailored to the underlying model and data.

Post-hoc approaches aim to provide local explanations for specific decisions, rather than attempting to explain the whole system behavior. One of the most representative examples for classification in recent years is LIME (**?**). The approach is simple: generate an explanation by approximating the underlying model by an interpretable one (e.g., a linear model with a only a few non-zero coefficients), learned on perturbations of the original instance. Typical perturbations can be removing words or hiding parts of an image. A similar model-agnostic approach is BETA (**?**), which opti-

Window	Start timestamp	Duration	# Events	Event	Frequency	Contribution
1	Day 1 10:07	65 min	1	Disk utilization	1	0.00
2	Day 3 10:02	395 min	9	Disk utilization Read transfer size	5 4	0.00 0.00
3	Day 4 2:07	195 min	2	Read transfer size	2	<0.01
4	Day 6 8:37	15 min	2	Read response time	2	<0.01
5	Day 10 15:07	25 min	1	Read response time	1	0.04
6	Day 11 18:22	65 min	2	Read transfer size	2	0.05
7	Day 13 2:47	135 min	5	Read response time Disk utilization	2 3	0.04 0.02

Table 4: Weights associated with 22 anomalous events clustered in 7 windows for a storage device predicted not to fail.

Window	Start timestamp	Duration	# Events	Event	Frequency	Contribution
1	Day 2 15:17	35 min	5	Peak backend write response time Read response time	2 3	0.05 0.00
2	Day 5 12:02	105 min	2	Peak backend write response time	2	0.06

Table 5: Highlighted weights associated with 5 peak backend write response time anomalous events clustered in 2 windows for a storage device predicted not to fail.

mizes for fidelity to the black-box model and interpretability of the explanation. (**?**) focuses on pixel-wise decomposition of nonlinear classifiers, which allows to visualize contributions of single pixels to predictions for kernel-based classifiers. (**?**) extracts explanations from latent factor recommendation systems by training association rules on the output of a matrix factorization black-box model. All approaches have been applied on text and images, but are not built to take into consideration temporal progressions in time or event series. *Ante-hoc approaches* are interpretable by design (**?**). Typical examples include decision trees, decision sets (**?**; **?**), fuzzy inference models (**?**) or additive models (**?**). However, none of these fit temporal data well.

The vast majority of explainable models for time series target their classification. (**?**) propose grammar-based decision trees to classify heterogeneous time series. (**?**; **?**) extract interpretable features from series, expressed as local shapelets, while (**?**) learn such shapelets via stochastic gradient learning and use them for early classification. In (**?**), the authors propose reversible and irreversible explainable tweaking, where given a time series and an opaque classifier, the objective is to find the minimum number of changes to the time series such that the classifier changes its decision. Closest to our problem is the method proposed in (**?**). There, the objective is to predict a future neural event based on a sequence of previously occurred events. Current approaches are mostly concerned with time-independent sequences, in which the actual time span between events is irrelevant and the difference between events is the difference between their order positions in the sequence. The authors extract and use the information provided by the time span between events in an RNN-based model to achieve some accuracy gain over baseline models. We also opt for an RNN architecture, but we additionally incorporate attention mechanisms (**?**) into the network to quantify how much an anomalous event contributed to a predicted critical incident.

Conclusions

Predictive modeling based on temporal data is key in many domains, from healthcare to IT and industries, particularly when it is concerned with critical incidents, such as failures. Providing explanations for these predictions is crucial, as it enables experts to gain trust in AI-powered models and take into consideration their outputs in the decision process. State of the art explainable models mostly focus on images and text and are not easily applicable to time or event series. We propose a deep learning approach that takes into consideration the irregularity and frequency of anomalous events extracted from time series and uses attention mechanisms to aggregate context information of these events in order to quantify how much information from each event flows into the network. A preliminary evaluation on 266081 events collected from real world storage environments shows that our approach is comparable in accuracy with traditional LSTMs, while at the same time being able to quantify the contribution of each past event recorded to a failure prediction.

Quis asperiores assumenda in non at aspernatur doloremque, eum at placeat adipisci, atque alias veritatis nostrum iure assumenda excepturi doloribus, non iusto maxime hic distinctio omnis nisi ipsa, deleniti sit ut recusandae obcaecati doloremque expedita?A repellendus hic consectetur natus suscipit est, dolorem quod debitis temporibus doloribus atque iste dolore nobis at, mollitia enim possimus porro dolores saepe quis nemo fugit sapiente aperiam incidunt.Assumenda voluptates temporibus nisi voluptatum et quis aspernatur unde, ipsa dolor fugit, dolores libero et vel tempora autem magnam perspiciatis labore hic, perferendis iste est numquam nostrum modi possimus recusandae maxime tempora hic.Deserunt cum quasi eos atque facilis minus, maxime eius sint quibusdam quam nulla deleniti soluta, pariatur laborum blanditiis facere rem dicta officia aut aliquid soluta, explicabo excepturi minima vitae fugiat enim voluptate.Reiciendis maiores quia veritatis ratione itaque quisquam perspiciatis consequuntur repellat, explicabo ipsum itaque tempore consequuntur accusamus illo distinctio mollitia expedita fuga optio.Nesciunt nihil