How much should you trust your AI?

A discussion of the trustworthiness of various AI-assisted methods

Florian Sobieczky, Anna-Christina Glock, Manuela Geiß, Aleksei Koretnikov, Michael Roßbory

– SCCH: Presentation given the HEAL Round Table on 30 Jan 2023

scch {
software
competence
center
hagenberg
}

Agenda

Change Point Detection (CPD)
Feature Extraction and Feature Selection
Explainable AI (XAI)

We discuss the issue of trustworthiness of AI in the **context of time series data:**

- **How to quantify trustworthiness?**
- **What does it mean in the case of time series?**
- **What *is* an AI's user's trust?**

How to define: Trust in AI?

scch { }

It is possible to produce pictures which are classified with high confidence as an image of an object which are completely unrecognizable by humans[1].

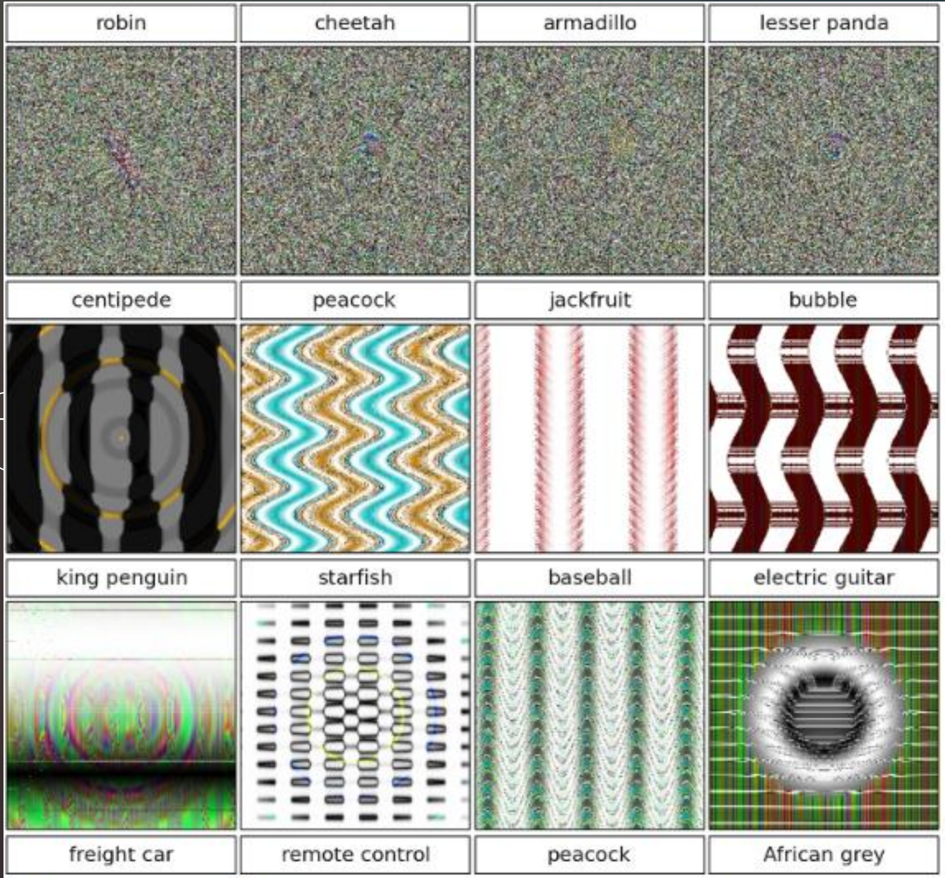


Figure 1. Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with $\geq$ 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Images are either directly (*top*) or indirectly (*bottom*) encoded.

AlexNet - convolutional neural network, trained on 1.3 Million images: ILSVRC 2012 ImageNet dataset

From: [1] Nguyen et. Al.““Deep Neural Networks Are Easily Fooled: High Confidence Predictions for unrecognize. Images.”

How to define: Trust in AI?

scch { }

How should we measure 'trustworthiness' of Machine Learning Models?

It is possible to produce pictures which are classified with high confidence as an image of an object which are completely unrecognizable by humans[1].

AlexNet - convolutional neural network, trained on 1.3 Million images: ILSVRC 2012 ImageNet dataset

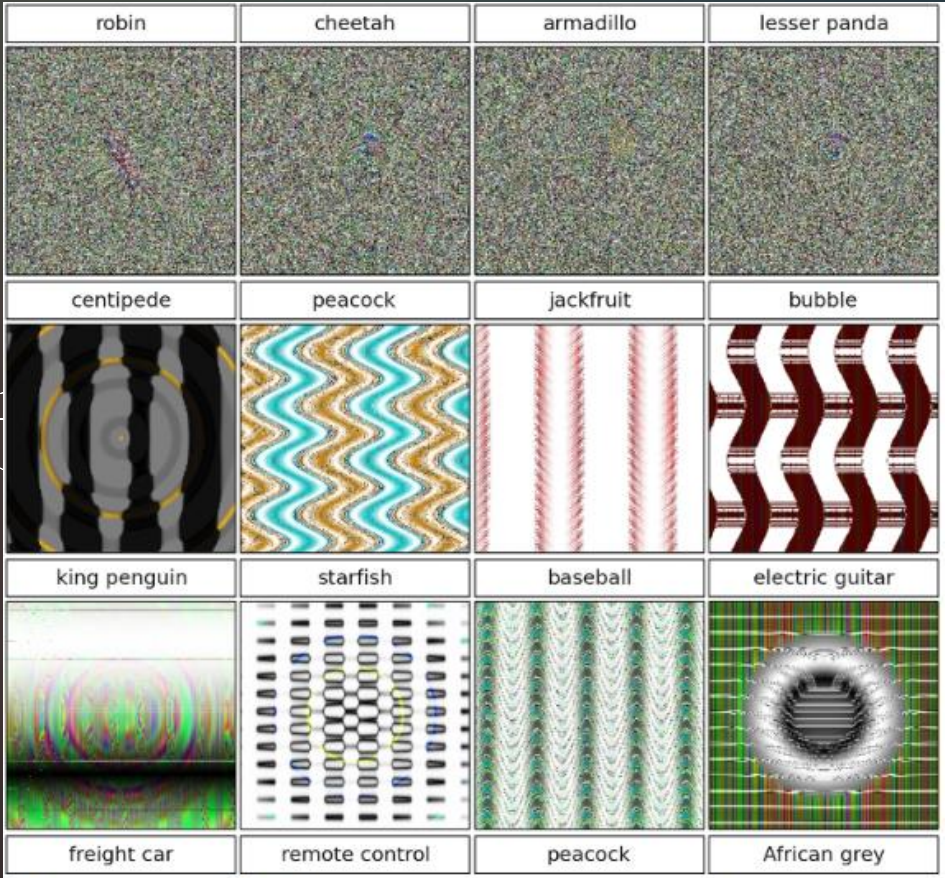


Figure 1. Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with $\geq$ 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Images are either directly (*top*) or indirectly (*bottom*) encoded.

From: [1] Nguyen et. Al.""Deep Neural Networks Are Easily Fooled: High Confidence Predictions for unrecognize. Images."

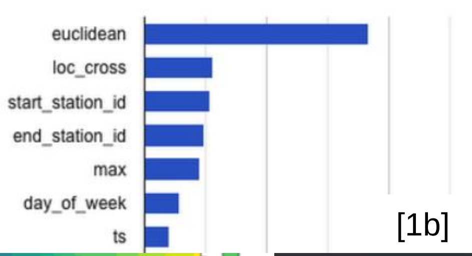
Explainability:
Different Concepts

scch { }

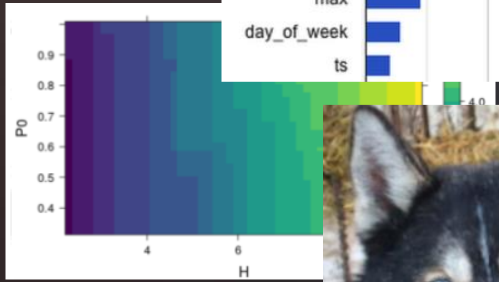
The correct prediction alone doesn‘t help indicate the general patters.

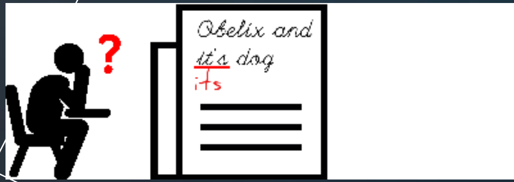
- Global vs. Local
- Model-Agnostic vs. Model-Specific
- Post-Hoc vs. Intrinsic …..

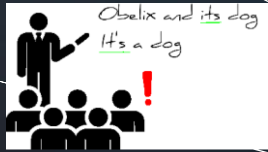
- Salience Maps (GradCam)
- PCP / ICE / ALE Plots
- Local Surrogates (LIME)
- Game Theory (Shapley Values)
- …..



[1b]



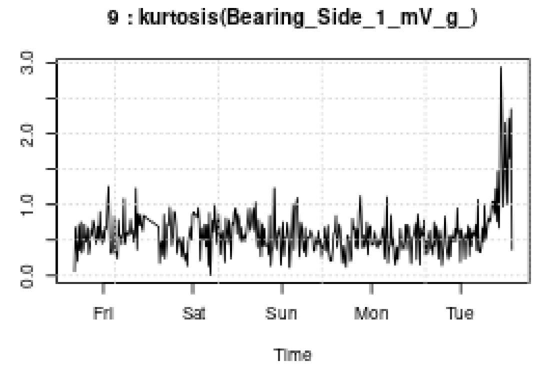


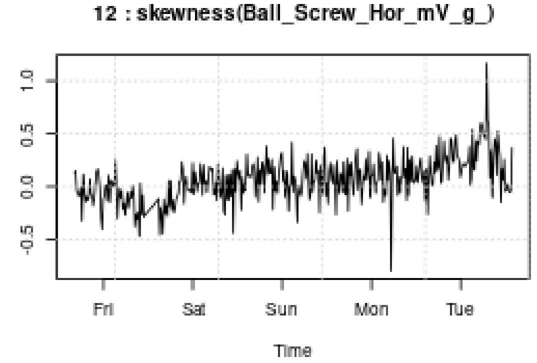


Data Examples

- Precision Machining
- Heart Rate Monitoring
- Production Data

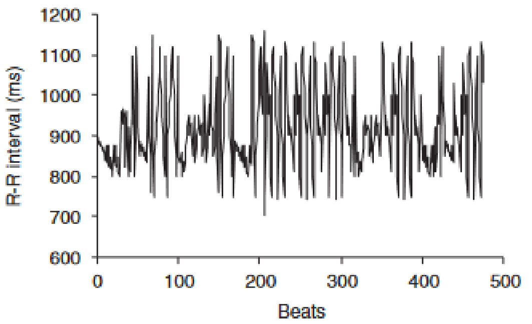
Ball Screws: The rotating motion of a cylndrical screw is translated into a longitudinal motion[2]. E.g.: High Prec. Positioning in CRC machining

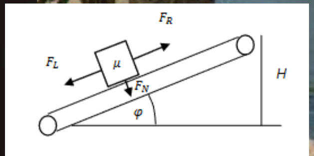




Heart Rate Variability (HRV)

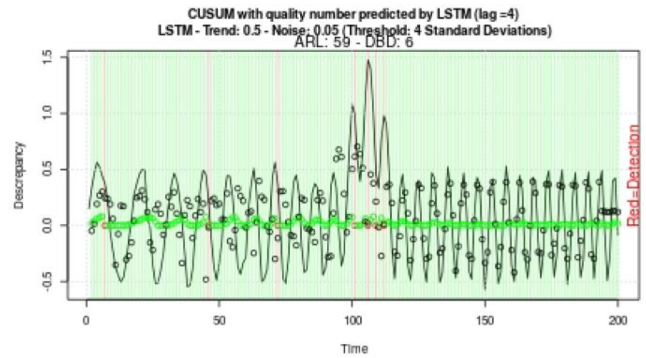
- Monitoring beat-to-beat variations
- Time-between-two-beats[3]
- R-R plots, ECG output, Power Spectra …
- Kaggle – Data[2]: Standard Deviation

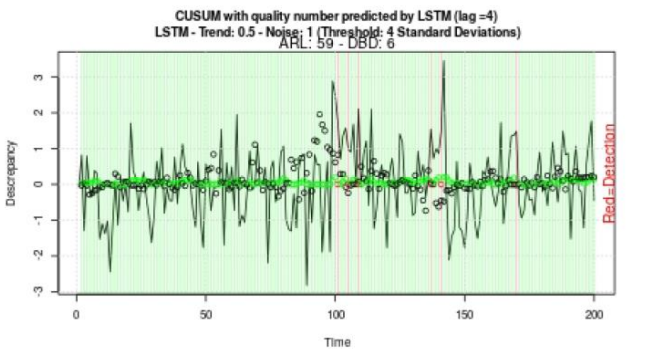




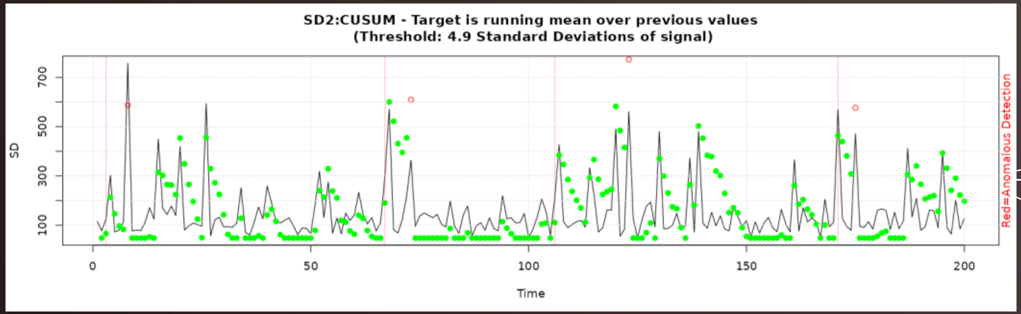
Production Process Data

- Quality Monitoring
- Different CPs
- Detecting CPs in presence of non-harmonic oscillations [4]





Change Point Detection



Biometrika (1970), 57, 1, p. 1
Printed in Great Britain

1

Inference about the change-point in a sequence of random variables

BY DAVID V. HINKLEY
Imperial College

SUMMARY

Inference is considered about the point in a sequence of random variables at which the probability distribution changes. In particular, we examine a normal distribution with changing mean. The asymptotic distribution of the maximum likelihood estimate is derived and also the asymptotic distribution of the likelihood ratio statistic for testing hypotheses about the change-point. These asymptotic distributions are compared with some finite sample empirical distributions.

$$X_t = \theta_0(t) + \epsilon_t \quad (t = 1, \ldots, \tau),$$
$$X_t = \theta_1(t) + \epsilon_t \quad (t = \tau + 1, \ldots, T)$$

- ‘Structural Breaks’ in Time Series Data
- Sequential tests allow quality monitoring (Wald [11])

  i. Accept H_0 ii. Accept H_1 iii. Wait for another Data Point
- Quality for CPD: #(False Positives) & Delay until Detection

  typic. measured in Average Run Length (in- & out-of-control) *in online case*.

Online Change Point Detection: Windowed CUSUM

scch { }

Rule 1. Take samples of fixed size N at regular intervals; assign a score x_k to the kth sample and plot the cumulative score $S_n = \sum_{k=1}^{n} x_k$ on a chart:

$$\text{Take action if } S_n - \min_{0 \leqslant i < n} S_i \geqslant h. \tag{5}$$

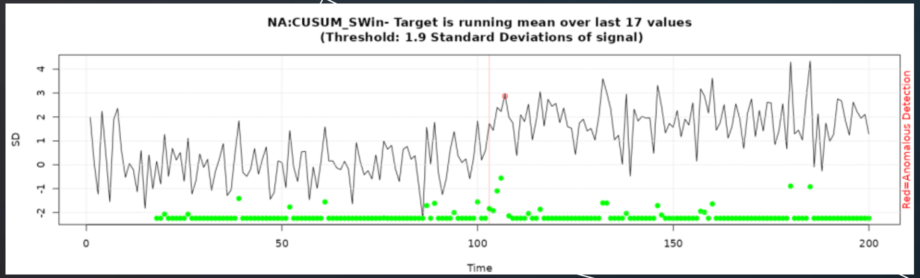
$$S_{n+1} = \max(0, S_n + X_n - \theta_n - k)$$

$$H_0: \mu_1 = \cdots = \mu_n \quad H_1: \mu_1 = \cdots = \mu_{r-1} < \mu_r = \cdots = \mu_n$$

$$\frac{\prod_{k=1}^{r-1} p_0(y_k) \cdot \prod_{k=r}^{n} p_1(y_k)}{\prod_{k=1}^{n} p_0(y_k)} = \prod_{k=r}^{n} \frac{p_1(y_k)}{p_0(y_k)}$$

$$\text{where :} \quad p_i(y) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{\left(-\frac{(y-\mu_i)^2}{2\sigma^2}\right)} \quad (i=0,1)$$

- Page's seminal paper[12] introduced the CUSUM method
- Lorden[13] (Asymp.Optim.), Barnard[14] (V-mask), Hinkley[15], Basseville&Benveniste[16], Bayesian Online CPD[17]
- Cumulative sum of fluctuations of signal X_n beyond target θ_n
- k is the 'allowance' – only consider exceedance beyond $\theta_n + k$
- 'Sequential': Only data up to time n taken into consideration and decision about outcome is potentially postponed



Classical Offline Change Point Detection

scch { }

- Uses Batches of Data
- Segments the time-line
- Data after proposed CP has equal eligibility

Successful *Offline* CPD Methods[5, 6]:

BINSEG[7] 1974 : Divide time-range recursively into dyadic sub-intervals
PELT[8] 2012: Cost-Function related partitioning – (precise, running long)
CPOP[9] 2019: Finds Changes in Slope
ECP[10] : Calculates discrepancy scores between data windows
AMOC[11] 1970: Maximum Likelihood estimator of CP of only one CP.
BOCPD[17] 2007: Bayesian *online* CPD

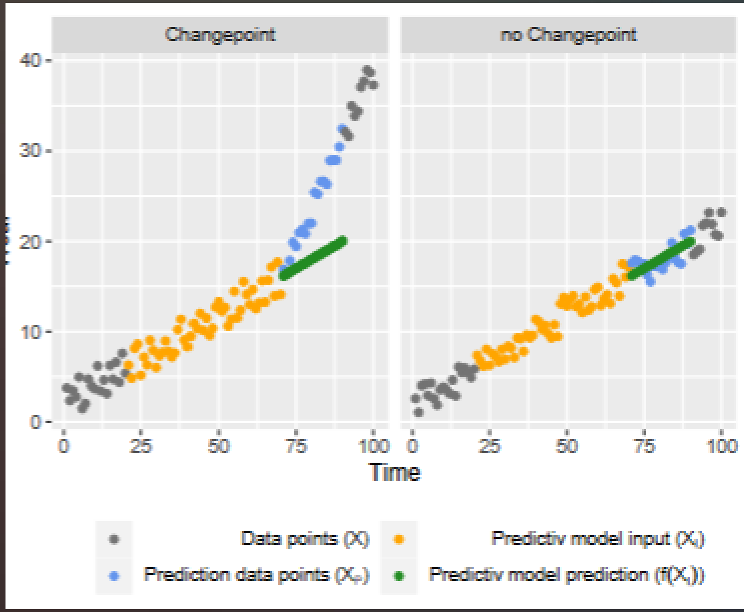
	Univariate			
	Default		Best	
	Cover	F1	Cover	F1
AMOC	0.702	0.704	0.746	0.799
BINSEG	**0.706**	**0.744**	0.780	0.856
BOCPD	0.636	0.690	**0.789**	**0.880**
BOCPDMS	0.633	0.507	0.744	0.620
CPNP	0.535	0.607	0.552	0.666
ECP	0.523	0.598	0.720	0.797
KCPA	0.062	0.111	0.626	0.683
PELT	0.689	0.710	0.725	0.787
PROPHET	0.540	0.488	0.576	0.534
RBOCPDMS	0.629	0.447		
RFPOP	0.392	0.499	0.414	0.531
SEGNEIGH	0.676	0.676	0.784	0.855
WBS	0.330	0.412	0.428	0.533
ZERO	0.583	0.669	0.579	0.662

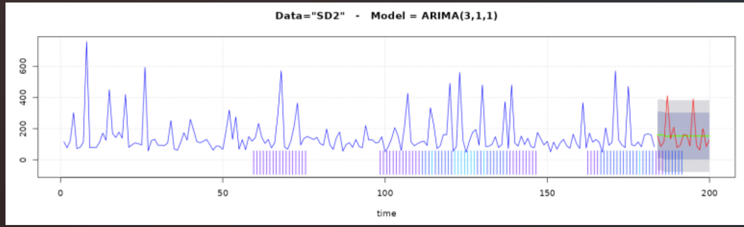
Observation: Online CPD methods are able to compete with offline ones in terms of precision and recall.

Classical Methods can be assisted by AI

scch { }

Predict & Compare:
- Prediction Model uses data on window as input (‘input window‘)
- Predicts on subsequent window (‘prediction window‘) of width b
- Comparison of true data with prediction reveals change point





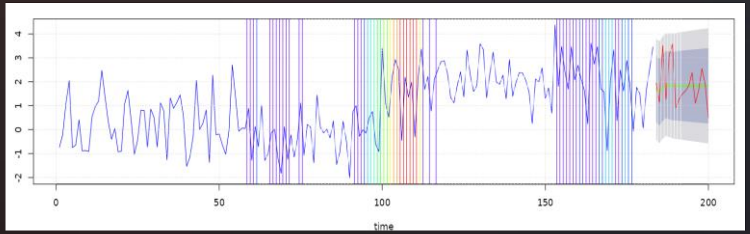


Table 1: This table shows the minimal and maximal value of the false-positive count (Fpc) and the average run length percent (ArlP) per method and dataset.

Method	Dataset	min Fp	max Fp	min ArlP	max ArlP
ARIMA CUSUM	1	0	20	30.50	82.45
ARIMA CUSUM	2	0	17	18.76	70.75
ARIMA CUSUM	3	0	10	1.11	63.18
ARIMA CUSUM	5	0	375	14.70	78.02
ARIMA CUSUM	7	0	17	21.48	88.80
Baseline	1	93	2465	14.51	86.45
Baseline	2	98	2465	0.67	41.94
Baseline	3	98	2465	0.35	19.83
Baseline	5	98	2465	1.55	6.01
Baseline	7	0	2465	0.00	82.57
Bayesian	1				
Bayesian	2	218	1091	1.95	7.66
Bayesian	3	102	255	3.99	10.10
Bayesian	5	581	2135	0.10	0.10
Bayesian	7	81	338	0.06	0.63
BFAST	1	0	118	2.52	98.44
BFAST	2	5	265	0.72	64.17
BFAST	3	0	18	0.16	80.66
BFAST	5	39	2282	0.01	55.68
BFAST	7	15	97	0.00	75.33
CUSUM	1	2	1044	2.52	70.46
CUSUM	2	0	884	0.72	38.97
CUSUM	3	0	185	0.06	21.38
CUSUM	5	0	8101	0.13	77.65
CUSUM	7	0	830	0.00	78.50
LSTM CUSUM	1	4	236	2.52	78.46
LSTM CUSUM	2	0	619	0.41	92.82
LSTM CUSUM	3	0	123	0.19	92.03
LSTM CUSUM	5	0	1235	1.94	77.48
LSTM CUSUM	7	0	151	0.61	91.63

- Predict & Compare[4a]
- Target θ_n is replaced by Prediction of Learning model:

$$\theta_n = \hat{f}\left(X_{[0,t-b\,]}, t\right)$$

- ‘CPD in Presence of Trends’

From [4a]: It shows that P&C competes well with other state-of-the art Online CPD methods (CUSUM[], BOCPD[], BFAST[]).

Assisting CPD with AI – trustworthy?

scch { }

Assumption: The signal X_n *is*

$$X_n = f_n + W_n + I_n$$

where

f_n is the trend,

W_n is some stationary noise,

I_n is the CP function (step, ramp…).

$$H_0 : \; I_n = 0, \qquad H_1 : \; I_n \neq 0$$

- If under H_0 prediction of learning model is close to ground truth with high probability

$$P[|\hat{f}(\boldsymbol{X}_{[\mathbf{0},t-b]}) - \boldsymbol{f}_{[t-b+1,t]}| > \boldsymbol{\epsilon}] \leq \boldsymbol{\delta}$$

(e.g., in the PAC-sense [19] $\epsilon, \delta > 0$ small)

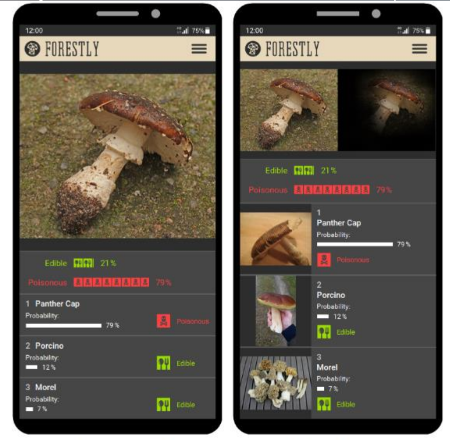
then P&M will not confuse CP with trend.

'Mushroom Picking in Austria assisted by AI' [20]: 'Trust cannot be a goal in itself' and '… avoid overtrust ….'

Quoted there: '*Trust in AI is 'the willingness of users to be vulnerable to the actions of some automated system to achieve some goal.*' [20, 21]

Result: If 'explanations' are provided for classification result, the user performs significantly better in deciding correctly about edibility of mushroom under the assistance of AI.

-> User's AI – literacy improves AI-assisted User Decisions



(a) Plain interface (b) XAI interface

B. Leichtmann, C. Humer, A. Hinterreiter, M. Streit, M. Mara: "Effects of Explainable Artificial Intelligence on Trust and Human Behavior in a High-Risk Decision Task,", https://doi.org/10.31219/osf.io/n4w6u

Assisting CPD with AI – trustworthy?

scch { }

Assumption: The signal X_n is

$$X_n = f_n + W_n + I_n$$

where

f_n is the trend,

W_n is some stationary noise,

I_n is the CP function (step, ramp…).

$$H_0 : I_n = 0, \qquad H_1 : I_n \neq 0$$

- If under H_0 prediction of learning model is close to ground truth with high probability

$$P[|\hat{f}(\boldsymbol{X}_{[0,t-b]}) - \boldsymbol{f}_{[t-b+1,t]}| > \boldsymbol{\epsilon}] \leq \boldsymbol{\delta}$$

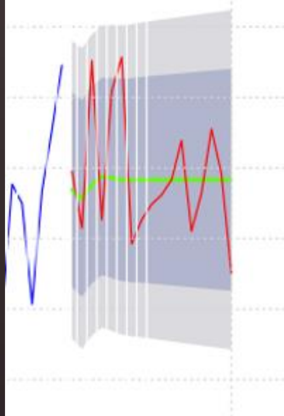
(e.g., in the PAC-sense [19] $\epsilon, \delta > 0$ small)

then P&M will not confuse CP with trend.

PAC means 'Probably Almost Correct':

- Concept in Statistical Learning Theory
- gives lower bound (1-δ) for the probability of the predictive model $\hat{f}: X \to Y$ to be ϵ-correct,
- bounds are uniform over all possible random measures (distributions) that may occur in nature.

D. Haussler generalized SLT beyond classification[19].



δ is the 'confidence parameter': It expresses the probability of the prediction being off by more than ϵ.

Def. *Confidence* is a bound for the probability of $\hat{f}$ being inaccurate. As such it is a natural candidate for the quantitative measure of the trustworthiness of a ML model.

Assisting CPD with AI – more so if explainable

scch { }

Assumption: The signal X_n *is*

$$X_n = f_n + W_n + I_n$$

where

f_n is the trend,

W_n is some stationary noise,

I_n is the CP function (step, ramp…).

$$H_0 : \; I_n = 0, \qquad H_1 : \; I_n \neq 0$$

- If under H_0 prediction of learning model is close to ground truth with high probability

$$P[|\hat{f}(\boldsymbol{X}_{[0,t-b]}) - \boldsymbol{f}_{[t-b+1,t]}| > \boldsymbol{\epsilon}] \leq \boldsymbol{\delta}$$

(e.g., in the PAC-sense [19] $\epsilon, \delta > 0$ small)

then P&M will not confuse CP with trend.

‘Mushroom Picking in Austria assisted by AI’ [20]: ‘Trust cannot be a goal in itself’ and ‘… avoid overtrust ….’.

Quoted there: ‘*Trust in AI is ‘the willingness of users to be vulnerable to the actions of some automated system to achieve some goal.*’ [20, 21]
Result: If ‘explanations’ are provided for classification result, the user performs significantly better in deciding correctly about edibility of mushroom under the assistance of AI.
-> User’s AI – literacy improves AI-assisted User Decisions

Def.: If a predictive model‘s $\hat{f}(\cdot) : \boldsymbol{R}^t \rightarrow \boldsymbol{R}^b$ user estimates the **confidence parameter** $\delta = \delta(t, \epsilon)$ at time t for a given accuracy parameter ϵ, then this expresses the user‘s trust in it.

-> PAC = Probably Almost Correct – Learning theory by Valiant, Vapnik, Pollard, Haussler[19].

Assisting CPD with AI – trustworthy?

scch { }

Assumption: The signal X_n *is*

$$X_n = f_n + W_n + I_n$$

where

f_n is the trend,

W_n is some stationary noise,

I_n is the CP function (step, ramp…).

$$H_0 : \ I_n = 0, \qquad H_1 \ : \ I_n \neq 0$$

- If under H_0 prediction of learning model is close to ground truth with high probability

  $$P[|\hat{f}(\boldsymbol{X}_{[\boldsymbol{0},\boldsymbol{t}-\boldsymbol{b}\,]}) - \boldsymbol{f}_{[\boldsymbol{t}-\boldsymbol{b}+\boldsymbol{1},\boldsymbol{t}]}\,| > \boldsymbol{\epsilon}] \leq \boldsymbol{\delta}$$

  (e.g., in the PAC-sense [19] $\epsilon, \delta > 0$ small)

  then P&M will not confuse CP with trend.

‘Mushroom Picking in Austria assisted by AI’ [20]: ‘Trust cannot be a goal in itself’ and ‘… avoid overtrust ….’

Quoted there: ‘*Trust in AI is ‘the willingness of users to be vulnerable to the actions of some automated system to achieve some goal.*’ [20, 21]

Result: If ‘explanations’ are provided for classification result, the user performs significantly better in deciding correctly about edibility of mushroom under the assistance of AI.

-> User’s AI – literacy improves AI-assisted User Decisions

- AI = Artificial Intelligence
- UI = User’s Intelligence
- User’s trust: $\hat{\delta}$ (user-estimated δ)
- -> Def.: Trust := Estimated Confidence

Trust, Explainability, Accuracy

scch { }

$$m \geqslant \frac{M^2}{2\varepsilon^2}\left(\ln|\mathbf{F}| + \ln\frac{2}{\delta}\right)$$

- $\boldsymbol{F}$ is the finite H
- $|\hat{f}| < M$ for all $\hat{f} \in \boldsymbol{F}$
- Under this condition
  $R < \delta$.

- AI = Artificial Intelligence
- UI = User's Intelligence
- User's trust: $\hat{\delta}$, user-estimated
- -> Def.: Trust = Confidence

D. Haussler's Learning generalisation of the PAC model [19] to regression problems involves:

- a hypothesis space $H = \{\hat{f}: X \to Y\}$
- the usual risk function $r = E[l(Y, \hat{f}(X))]$ (empirical & ideal)
- a regret function $L(P, \hat{f})$
- a learning method $A: X^m \times Y^m \to \hat{f}$
- a "big L"-risk (expected regret) $R = \int L(P, A(X,Y))dP^m(X,Y)$ where the expectation is taken over the m-dim. training set.
- E.g.: $L(P, A) = I_{<r^*+\epsilon}(\hat{r})$ ➔ $R = P[|\hat{r} - r^*| > \epsilon]$

Learning bounds limit the minimum training set size m.
(these are uniform bounds of the expected regret over all P.)

Trust, Explainability, Accuracy

scch { }

$$m \geqslant \frac{M^2}{2\varepsilon^2}\left(\ln|\mathbf{F}| + \ln\frac{2}{\delta}\right)$$

- $\boldsymbol{F}$ is the finite H
- $|\hat{f}| < M$ for all $\hat{f} \in \boldsymbol{F}$
- Under this condition $R < \delta$.

- AI = Artificial Intelligence
- UI = User's Intelligence
- User's trust: $\hat{\delta}$, user-estimated
- -> Def.: Trust = Confidence

D. Haussler's Learning generalisation of the PAC model [19] to regression problems involves:

- a hypothesis space $H = \{\hat{f}: X \to Y\}$
- the usual risk function $r = E[l(Y, \hat{f}(X))]$ (empirical & ideal)
- a regret function $L(P, \hat{f})$
- a learning method $A: X^m \times Y^m \to \hat{f}$
- a "big L"-risk (expected regret) $R = \int L(P, A(X,Y))dP^m(X,Y)$ where the expectation is taken over the m-dim. training set.
- E.g.: $L(P,A) = I_{<r^*+\epsilon}(\hat{r})$ ➔ $R = P[|\hat{r} - r^*| > \epsilon]$

$R < \delta$: High trust in a learning method is small expected regret.

Trust, Explainability, Accuracy

scch { }

$$m \geqslant \frac{M^2}{2\varepsilon^2}\left(\ln|\mathbf{F}| + \ln\frac{2}{\delta}\right)$$

- $\boldsymbol{F}$ is the finite H
- $|f| < M$ for all $f \in \boldsymbol{F}$
- Under this condition $R < \delta$.

Haussler‘s Learning generalisation of the PAC model [19] to regression problems (also: density estimation and parameter estimation) involves

- a hypothesis space $H = \{h: X \to Y\}$
- the usual risk function $r = E[l(Y, h(X))]$ (empirical & ideal)
- a regret function $L(P, h)$ which is small when the empirical risk is close to the ideal risk
- a “big L“-risk (expected regret) $R = \int L\big(P, A(X,Y)\big) dP^m(X,Y)$ where the expectation is taken over the m-dim. training set.

Learning bounds limit the minimum training set size m.
(these are uniform bounds of the expected regret over all P.)

Conclusion: Trust in a learning method is small expected regret.

- AI = Artificial Intelligence
- UI = User’s Intelligence
- User’s trust: $\hat{\delta}$, user-estimated
- -> Def.: Trust = Confidence

Conclusion

scch { }

Our questions:

- **How to quantify trustworthiness? -> Confidence**
- **What does it mean in the case of time series? -> Uniform Bounds of probability to be correct**
- **What *is* an AI's user's trust? -> The estimate of the corresponding Confidence**

- **Trustworthiness** is confidence δ (in dependence of accuracy ϵ).
- **Trust** is estimated confidence $\hat{\delta}$.
- The statistical learning theory provides rigorous terms for defining **trust in learning methods** to be estimated bound of expected regret $R(P)$.

What about AlexNet?

Confidence estimates are determined by the DNN – Human estimate differs!

5. Conclusion

We have demonstrated that discriminative DNN models are easily fooled in that they classify many unrecognizable images with near-certainty as members of a recognizable class. Two different ways of encoding evolutionary algorithms produce two qualitatively different types of unrecognizable "fooling images", and gradient ascent produces a third. That DNNs see these objects as near-perfect examples of recognizable images sheds light on remaining differences between the way DNNs and humans recognize objects, raising questions about the true generalization capabilities of DNNs and the potential for costly exploits of solutions that use DNNs.

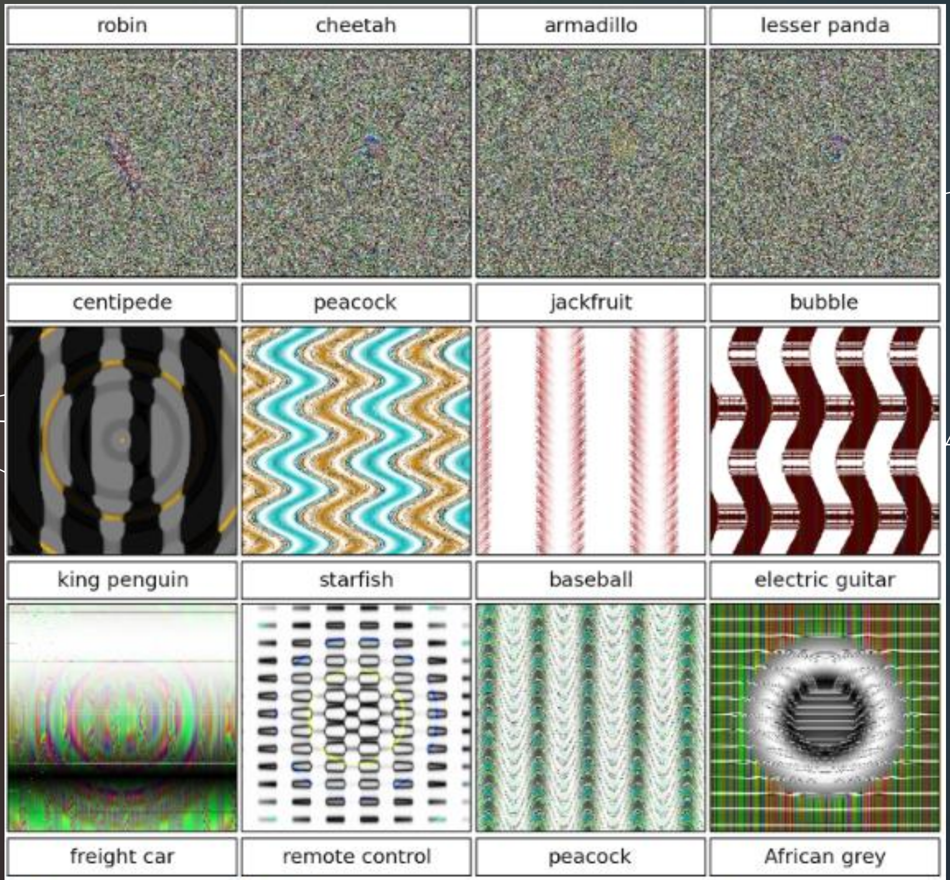


Figure 1. Evolved images that are unrecognizable to humans, but that state-of-the-art DNNs trained on ImageNet believe with $\geq$ 99.6% certainty to be a familiar object. This result highlights differences between how DNNs and humans recognize objects. Images are either directly (*top*) or indirectly (*bottom*) encoded.

AlexNet - convolutional neural network, trained on 1.3 Million images: ILSVRC 2012 ImageNet dataset

From: [1] Nguyen et. Al.""Deep Neural Networks Are Easily Fooled: High Confidence Predictions for unrecognize. Images."

Literature

scch { }

[1]] Nguyen, Anh, Jason Yosinski, and Jeff Clune. “Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images.” arXiv, April 2, 2015. http://arxiv.org/abs/1412.1897.
[1a] PredMAIn – Interreg Project: European Interreg Austria-Czech Republic project ”PredMAIn (ATCZ27)
[1b] Amy Unruh, Sarah Robinson: **Explaining model predictions on structured data, Google Cloud, AI & Machine Learning, 2020** https://cloud.google.com/blog/products/ai-machine-learning/explaining-model-predictions-structured-data/
[2] Anand Saurav: Kaggle: “Heart Rate Prediction“, https://www.kaggle.com/datasets/saurav9786/heart-rate-prediction
[3] Juul Achten, Asker E. Jeukendrup“Heart Rate Monitoring Applications and Limitations”, Sports Med 2003; 33 (7): 517-538, 0112-1642/03/0007-0517/$30.00/0
[4] S. Mahmoud, J. Martinez-Gil, B. Freudenthaler, P. Praher, A. Girkinger. Deep Learning Rule for Efficient Changepoint Detection in the Presence of non-linear Trends. 1st International Workshop on Time Ordered Data (ProTime2021), DOI https://doi.org/10.1007/978-3-030-87101-7_18, 9, 2021.
[4a] A.-C. Glock, F. Sobieczky, T. Wopelka, J. Fürnkranz, P. Filzmoser, M. Jech: “AI assisted CPD for heterogeneous data“, (Preprint)
[5] A.-C. Glock, F. Sobieczky, M. Jech. Detection of anomalous events in the wear-behaviour of continuously recorded sliding friction pairs. ÖTG-Tagungsbericht, ISBN 978-3-901657-62-7, pages 30-40, 11, 2019.
[6] G.J.-J. van den Burg, C.K.I. Williams: “An Evaluation of Change Point Detection Algorithms”, arXiv:2003.06222 [cs, stat]
[7] A. J. Scott and M. Knott. A cluster analysis method for grouping means in the analysis of variance. Biometrics, 30(3):507–512, 1974.
[8] R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. Journal of the American Statistical Association, 107(500):1590–1598, 2012.
[9] P. Fearnhead and G. Rigaill. Changepoint detection in the presence of outliers. Journal of the American Statistical Association, 114(525):169–183, 2019. - cpop – R-library
[10] D. S. Matteson and N. A. James. A nonparametric approach for multiple change point analysis of multivariate data. Journal of the American Statistical Association, 109(505):334–345, 2014.
[11] Wald, A., and J. Wolfowitz. “Optimum Character of the Sequential Probability Ratio Test.” The Annals of Mathematical Statistics 19, no. 3 (1948): 326–39.
[12] S. Page, “Continuous Inspection Schemes’, Biometrika, Vol. 41, No. 1/2 (1954), pp. 100-115
[13] G. Lorden: “Nearly optimal sequential tests for finite parameter values”, The Annals of Statistics, Vol. 5, No. 1, pp. 1-17, 1977
[14] Barnard, G. A. “Control Charts and Stochastic Processes.” Journal of the Royal Statistical Society: Series B (Methodological) 21, no. 2 (July 1959): 239–57.
[15] D. V. Hinkley. Inference about the change-point in a sequence of random variables. Biometrika, 57(1):1–17, 4 1970.
[16] Basseville, Michèle, and A. Benveniste, eds. Detection of Abrupt Changes in Signals and Dynamical Systems. Lecture Notes in Control and Information Sciences 77. Berlin: Springer, 1986.
[17] Adams, Ryan Prescott, and David J. C. MacKay. “Bayesian Online Changepoint Detection.” ArXiv:0710.3742 [Stat], October 19, 2007. http://arxiv.org/abs/0710.3742.
[18] Verbesselt, Jan, Rob Hyndman, Glenn Newnham, and Darius Culvenor. “Detecting Trend and Seasonal Changes in Satellite Image Time Series.” Remote Sensing of Environment 114, no. 1 (2010): 106–15.
[19] D. Haussler: ‘Decision Theoretic Generalizations of the PAC Model for Neural Net and Other Learning Applications’, INFORMATION AND COMPUTATION 100, 78-150 (1992)
[20] B. Leichtmann, C. Humer, A. Hinterreiter, , M. Streit, M. Mara: “Effects of Explainable Artificial Intelligence on Trust and Human Behavior in a High-Risk Decision Task,”, https://doi.org/10.31219/osf.io/n4w6u
[21] Hannibal, G., Weiss, A., & Charisi, V. (2021). “The robot may not notice my discomfort” – examining the experience of vulnerability for trust in human-robot interaction. 2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 704–711. https://doi.org/10.1109/RO-MAN50785. 2021.9515513
[22] Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50–80
[23] Zhang, Y., Liao, Q. V., Bellamy, R. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. Proc. 2020 Conf- on Fairness, Accountability, and Transparency, 295–305.

! ! ! Thank you for your interest and consideration ! ! !

scch { software competence center hagenberg }

Dr. Florian Sobieczky
Area Manager Software Science
+43 50 343 839
florian.sobieczky@scch.at

Florian Kromp
Researcher, Senior Data Scientist
+43 50 343 809
florian.kromp@scch.at

Anna-Christina Glock
Researcher, Data Scientist
+43 50 343 840
anna-christina.glock@scch.at

SCCH is an initiative of

JKU JOHANNES KEPLER UNIVERSITÄT LINZ

SCCH is located in

softwarepark hagenberg

Dr. Manuela Geiß
Senior Researcher Data Science
+43 50 343 848
manuela.geiss@scch.at

Aleksei Karetnikov
Researcher, Data Scientist
+43 50 343 901
aleksei.karetnikov@scch.at

Michael Rossbory
Senior Project Manager Data Science
+43 50 343 860
michael.rossbory@scch.at

www.scch.at