

# Binaire

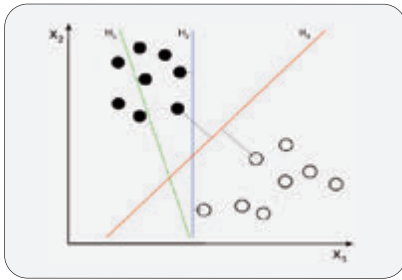
DAS MAGAZIN DES FORSCHUNGSZENTRUMS L3S

WWW.BINAIRE.DE

AUSGABE №  
10/11111100010

# MASCHINELLES LERNEN





Wie Maschinen lernen und verstehen. Die Grafik oben zeigt das Prinzip der binären Klassifikation. Das lässt sich auch auf Farbkleckse in der Malerei anwenden (Titelseite). Man sieht mehrere lineare Trenngeraden, die rote ist jeweils die bestmögliche. → Quelle: Wikipedia

### **BINÄRE KLASSIFIKATION**

... ist eine der grundlegenden Aufgaben des maschinellen Lernens. Das Ziel ist, Input in zwei Klassen einzuteilen. Bilder können zum Beispiel danach klassifiziert werden, ob sie ein Gesicht abbilden oder nicht. Dazu identifizieren wir geeignete Merkmale als Beschreibung unserer Input-Daten und stellen sie unserem ML-System zur Verfügung. Das Besondere: Wir müssen keine Regeln festlegen und kein Programm schreiben. Durch eine Vielzahl an Beispielbildern lernt das System selbständig, die Merkmale mit der jeweiligen Klasse zu assoziieren. Dabei müssen wir jedoch darauf achten, dass diese Beispiele tatsächlich die reale Welt repräsentieren, sonst spiegelt das gelernte Modell nur einen eingeschränkten Teil der Wirklichkeit wider. Das System würde in nicht-repräsentierten Situationen falsch reagieren oder zu Entscheidungen führen, die Minderheiten diskriminieren.



### **DAS FORSCHUNGS- ZENTRUM L3S**

L3S-Forscher entwickeln im Bereich **Web-Science** und **digitale Transformation** zukunftsweisende Methoden und Technologien, die einen intelligenten und nahtlosen Zugriff auf Informationen über das Web ermöglichen, Individuen und Gemeinschaften in allen Bereichen der Wissensgesellschaft vernetzen und das Internet an die reale Welt und ihre Einrichtungen anbinden. Das L3S erforscht die Auswirkungen des digitalen Wandels, um aus den Erkenntnissen Handlungsoptionen, -empfehlungen und Innovationsstrategien für die Wirtschaft, die Politik und Gesellschaft herzuleiten. Durch Forschung, Entwicklung und Beratung trägt das L3S gemeinsam mit seinen Partnern zur digitalen Transformation insbesondere in den Bereichen Mobilität, Gesundheit, Produktion und Bildung bei.



# Intelligenz mit Verantwortung

Liebe Leserin, lieber Leser,

künstliche Intelligenz und maschinelles Lernen, Herausforderungen und Möglichkeiten – das ist das Leitthema der zweiten Ausgabe unserer *Binaire*. Wohl kein anderes Thema der Informatik hat in den letzten Jahren so sehr die Öffentlichkeit bewegt wie dieses.

Künstliche Intelligenz hilft dabei, in der Medizin die Erfolgchancen einer bestimmten Behandlung für Patienten zu prognostizieren, etwa von Antibiotikagaben bei Infektionen mit multiresistenten Keimen. Autonom fahrende Autos müssen sich in der realen Welt zurechtfinden und sollen Unfälle vermeiden helfen. Intelligente Produktion macht Unternehmen weltweit flexibler, Produkte personalisierter und Maschinen zuverlässiger. Digitale Bildung ermöglicht zielgruppenspezifische Lehre und personalisiertes Lernen.

Gleichzeitig müssen wir die sozialen, wirtschaftlichen und ethischen Rahmenbedingungen und Auswirkungen von künstlicher Intelligenz erforschen, um intelligente, verlässliche und verantwortungsvolle Systeme zu schaffen, mit denen wir unsere Zukunft positiv gestalten können.

Intelligenz ohne Ethik, digitalisierte Entscheidungen ohne Erklärungen, Unfälle autonomer Fahrzeuge, die wir nicht verstehen, oder Diskriminierung durch Big Data und Algorithmen sind Aspekte, die wir in unserer Forschung aktiv angehen und ausschließen müssen.

Wir freuen uns, gemeinsam mit Ihnen unsere Zukunft gestalten zu können und Ihnen bei der Einführung in die vielfältige Welt der künstlichen Intelligenz zu helfen. Sprechen Sie uns an, wenn Sie an einem Thema besonderes Interesse haben, wir freuen uns auf Diskussionen und Projekte mit Ihnen!

Eine spannende Lektüre wünscht Ihnen



Prof. Dr. techn. Wolfgang Nejdil



ESSENZ

**DURCH FORSCHUNG,  
ENTWICKLUNG  
UND BERATUNG**

gestaltet das L3S  
gemeinsam mit seinen Partnern  
die digitalen Transformation  
insbesondere in den Bereichen:

- Intelligente Produktion
- Digitale Bildung
- Intelligente Mobilität
- Personalisierte Medizin

»Gerade der Einsatz  
von Algorithmen in der  
öffentlichen Sicherheit verdeutlicht,  
dass bereits vorab Regeln gesetzt werden  
müssen, um Diskriminierung und  
Intransparenz zu vermeiden.«

PROF. DR. TINA KRÜGEL · LL.M  
Rechtsanwältin, Juniorprofessorin für Informationsrecht  
(insbesondere Datenschutzrecht) und L3S-Mitglied

**ÜBERSICHT**

BINAIRE - AUSGABE 2 / 2018

|                         |  |            | dezimal | binär |
|-------------------------|--|------------|---------|-------|
| EDITORIAL               | Intelligenz mit Verantwortung            | → Seite 03 | •       | 11    |
| EVENTS                  | Termin   Veranstaltungsberichte          | → Seite 05 | •       | 101   |
| TITELTHEMA              | Lernende Computersysteme ...             | → Seite 08 | •       | 1000  |
| TITELTHEMA              | Wie sieht ein selbstfahrendes Auto?      | → Seite 12 | •       | 1100  |
| JURISTISCHE FRAGEN      | Rechtliche Implikationen ...             | → Seite 14 | •       | 1110  |
| INTERPRETIERBARKEIT ... | Wie entstehen Suchmaschinenrankings?     | → Seite 16 | •       | 10000 |
| WAHRNEHMUNG             | Nicht die Orientierung verlieren         | → Seite 17 | •       | 10001 |
| DATENSAMMLUNGEN         | Blick in die Vergangenheit               | → Seite 18 | •       | 10010 |
| SPRACHE IM WANDEL       | Im Netz der Meinungen                    | → Seite 20 | •       | 10100 |
| WISSENSWERTES           | Die Zahl   Zahlensalat und Einheitenbrei | → Seite 21 | •       | 10101 |
| PERSONEN UND ...        | Ankündigung   Auszeichnung               | → Seite 22 | •       | 10110 |
| IMPRESSUM               | Kontakt                                  | → Seite 23 | •       | 10111 |

## TERMIN

## CEBIT®

11. – 15. 6. 2018

**L3S auf der CEBIT**

Das L3S widmet sich auch auf dem Niedersächsischen Gemeinschaftsstand (Halle 16, Stand D18) dem Trend zu Big Data und maschinellem Lernen. CEBIT-Besucher können sich anhand interaktiver Exponate, Lernmodule und Spiele über die vielfältigen Möglichkeiten des maschinellen Lernens informieren. Das L3S stellt unter anderem das Qualifizierungsangebot der *Applied Machine Learning Academy (AMA)* vor. AMA setzt den Schwerpunkt auf den Einsatz des maschinellen Lernens in der Produktion, denn insbesondere im Mittelstand gilt der Mangel an Fachkräften als eines der größten Hemmnisse für den erfolgreichen Einstieg in die intelligente Digitalisierung.

→ <https://ama-academy.eu/>

Besucher und Wissenschaftler im Austausch auf der CEBIT.



### Wissensmanagement und -entdeckung mit Wissensgraphen

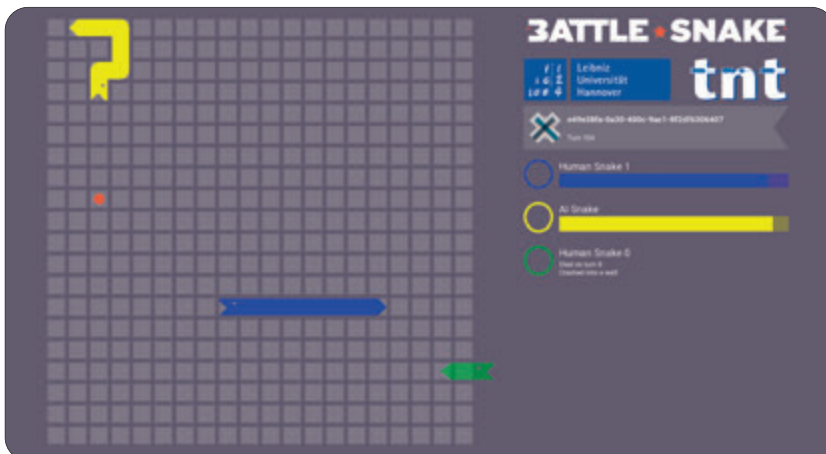
Ob im Gesundheitswesen oder in der Fertigung – Big-Data-Tools werden zu Analyse- und Management-Zwecken überall dort eingesetzt, wo große Datenmengen entstehen. Analyseergebnisse können aber auch falsch oder irreführend sein, beispielsweise aufgrund von ungenauen oder fehlenden Daten. Das L3S und die *Technische Informationsbibliothek* demonstrieren auf der CEBIT, wie Big-Data-Technologien die Datenanalyse auf eine neue Qualitätsstufe bringen können.

→ <https://www.L3S.de/de/projects/iasis>  
 → [www.iit.demokritos.gr/project/bigmedilytics](http://www.iit.demokritos.gr/project/bigmedilytics)

### Maschinelles Lernen in der wissenschaftlichen Informationsversorgung

In Zeiten von Big Data und der überwältigenden Flut an Informationen in Unternehmen werden intelligente Systeme benötigt, die die richtige Information für jedes Problem liefern können. Auf der CEBIT präsentiert das L3S das Vorzeigeprojekt *Fachinformationsdienst Pharmazie (FID)*, das mithilfe von Deep-Learning-Technologien einen umfassenden und serviceorientierten Zugang zu pharmazeutischen Informationsressourcen (Publikationen, Forschungsdatensätze, Patente, Software etc.) bereitstellt. Wissenschaftliche Produkte stehen also nicht mehr für sich allein, sondern bilden ein Netzwerk, ohne dessen tieferes Verständnis kaum noch Innovation möglich ist.

→ [www.pubpharm.de](http://www.pubpharm.de)



Am L3S-Stand haben Besucher bei einer Runde Battlesnake die Möglichkeit, die von Studierenden entwickelte KI herauszufordern und natürliche gegen künstliche Intelligenz antreten zu lassen. Die KI steuert eine virtuelle Schlange, die über ein Spielfeld navigiert werden muss. Es befinden sich stets mehrere Schlangen im Spiel und das Ziel der eigenen Schlange ist es, jene der Gegenspieler zu überleben.

## KONTAKT:

Susanne Oetzmann

Redaktion Binaire

oetzmann@L3S.de

<http://www.binaire.de>



13. 4. 2018

### Erste AMA-Regionalkonferenz

Die *Applied Machine Learning Academy (AMA)* legt los: Am 13. April 2018 fand in der *Robotation-Academy* auf dem hannoverschen Messegelände die erste Regionalkonferenz von AMA zum Thema »Maschinelles Lernen für die intelligente Produktion« statt. Sie ist der Beginn einer Veranstaltungsreihe, die sich an die Produktionswirtschaft und an Vertreter der Forschung richtet, um Erfahrungen mit dem Einsatz zukunftsweisender Digitalisierungstechnologien auszutauschen.

Auf die rund 50 Teilnehmer der Konferenz wartete ein vielfältiges Programm: Vorträge aus Forschung und Industrie, eine Führung durch die Generalfabrik in der *Robotation-Academy* mit Vorstellung der Produktion-4.0-Ansätze sowie Gelegenheit zur Diskussion.

→ <https://ama-academy.eu>



Prof. Dr. techn. Wolfgang Nejdlt begrüßt die Teilnehmer der AMA-Regionalkonferenz.



Das Wingfield-Team aus Hannover.

17. 4. 2018

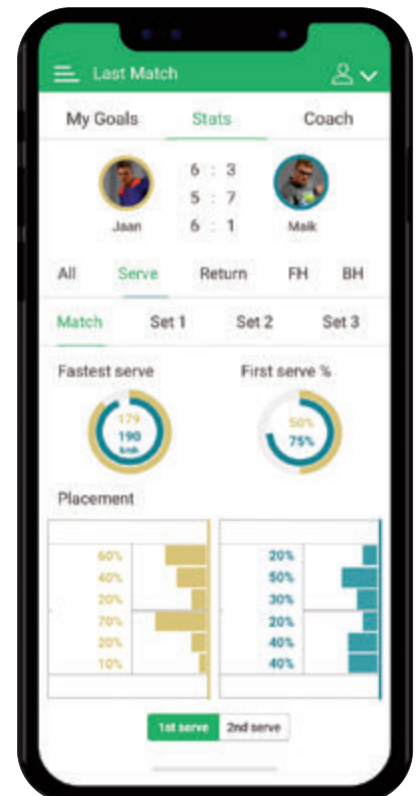
### Meetup mit Tennis-Startup

50 Teilnehmer waren beim Meetup der »Hannover Data Science«-Gruppe im L3S dabei. Das Gründerteam *Wingfield* stellte eine Beispielanwendung für künstliche neuronale Netze vor: Die Gründer klassifizieren Bewegungsabläufe bei Rückschlagsportarten. Mit moderner Bildverarbeitung und künstlicher Intelligenz arbeitet das Start-up an der Digitalisierung des Tennisplatzes. Zukünftig werden Fehlerquellen und Verbesserungspotentiale automatisch erkannt und persönliche Coaching-Tipps auf den Smartphones der Spieler bereitgestellt. Das Start-up wird seit August 2017 durch das Exist-Gründerstipendium gefördert. Der Markteintritt ist für Ende des Jahres geplant.

→ <https://www.wingfield.io>



Die Smartphone-App von *Wingfield* gibt Tennisspielern Coaching-Tipps. → Fotos: *Wingfield*





## VERANSTALTUNGSBERICHTE

20. / 21. 4. 2018

**Data-Science-Hackathon**

34 Teilnehmer haben beim Data-Science-Hackathon 2018 Durchhaltevermögen bewiesen: In einer 24-Stunden-Schicht entwickelten sie in Teams kreative Ideen, um aus vorgegebenen Datensätzen zu den Themen Mobilität, akademische Suchmaschinen und Social Streams Informationen zu extrahieren, vorherzusagen und zu visualisieren. Zum Abschluss präsentierten die Teams ihre Ergebnisse und wurden von einer fünfköpfigen Fachjury bewertet.



Teilnehmer bei der Ergebnispräsentation.

Der Hackathon wurde von der Fakultät für Elektrotechnik und Informatik in Zusammenarbeit mit dem Gründungsservice *starting business* der Leibniz Universität Hannover und der *hannover-impuls GmbH* ausgerichtet. Zu den Organisatoren zählten Prof. Eirini Ntoutsis und Prof. Ralph Ewerth vom L3S.

Die *heinekingmedia GmbH* stellte ihre Räumlichkeiten mit Dachterrassen im Zentrum Hannovers zur Verfügung und stiftete Preise im Wert von mehr als 500 €.

L3S-Wissenschaftler stellen den Roboter *SmokeBot* auf der Hannover Messe vor.

23. – 27. 4. 2018

**SmokeBot auf der Hannover Messe**

Wissenschaftler des L3S haben auf der diesjährigen Hannover Messe das Projekt *SmokeBot* vorgestellt. *Smokebot* soll die Umgebungswahrnehmung mobiler Roboter bei schlechten Sichtverhältnissen verbessern.



Die Technik kann zum Beispiel Feuerwehr oder THW in Situationen wie Tunnelkatastrophen mit dichtem Rauch und starker Hitze unterstützen. Zum Einsatz kommen Radar, Wärmebildkameras oder Gassensoren. Das Projekt des EU-Förderprogramms *Horizon 2020* läuft in Zusammenarbeit mit internationalen und nationalen Forschungspartnern und der Feuerwehr Dortmund.

→ <https://www.L3S.de/de/projects/SmokeBot>

## KONTAKT:

Susanne Oetzmann

Redaktion Binaire

oetzmann@L3S.de

<http://www.binaire.de>



Blindes Vertrauen? Unter welchen Bedingungen übergeben wir unser Leben selbstlernenden Computern? Intelligente Assistenzsysteme können Menschen künftig den Alltag sehr erleichtern. Diese Helfer werden selbstverständlich praktischer und unscheinbarer als in dieser Fotomontage sein. → Fotos: Fotolia by Adobe





**WIE LERNEN MASCHINEN?**

---

# Lernende Computersysteme verändern unser Leben

**Die Prinzipien des maschinellen Lernens** (ML) lassen sich am besten anhand eines einfachen Beispiels veranschaulichen: Nehmen wir an, wir haben einen industriellen Produktionsprozess mit Schrauben, Nägeln und Reißzwecken und möchten diese mit einem Kamera-basierten Computersystem automatisch sortieren. In einem klassischen Ansatz würde sich der Entwickler Eigenschaften und Regeln überlegen und das System entsprechend programmieren: Eine Schraube hat eine rundlichere Form mit einem Durchmesser von einem Zentimeter, die Nägel hingegen sind zwei Zentimeter lang und haben keine rundliche Form usw. Ein ML-System benötigt diese programmier-

ten Regeln nicht – es lernt sie selbständig! Im Wesentlichen kann dies auf zwei Arten geschehen: durch überwachtes beziehungsweise unüberwachtes Lernen. Wird obiges Beispiel als überwachtes Lernaufgabe behandelt, benötigt das System charakteristische Beispielobjekte als »Trainingsdaten«, verknüpft mit der Information zu welcher Klasse jedes gehört (»Nagel«, »Schraube«, »Reißzwecke«). Zudem teilen wir dem System mit, auf welche Objekteigenschaften (»Features«) es achten soll – hier beispielsweise Länge, Breite und Rundlichkeit. Hiermit lernt das ML-System dann Regeln, um eine korrekte Klassenzuordnung vorzunehmen. In unüberwachten Lernszenarien wird dem ↗



Go gilt als eines der komplexesten Brettspiele der Welt, obwohl es nur aus wenigen Regeln besteht. —> Foto:Fotolia by Adobe

System ebenfalls mitgeteilt, welche Features relevant sind. Es werden aber keine Klassen vorgegeben; oft sind diese unbekannt. Vielmehr soll das System bisher unentdeckte Muster in den Daten finden. Im Beispiel kann ein solches Verfahren helfen, **typische Produktionsfehler** zu erkennen: Fehlerhafte Bauteile werden nach Ähnlichkeit in einem Cluster gruppiert, häufig auftretende Abweichungen aufgezeigt. In interaktiven Systemen werden solche Methoden zur Visualisierung von Zusammenhängen und der Exploration hochdimensionaler Daten eingesetzt.



## REVOLUTIONÄRE FORTSCHRITTE MIT TIEFEN NEURONALEN NETZEN

ML ermöglicht die Automatisierung einfacher, eintöniger Aufgaben. Die Systeme können aber auch komplexere Aufgaben lösen und potenziell menschliche Leistung erreichen – vielleicht eines Tages sogar weit darüber hinaus? In den letzten Jahren hat es in der Mustererkennung große Fortschritte durch »tiefe«, das heißt sehr große, neuronale Netze gegeben (»Deep Learning«).

**Für Aufsehen sorgte kürzlich ein System, das den weltbesten menschlichen Spieler im Brettspiel Go geschlagen hat, mit zum Teil verblüffenden Zügen.**

In Bereichen wie Spracherkennung oder Computer-Vision hat es ebenfalls enorme Fortschritte gegeben. Deep-Learning-Systeme erreichen oder übertreffen dabei menschliche Fähigkeiten, zum Beispiel bei der Gesichtserkennung oder der Erkennung des Aufnahmeorts von Bildern.

## ZUKÜNFTIGE ENTWICKLUNG UND AUSWIRKUNGEN

Tatsächlich sind wir bereits heute an vielen Stellen mit ML-Systemen in Kontakt – oft ohne es wahrzunehmen. So werden Lösungen ähnlich dem obigen Beispiel bereits seit geraumer Zeit eingesetzt. Darüber hinaus **begegnen uns ML-Systeme tagtäglich** auch in Kaufempfehlungen im Onlineshop, personalisierten Newslettern oder Websuchergebnissen, die auch Standort und bekannte Vorlieben berücksichtigen. Es ist absehbar, dass diese Tendenz anhält: Maschinen werden fähig sein, immer komplexere Aufgaben zu übernehmen, und

hierzu immer weniger menschliche Anleitung benötigen. Forscher testen heute ML-Systeme in vielfältigen Aufgabenstellungen unter komplexen, herausfordernden Bedingungen – zum Beispiel im Straßenverkehr – und nähern ihre Performance der menschlichen an. Ein nächster Schritt wird sein, neuronale Netze für beliebige Aufgabenstellungen zu trainieren, im Sinne »allgemeiner Intelligenz«. Bisher sind Netzarchitekturen für die jeweilige Aufgabe optimiert. Je weiter Maschinen in alltägliche Bereiche eindringen, desto mehr wird sich dies **auf unser aller Leben auswirken**. ML wird uns von eintönigen Aufga-

ben befreien, unser Lernen und Arbeiten grundlegend verändern (am L3S erforscht z. B. im Projekt *Salient*) und Menschen mit gesundheitlichen Einschränkungen mehr Teilhabe ermöglichen (siehe Projekt *InclusiveOCW*), auch in Verbindung mit Robotik. Es stellt sich allerdings auch die Frage der Verantwortung: Was geschieht, wenn ein ML-System fehlerhafte Entscheidungen trifft und dadurch Menschenleben negativ beeinflusst? Damit beschäftigen sich am L3S die Experten aus der Rechtsinformatik (siehe auch die Seiten 14 und 15 dieser Ausgabe). ¶

→ <https://www.L3S.de/de/projects/inclusiveocw>  
→ <https://L3S.de/de/projects/salient>



KONTAKT:  
Prof. Dr. Ralph Ewerth  
Ewerth@L3S.de

\\ Prof. Dr. Ralph Ewerth ist seit 2016 Mitglied im L3S und leitet die Forschungsgruppe *Visual Analytics* an der TIB. Er forscht in den Bereichen Visual Analytics, Multimedia-Retrieval und Search-as-Learning. \\

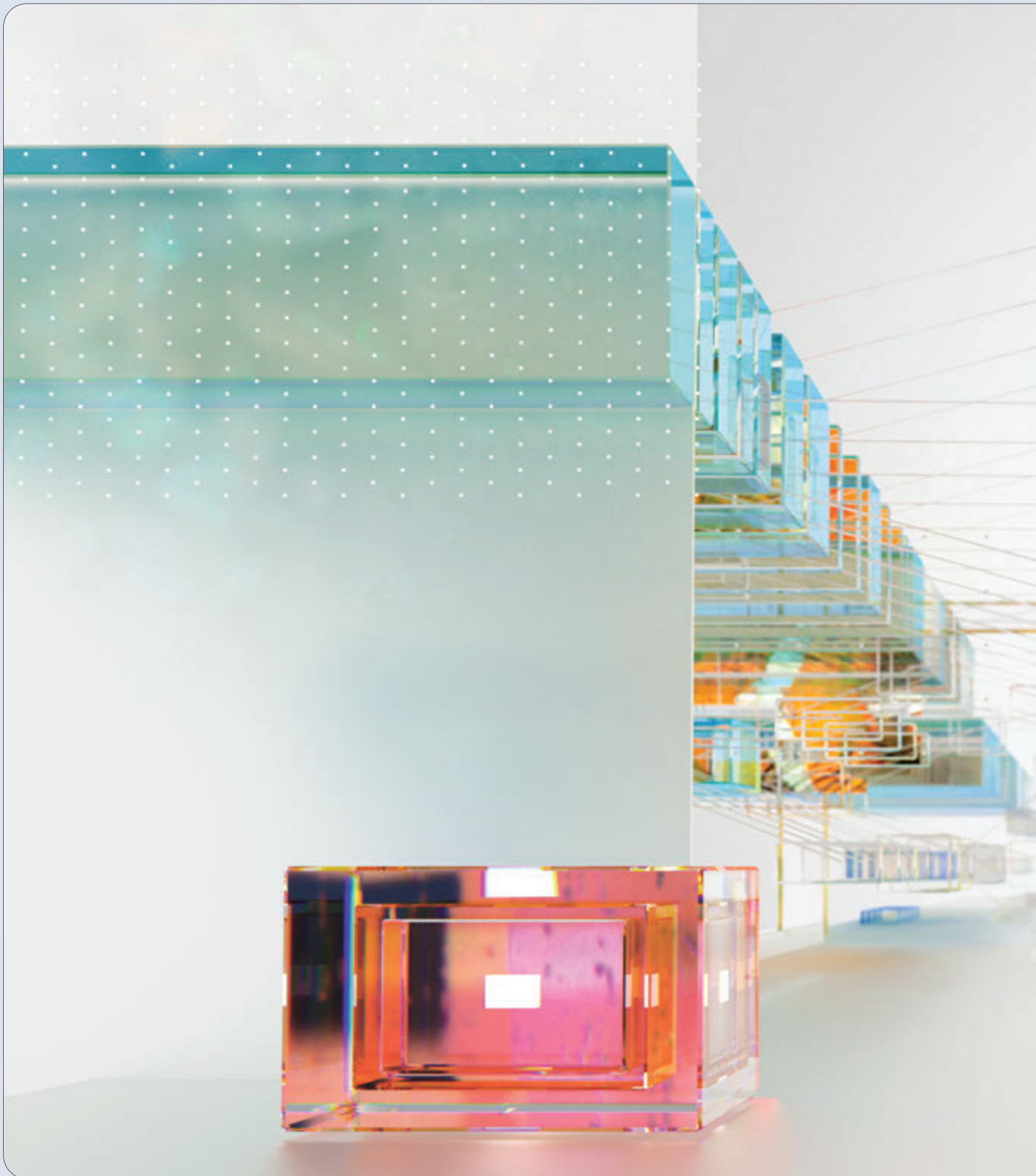


MITARBEIT:  
Dr. Anett Hoppe

\\ Dr. Anett Hoppe arbeitet seit 2016 in der Forschungsgruppe *Visual Analytics* als Postdoktorandin. Sie forscht in den Bereichen Search-as-Learning und Digitale Bibliotheken. \\







### **WIE SIEHT EIN SELBSTFAHRENDES AUTO?**

Um sicher navigieren zu können, müssen sich selbstfahrende Autos ein Bild von ihrer Umgebung konstruieren. Dafür verbindet ein Algorithmus Echtzeit-Infor-

mationen von unterschiedlichen Sensoren wie Kamera, Infrarot, Radar und Ultraschall. In Hunderten von Verarbeitungsstufen werden die Daten in einem künstlichen neuronalen Netz aufbereitet, gefiltert und mit zuvor Gelerntem abgeglichen.



Diese 3D-Illustration basiert auf dem Bilderkennungssystem *Inception* von *Google*. Zu sehen ist, wie der Informationsfluss (rechts) aufgenommen und in ein Modell der Straße verarbeitet wird (links oben). Andere Fahrzeuge werden durch die roten Kuben dargestellt. —> Bild: Field

Das Londoner Designstudio *Field* hat sich auf die künstlerische Visualisierung von Technologien wie KI spezialisiert, um ein breites Publikum zur Auseinandersetzung mit dieser Thematik anzuregen.

→ <https://www.field.io>

# Rechtliche Implikationen des maschinellen Lernens

Wenn wir über maschinelles Lernen (ML) sprechen, ist eine Tatsache evident, sogar fast trivial: Um Algorithmen zu trainieren, werden Daten benötigt. Auf Grundlage großer Datenbestände können Algorithmen erfolgreich Muster erkennen und so ihr »Wissen« erweitern. Fraglich ist, wem diese Daten »gehören«, ob Ausschließlichkeitsrechte an ihnen bestehen und ob sie als Wirtschaftsgut tauglich sind. Aber weder die rechtliche Konstruktion des Sacheigentums noch das Immaterialgüterrecht sind auf Daten

anwendbar. Mögliche Lösungen sind derzeit Teil des wissenschaftlichen Diskurses.

## WEM GEHÖREN DIE DATEN?

Daten aus autonomen Fahrzeugen sind ein Beispiel, an dem sich die unterschiedlichen Interessen aufzeigen lassen. Softwarefirmen, Autohersteller und andere Dienstleister würden gern exklusiv auf diese Daten zugreifen, denn sie verheißen neue Geschäftsmodelle. Die betroffenen Fahrer autonomer Fahrzeuge bangen hingegen um

den Schutz ihrer Privatsphäre. Sie soll insbesondere durch die ab 25. Mai geltende **Datenschutzgrundverordnung** besser geschützt werden. Hier sind Informationspflichten und Betroffenenrechte ebenso verankert wie Prinzipien der Transparenz, Zweckbindung und Datenminimierung. Im Projekt *ABIDA* wollen *L3S*-Juristen zusammen mit Partnern aus anderen Disziplinen Lösungen entwickeln, die die unterschiedlichen Interessen ausgleichen. Ein weiterer Anwendungsbereich für maschinelles Lernen ist die **öffentliche Sicherheit**. Das Projekt *iBorderCtrl* entwickelt derzeit ein »smartes« System, das die Grenzübertritte an den Schengen-Außengrenzen effektiver gestalten soll. Auch bei diesem System, das aus Software- und Hardware-Komponenten besteht – etwa Scanner für die Validierung von Ausweisdokumenten oder zur biometrischen Identifikation von Personen – kommt maschinelles Lernen zum Einsatz: Zunächst werden die relevanten Informationen digital erhoben und verarbeitet. Dies ermöglicht die systemati-



Viele Unternehmen haben Interesse an den Daten aus (selbstfahrenden) Fahrzeugen, weil sie vielfältige Geschäftsmodelle ermöglichen. → Foto: Fotolia by Adobe





sche Auswertung von Grenzübertritten und lässt Rückschlüsse etwa über besonders risikobehaftete Personengruppen oder neu auftretende Muster beim Versuch illegaler Grenzübertritte zu. Für jeden Reisenden kann auf diese Weise bereits im Vorfeld ein individueller Risikowert errechnet und entsprechend gezieltere Kontrollen durchgeführt werden. Das System lernt daraus, verbessert sich fortlaufend selbst und liefert immer genauere Ergebnisse.

### **DISKRIMINIERUNG DURCH ALGORITHMEN**

Aus rechtlicher Sicht birgt solch ein Ansatz diverse Risiken: Zunächst hängen die Ergebnisse immer von einer hinreichenden Datenqualität ab. Ferner können Algorithmen Fehlschlüsse ziehen und etwa ein Risiko größer bewerten, als es tatsächlich ist. Diskriminierungen durch Algorithmen sind also nicht auszuschließen. Selbst wenn sich insgesamt eine Steigerung der Effizienz bei Grenzkontrollen ergeben sollte, stehen sich hier das Wohl der Allgemeinheit – schnellere und sicherere Grenz-

kontrollen – und das Wohl des Einzelnen gegenüber. Gerade der Einsatz von Algorithmen in der öffentlichen Sicherheit verdeutlicht, dass bereits vorab Regeln gesetzt werden müssen, **um Diskriminierung und Intransparenz zu vermeiden**. Das gilt ebenso für den privaten Bereich. Im Rahmen von *Mobilise – Mobiler Mensch* befassen sich Wissenschaftler mit den Anforderungen an Algorithmen. Bis dato können nicht einmal die Entwickler selbstlernender Systeme nachvollziehen, wie das System genau zu einer Entscheidung gelangt ist. Kommt es zu Schäden, etwa

weil das Verhalten von Mensch und Maschine divergiert, muss das Geschehen nachvollzogen werden können. Ungeklärt ist auch die Frage, wer für Schäden einzustehen hat, die durch selbstlernende Algorithmen verursacht wurden. ML wird aus diesen und vielen weiteren Gründen nicht nur zahlreiche Rechtswissenschaftler beschäftigen, sondern auch weitreichende Veränderungen in anderen Bereichen, etwa der Versicherungsbranche, nach sich ziehen. ¶

→ <https://www.iborderctrl.eu>

→ <http://abida.de>

→ <http://L3S.de/de/projects/mobilise>

#### KONTAKT:

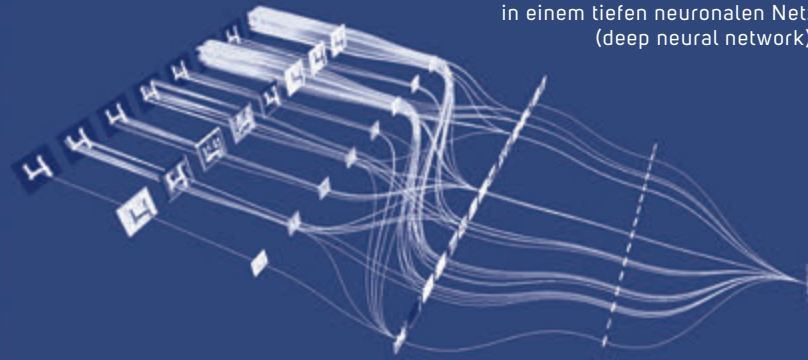
Prof. Dr. Tina Krügel, LL.M

Kruegel@L3S.de



\\ Tina Krügel ist Rechtsanwältin und seit April 2014 Juniorprofessorin für Informationsrecht, insbesondere Datenschutzrecht, an der *Leibniz Universität Hannover*. Seit 2016 ist sie Mitglied im L3S. \\

Diese Grafik zeigt die Aktivierungen in einem tiefen neuronalen Netz (deep neural network).



ALGORITHMEN VERSTEHEN

# Wie entstehen Suchmaschinen-rankings?

**Entscheidungen werden immer häufiger von Algorithmen getroffen**, zum Beispiel in der Medizin, der Automobilindustrie oder im Handel. Das zeigt, wie sehr sich die Leistungsfähigkeit komplexer Modelle des maschinellen Lernens mittlerweile verbessert hat. Auch die meisten modernen Suchmaschinen sind auf maschinelles Lernen angewiesen, damit jeder Nutzer wirklich die Informationen erhält, nach denen er sucht. Maschinelles Lernen ermöglicht es, die Suchergebnisse je nach Nutzerinteresse in eine individuelle Rangfolge zu bringen. Dabei werden verschiedene Faktoren berücksichtigt: zum Beispiel eine Vielzahl von Signalen, die auf Informationen in den Dokumenten basieren, die Relevanz der Dokumente und das Nutzerverhalten. Das ist sehr komplex und **nicht mehr nachvollziehbar**, Modelle der Entscheidungsfindung sind dann eine Art Black Box. Der Nutzer versteht auch nicht ansatzweise, wie das Ergebnis zustande kommt. Wenn also ein Algorithmus Informationen gewichtet, um Vorhersagen, Klassifizierungen und Ranglisten zu erstellen, dann kommt der Transparenz eine besondere Bedeu-

tung zu, um mögliche Diskriminierungen im Auge zu haben und das Vertrauen in das System durch Interpretierbarkeit zu stärken. Interpretierbarkeit heißt hier, das Ergebnis für jeden verständlich erklären zu können.

**Aber wie kann Transparenz und Erklärbarkeit hergestellt werden?** In letzter Zeit wird immer intensiver daran geforscht, wie ein maschinelles Lernsystem die Erklärungen gleich mitliefern kann, zum Beispiel bei der Klassifizierung von Bildern und Texten, bei maschinellen Übersetzungen sowie bei Empfehlungssystemen. Bislang fehlen Untersuchungen im Bereich der Web-Suche, also bei der Suche nach komplexen Inhalten. Die für den maschinellen Lernprozess erforderlichen großen Datenmengen und komplexen Modelle erschweren die Erklärbarkeit der Ergebnisse. Am L3S arbeitet Prof. Dr. Avishek Anand mit seinem Team daran, die Fundamente für die Interpretierbarkeit der Web-Suche zu legen, und hat dafür einen *Amazon Research Award* erhalten. ¶

→ <https://L3S.de/~singh/interpret>

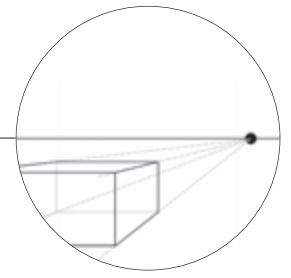


KONTAKT:

Prof. Dr. Avishek Anand

Anand@L3S.de

\\ Avishek Anand ist Juniorprofessor an der *Leibniz Universität Hannover* und Mitglied des *Forschungszentrums L3S*. Er befasst sich mit Fragen von Information-Retrieval und Web-Suche: Wie können wir Entscheidungen erklären, die durch Suchmaschinen getroffen wurden? Wie können wir zeitbezogene Informationen einbinden, um das Suchergebnis zu verbessern? \\



# Nicht die Orientierung verlieren



Fluchtpunkte können auch für Augmented Reality verwendet werden. → Foto: David Perez/Flickr + eigene Bearbeitung

## In einigen Bereichen ist das menschliche Gehirn der Maschine immer noch überlegen:

Für Menschen ist es ziemlich einfach, die grobe Struktur einer Szene auf einen Blick zu erfassen. Das Gehirn identifiziert dafür Fluchtpunkte im Bild der Umwelt. Sie geben Aufschluss über die Orientierung und Anordnung von Objekten und entstehen, wenn im dreidimensionalen Raum parallele Linien auf eine zweidimensionale Ebene projiziert werden – zum Beispiel auf die Netzhaut des Auges oder den Sensor einer Digitalkamera. Die vormals parallelen Linien werden so transformiert, dass sie sich anschließend in einem gemeinsamen Fluchtpunkt schneiden.

Fluchtpunkte können zur Extraktion von Informationen aus Bildern genutzt werden, die **für komplexe Anwendungen wie der Steuerung autonomer Fahrzeuge** nötig sind. Nicht nur die Ausrichtungen und relativen Abstände von Objekten wie Autos und Gebäuden, sondern auch die Orientierung des Fahrzeugs zu seiner Umwelt können mit Hilfe von Fluchtpunkten gemessen werden, die aus einer einzigen Kamera gewonnen werden.

Trotz mehr als 20 Jahren Forschung gibt es bislang **keine ausreichend robusten Algorithmen** zur automatischen Fluchtpunkt-detektion. Das menschliche Gehirn bezieht viel Kontextinformation und Vorwissen ein, die sich nur mit sehr hohem Aufwand in einen klassischen Bildverarbeitungsalgorithmus gießen ließen. L3S-Wissenschaftler am *Institut für Informationsverarbeitung* erforschen daher, wie künstliche neuronale Netze dieses Wissen erlernen und zum Erkennen von Fluchtpunkten nutzen können. Bekannte Netzarchitekturen und Datensätze zum Anlernen der neuronalen Netze haben sich bislang als untauglich für diese Problemstellung erwiesen. Die Forschungsschwerpunkte am L3S liegen daher auf der Erkundung neuartiger Netzstrukturen und der effektiven Erzeugung und Anreicherung von Daten, mit denen sie trainiert werden können. ¶

→ <https://www.L3S.de/de/projects/covmap>

## KONTAKT:

Florian Kluger, M.Sc.

kluger@L3S.de



\\\ Florian Kluger ist wissenschaftlicher Mitarbeiter am *Institut für Informationsverarbeitung*. Er beschäftigt sich mit Machine-Learning und dessen Anwendungen im Bereich automatischer Bildinterpretation. \\\





BIG DATA UND DAS VERLORENE WEB

## Blick in die Vergangenheit

Das Internet vergisst nicht. Trotzdem sind viele Inhalte auf Anhieb unauffindbar – alte Webseiten zum Beispiel. Dabei gibt es Organisationen, die alte Seiten für die Nachwelt aufbewahren und zur Verfügung stellen, allen voran das *Internet Archive* in Kalifornien. Am *L3S* beschäftigen sich Forscher bereits seit 2014 intensiv mit Web-Archiven und erarbeiten neue **Anwendungsfälle**, **Zugriffsmethoden** und **Analyseverfahren** für diese interessanten Datensammlungen. Das *L3S* hat Zugang zu einer lokalen Kopie des gesamten deutschen Webs unter der Domäne *.de*, welches das *Internet Archive* bereits seit 1996 archiviert. Eine Fragestellung, die die Forscher dabei beschäftigt: Wie kann der Zugang zu diesen archivierten Seiten für jeden von uns einfacher werden? Besonders durch den zeitlichen Aspekt, der im aktuellen Web kaum eine Rolle spielt, in einem Archiv jedoch höchste Priorität hat, ergeben sich neue Anforderungen an Suchmaschinen. Es ist nicht mehr nur wichtig, eine möglichst relevante Seite zu finden, sondern außerdem eine bestimmte Version einer Seite, die sich möglicherweise in der Zwischenzeit verändert hat oder gar nicht mehr verfügbar ist. Zwei Such-



maschinen, die vor diesem Hintergrund als Prototypen entstanden sind, jedoch noch weiterentwickelt und verbessert werden, können unter <http://alexandria-project.eu/archivesearch> (*Archive-Search*) sowie <http://tempas.L3S.de> (*Tempas – Temporal Archive Search*) bereits ausprobiert werden. Das Interesse an Web-Archiven nimmt auch in anderen Wissenschaftsdisziplinen zu.

**Für Historiker, Politikwissenschaftler und andere, die früher weitestgehend mit analogen Daten gearbeitet haben, gewinnt das Web immer mehr an Bedeutung – und damit auch die archivierten Websites.**

Bei den riesigen Datenmengen ist es allerdings nicht mehr möglich, alle Dokumente zu lesen. Also sind neue **Datenverarbeitungsmethoden** notwendig. Auch daran forscht das *L3S* und bietet mit *ArchiveSpark* eines der meistgenutzten Tools für den effizienten Zugriff auf »historische« Web-Kollektionen.



Die Websites der CEBIT aus den Jahren 2000, 2002, 2009 und 2011. → Screenshots: web.archive.org

pa auf diese Weise zur Verfügung. Die wachsende Rolle von Daten beziehungsweise Big Data ist nicht nur in der Forschung, sondern insbesondere auch in der **Wirtschaft** spürbar. Das Forschungszentrum *L3S* bietet mit seinen Arbeiten aus den Projekten *Alexandria* und *SoBigData* die optimale Grundlage für einen leichten Einstieg in dieses komplexe Themenfeld. Da sich das Web zum primären Medium entwickelt, um Nachrichten, Informationen und Daten zu teilen, wird auch die Bedeutung von Web-Archiven als Zeugen dieser Entwicklung weiter zunehmen und auch in Bereiche Einzug halten, wo dies heute noch gar nicht absehbar ist. Diesen Weg zu ebnen, ist das Ziel von *Alexandria*. Die *SoBigData*-Infrastruktur als zentraler Anlaufpunkt für alle Fragen rund um Big Data ermöglicht, die Forschungsergebnisse anderen zugänglich zu machen und gemeinsam in Europa Ansätze für die Arbeit mit diesen und anderen Daten zu erarbeiten. ¶  
 → <http://alexandria-project.eu>  
 → <http://www.sobigdata.eu>

Die Software entwickeln die Wissenschaftler am *L3S* gemeinsam mit dem *Internet Archive*, um auf möglichst einfache Weise Datenanalysen jeglicher Art in Web-Archiven zu ermöglichen. Ein weiteres Projekt, das sich ebenfalls mit großen Datenmengen beschäftigt, ist *SoBig-Data*. Gemeinsam mit anderen Wissenschaftlern aus Italien, Großbritannien, den Niederlanden, Estland, Finnland und der Schweiz entwickelt das *L3S* eine **europäische Forschungsinfrastruktur** für Big Data.

Dabei werden Datensätze aus unterschiedlichen Quellen sowie verschiedene Tools für die Arbeit mit diesen Daten in einer offenen Plattform integriert. Zusätzlich werden Leitfäden sowohl für die praktische Arbeit mit den Daten als auch mit besonderem Fokus auf die rechtlichen und Datenschutz-relevanten Aspekte sowie umfassende Beispiele und Vorlagen über die *SoBigData*-Plattform geteilt. Auch die oben genannten Arbeiten zu Web-Archiven stellt das *L3S* anderen Wissenschaftlern aus ganz Euro-



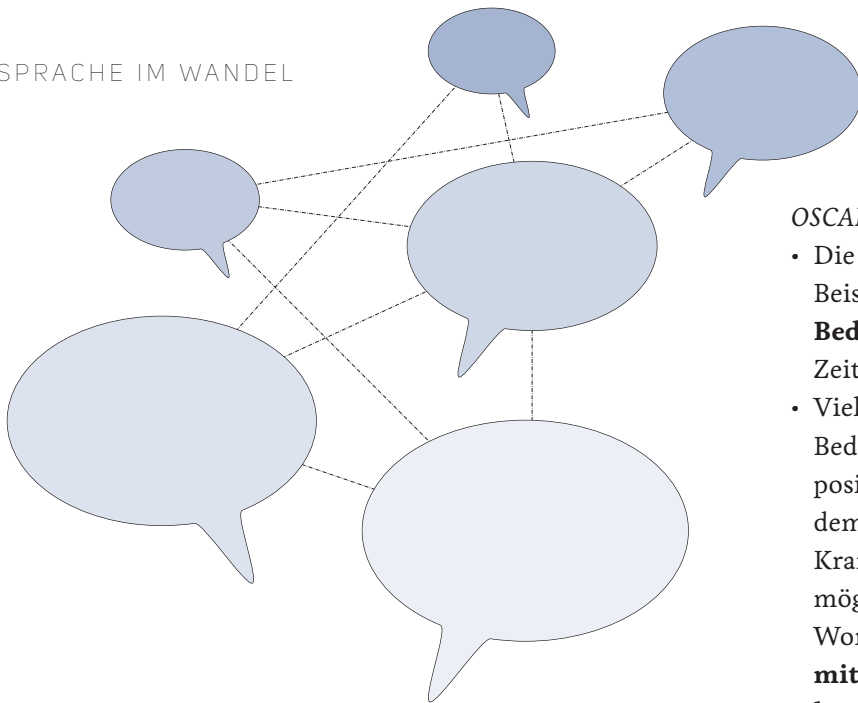
KONTAKT:  
 Prof. Dr. techn. Wolfgang Nejdil  
 Nejdil@L3S.de

\\ Wolfgang Nejdil leitet das *Forschungszentrum L3S* und forscht u. a. in den Bereichen Suche und Information-Retrieval im Rahmen seines ERC-Advanced-Grant *Alexandria*. \\



KONTAKT:  
 Prof. Dr. Avishek Anand  
 Anand@L3S.de

\\ Avishek Anand leitet am *L3S* das Projekt *SoBig-Data*. Seine Forschungsinteressen liegen im Bereich Information-Retrieval und maschinelles Lernen. \\



MASCHINELLES LERNEN  
UND SOZIALE MEDIEN

# Im Netz der Meinungen

**Was denken andere Leute?** Diese Frage hat uns schon immer interessiert und unsere Entscheidungen beeinflusst. Seit dem Aufkommen des Web 2.0 nutzen viele Menschen soziale Medien, um ihre Meinung über ihren Bekanntenkreis hinaus zu allen möglichen Themen zu posten. Mit den Meinungen und Stimmungen im Web befasst sich der Bereich des **Opinion-Mining** und der **Sentimentanalyse**. Die Erkenntnisse sind wichtig für Produktentwicklung, Marketing, Eventplanung, politische Kampagnen. Die klassische Sentimentanalyse nutzt dafür statische Daten. Allerdings können sich Meinungen und Themen in sozialen Medien über die Zeit verändern. Im Projekt OSCAR entwickelt Prof. Dr. Eirini Ntoutsis mit ihrem Team daher sogenannte Opinion-Stream-Mining-Methoden, die mit Änderungen umgehen können und das Modell der Sentimentanalyse kontinuierlich an die dem Modell zugrundeliegende sich verändernde Grundgesamtheit anpassen.

OSCAR bewältigt dabei drei Herausforderungen:

- Die Daten und das Vokabular ändern sich: Zum Beispiel verdrängen neue Wörter alte oder die **Bedeutung eines Begriffs verändert sich im Zeitablauf**.
- Viele Wörter haben je nach Kontext eine andere Bedeutung. Das Wort »ernst« kann zum Beispiel positiv oder negativ verstanden werden, je nachdem, ob es sich auf eine Beziehung oder eine Krankheit bezieht. Da es nicht möglich ist, alle möglichen Kontexte zu erfassen, in denen ein Wort auftauchen könnte, **müssen Algorithmen mit der Mehrdeutigkeit von Wörtern umgehen können**.
- Auf vielen Plattformen äußern Nutzer ihre Meinungen ohne explizit die Einstellung zu spezifizieren, die sie mit ihrem Text verbinden. Um Modelle des maschinellen Lernens trainieren zu können, sind aber **gekennzeichnete Daten** notwendig.

Das Ergebnis von OSCAR ist ein komplettes Framework aus Ensemble-Learning-Methoden, die mit verschiedenen Formen von Veränderungen umgehen können. Ein solches Framework kann auch für andere Aufgaben genutzt werden, zum Beispiel für vorausschauende Instandhaltung oder Netzwerküberwachung. ¶

→ <https://www.L3S.de/en/projects/oscar>

## KONTAKT:

Prof. Dr. Eirini Ntoutsis

Ntoutsis@L3S.de



\\ Eirini Ntoutsis ist Professorin am *Institut für Verteilte Systeme der Leibniz Universität Hannover* und seit 2016 Mitglied im L3S. Ihre Forschungsinteressen liegen in den Bereichen Data Mining und Maschinelles Lernen. \\





15 PB entsprechen dem Text auf einem einseitig beschriebenen Papierstapel von 350.000 km Höhe (100 Seiten sind 1 cm hoch), der fast von der Erde bis zum Mond reichen würde (mittlerer Abstand Erde – Mond = ca. 384.400 km).

### ZAHLENSALAT UND EINHEITENBREI

| DEZIMAL               | EINHEIT       | EINHEIT        | BINÄR                 |
|-----------------------|---------------|----------------|-----------------------|
| 1                     | Petabyte (PB) | Pebibyte (PiB) | 1                     |
| 1.000                 | Terabyte (TB) | Tebibyte (TiB) | 1.024                 |
| 1.000.000             | Gigabyte (GB) | Gibibyte (GiB) | 1.048.576             |
| 1.000.000.000         | Megabyte (MB) | Mebibyte (MiB) | 1.073.741.824         |
| 1.000.000.000.000     | Kilobyte (KB) | Kibibyte (KiB) | 1.099.511.627.776     |
| 1.000.000.000.000.000 | Byte (B)      | Byte (B)       | 1.125.899.906.842.624 |
| 8.000.000.000.000.000 | Bit (b)       | Bit (b)        | 9.007.199.254.740.992 |

Wer sich schon mal gewundert hat, dass Datenspeicher mit unterschiedlichen Speichergößen bezeichnet werden: Es gibt ein *dezimales Kilobit* mit 1.000 Bit und ein *binäres Kilobit* mit 1.024 Bit \* – das liegt an den verschiedenen Zahlensystemen von Mensch und Computer. Während Menschen meistens im Dezimalsystem (0 bis 9) auf Basis von 10

rechnen, kennen Computer nur die Zustände »1« und »0« und rechnen im Dualsystem auf Basis von 2. Die Bedeutung der Präfixe Kilo, Mega und Giga sind im internationalen Einheitensystem *SI (Système international d'unités)* verbindlich festgelegt – auf Basis des dezimalen Zahlensystems. Um eine eindeutige Unterscheidung zu treffen, gibt es

seit 1996 zusätzliche Präfixe, die auf dem dualen Zahlensystem beruhen und extra für binäre Werte festgelegt wurden.

\* Die SI-Präfixe sind für ein Vielfaches von 10 hoch 3 (= 1.000) definiert, die Binär-Präfixe stehen für ein Vielfaches von 2 hoch 10 (= 1.024).

## ANKÜNDIGUNG



19. und 20. 6. 2018

### SysInt – Internationale Konferenz zur Digitalisierung

Wissenschaft und Industrie diskutieren neue Trends: Am Produktionstechnischen Zentrum Hannover (PZH) der Leibniz Universität in Garbsen findet am 19. und 20. Juni die 4th International Conference on System-Integrated Intelligence statt. Die Konferenz bietet Wissenschaft und Industrie ein gemeinsames Forum rund um den Einsatz intelligenter, flexibler und vernetzter Systeme in Produkten und in der Produktion. Zu den

Keynote-Rednern gehört der Direktor des L3S Prof. Dr. Wolfgang Nejdl. Unter dem Titel »Are you looking for an AI that works?« spricht Nejdl über Methoden der künstlichen Intelligenz und des maschinellen Lernens für die intelligente Produktion, die Fortschritte in den letzten fünf bis zehn Jahren sowie über Herausforderungen und (falsche) Strategien, die zu Problemen beim Einsatz moderner KI-Methoden und maschinellen Lernens führen können. Erstmals gibt es bei dieser Konferenz eine Industrie-Session mit Vorträgen aus der Praxis zu den Themen Industrie 4.0 und Digitalisierung: Vertreter aus der Industrie berichten über aktuelle

Herausforderungen und umgesetzte Lösungen. Einige Schwerpunktthemen der Veranstaltung sind: Die Integration von intelligenten Funktionen in Systeme, Bauteile, Produkte und Produktionssysteme, neue Sensortechnologien und Sensorwerkstoffe, intelligente Produkte samt cyber-physischer Systeme (CPS) sowie intelligente Produktions- und Logistikprozesse. Während der Konferenz können die Teilnehmer Digitalisierung in Live-Demonstrationen im Versuchsfeld am PZH erleben. ¶

ORT:

Leibniz Universität Hannover

An der Universität 2

30823 Garbsen

→ <http://www.sysint-conference.org>

## AUSZEICHNUNG

19. 4. 2018

### GFFT-Förderpreis »Beste Dissertation« für Besnik Fetahu

L3S-Wissenschaftler Dr. Besnik Fetahu wurde auf der Jahrestagung der *gemeinnützigen Gesellschaft zur Förderung des Forschungstransfers e. V. (GFFT)* mit dem GFFT-Förderpreis für die beste Dissertation ausgezeichnet. Die GFFT-Förderpreise werden jährlich für hervorragende Innovationen verliehen.

Zur Begründung heißt es, Besnik Fetahu habe in seiner Doktorarbeit »Approaches for Enriching and Improving Textual Knowledge Bases« einen innovativen Ansatz entwickelt, um Wikipedia-Artikel automatisch mit relevanten, fehlenden und wichtigen Zitierungen aus Informationsmedien zu ergänzen. Den Autoren werden dafür aktuelle Zitierungen vorgeschlagen. Damit wird die Nachweisbarkeit unterstützt, eines der zentralen Prinzipien von Wikipedia. Vor dem Hintergrund, dass Wikipedia täglich von Millionen Menschen und

zahlreichen Web-Applikationen genutzt wird und der Zugang zu Wissen und Information ein zentraler Grundpfeiler vieler Unternehmen ist, können Besnik Fetahus Forschungsergebnisse eine Vielzahl von Geschäftsszenarien und Anwendungsfällen maßgeblich beeinflussen. ¶



Dr. Besnik Fetahu (Bildmitte) mit dem Vorsitzenden der Jury Dr. Thomas Kunstmann, Geschäftsbereichsleiter *Travel & Logistics, msg systems ag* (rechts) und Dr. Gerd Große, Vorstandsvorsitzender des *GFFT e.V.* und Geschäftsführer der *GFFT Technologies GmbH* (links).  
—> Foto: Anne

## IMPRESSUM

---



HERAUSGEBER:

**Forschungszentrum L3S**  
**Leibniz Universität Hannover**  
Appelstraße 9a  
30167 Hannover

VERANTWORTLICH:

Prof. Dr. techn. Wolfgang Nejdl •  
Geschäftsführender Direktor

REDAKTION:

Dipl.-Geogr. Susanne Oetzmann  
Telefon: + 49.(0)511.762-17734  
Fax: + 49.(0)511.762-17779  
E-Mail: oetzmann@L3S.de

KONZEPT & DESIGN:

Dipl.-Des. Priska Tosch  
[www.tosch-kommunikation.de](http://www.tosch-kommunikation.de)

DRUCK:

auf 100% Recyclingpapier  
Ströher Druckerei und Verlag  
GmbH & Co. KG  
[www.stroeher-druck.de](http://www.stroeher-druck.de)



BILDQUELLEN:

Forschungszentrum L3S,  
wenn nicht anders vermerkt.  
Titelbild-Gestaltung:  
Priska Tosch

---

[www.L3S.de](http://www.L3S.de)

---







**Applied  
Machine Learning  
Academy**

# **Maschinelles Lernen für die Industrie**

Lehrangebot an der  
Leibniz Universität Hannover

Die Digitalisierung  
erfordert Weiterlernen.  
Dabei gibt es Spannendes zu entdecken!  
<https://ama-academy.eu>

