

Christian Kohl – Beratung & Projektmanagement

Beispiele aus der Verlagsbranche, in denen Verfahren, die zumindest teilweise aus dem Bereich „KI“ stammen, zum Einsatz kommen

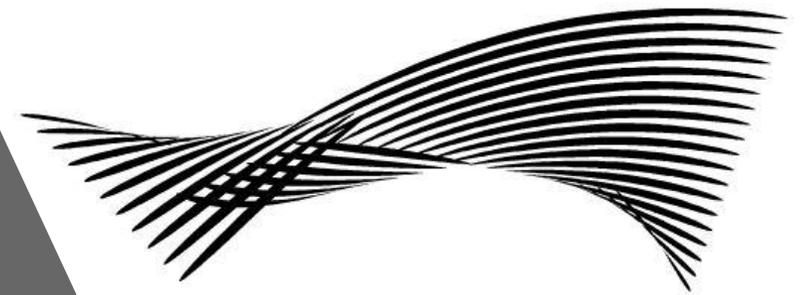
Markupforum

Stuttgart, 16.11.2018

Christian Kohl - Beratung und Projektmanagement

christian@kohl.consulting

www.kohl.consulting





Konkrete und erfolgreiche Anwendungen aus dem „inhaltlichen Bereich“ – es geht nicht um Marketing- oder Prozessautomatisierung

Appetitanreger – was tut sich anderswo?

Anregung für soziologische / psychologische Forschungen: Die „Sexyness“ eines Themas verschwindet, sobald man mehr darüber erfährt ;-)

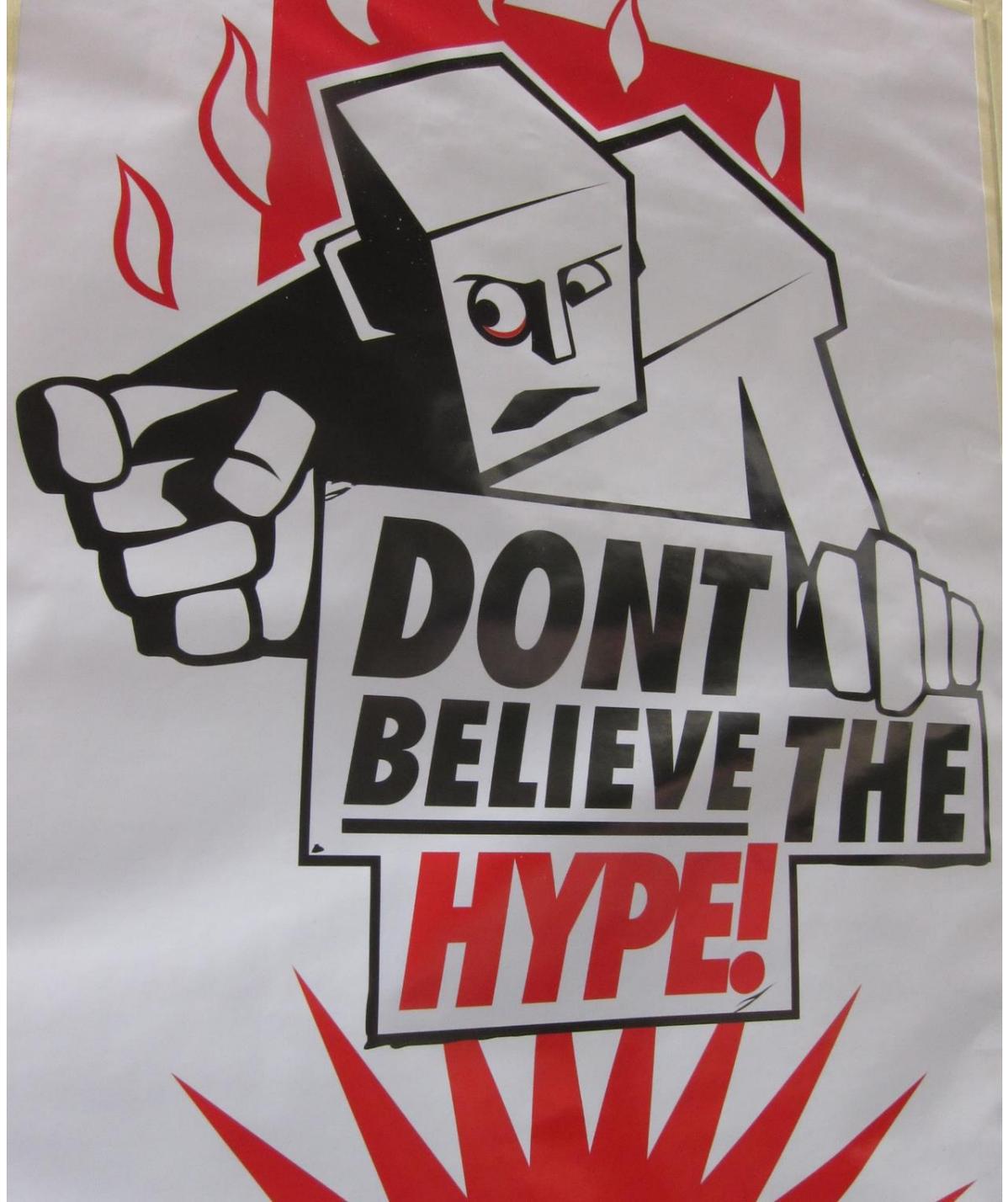
Motivation



“Any sufficiently advanced technology is indistinguishable from magic”

(Arthur C. Clarke)

Even an application of basic statistics will be celebrated as “intelligent” if it “feels” that way.





“A lot of cutting edge AI has filtered into general applications, often without being called AI because once something becomes useful enough and common enough it's not labelled AI anymore.”

(Nick Bostrom, Source: <http://edition.cnn.com/2006/TECH/science/07/24/ai.bostrom/>, last accessed: 25.05.2017)

“AI is whatever hasn't been done yet.”
 (“Tesler's Theorem”)

“The minute we can automate a task, we downgrade the relevant skill involved to one of mere mechanism.”

(Gideon Lewis-Kraus: The Great A.I. Awakening. In: The New York Times Magazine. December 2016. Last accessed: 15.05.2017)



Die größten Förderer sind Militär und Regierungen

IT **NEVER** BLINKS, SLEEPS OR EATS

INTRODUCING THE GLOBAL DIGITAL DEFENSE NETWORK: **SKYNET**

The United States Air Force, in partnership with Cyberdyne Systems, is pleased to announce the launch of the Global Digital Defence Network, or SKYNET. In a world of global threats, terrorist mobility and third world nuclear arsenals, Skynet ensures an ever watchful and instantly reacting digital response that eliminates human error.

You are cordially invited to join us on April 19th to celebrate the official "live" launch of Skynet. This will be a date in history that you won't want to miss. The world will never be the same.

RSVP your US Air Force liaison or Cyberdyne Sales Representative today.

Image Source: Windell Oskay, <https://www.flickr.com/photos/oskay/1328610525/>, License: CC BY 2.0, <https://creativecommons.org/licenses/by/2.0/>.



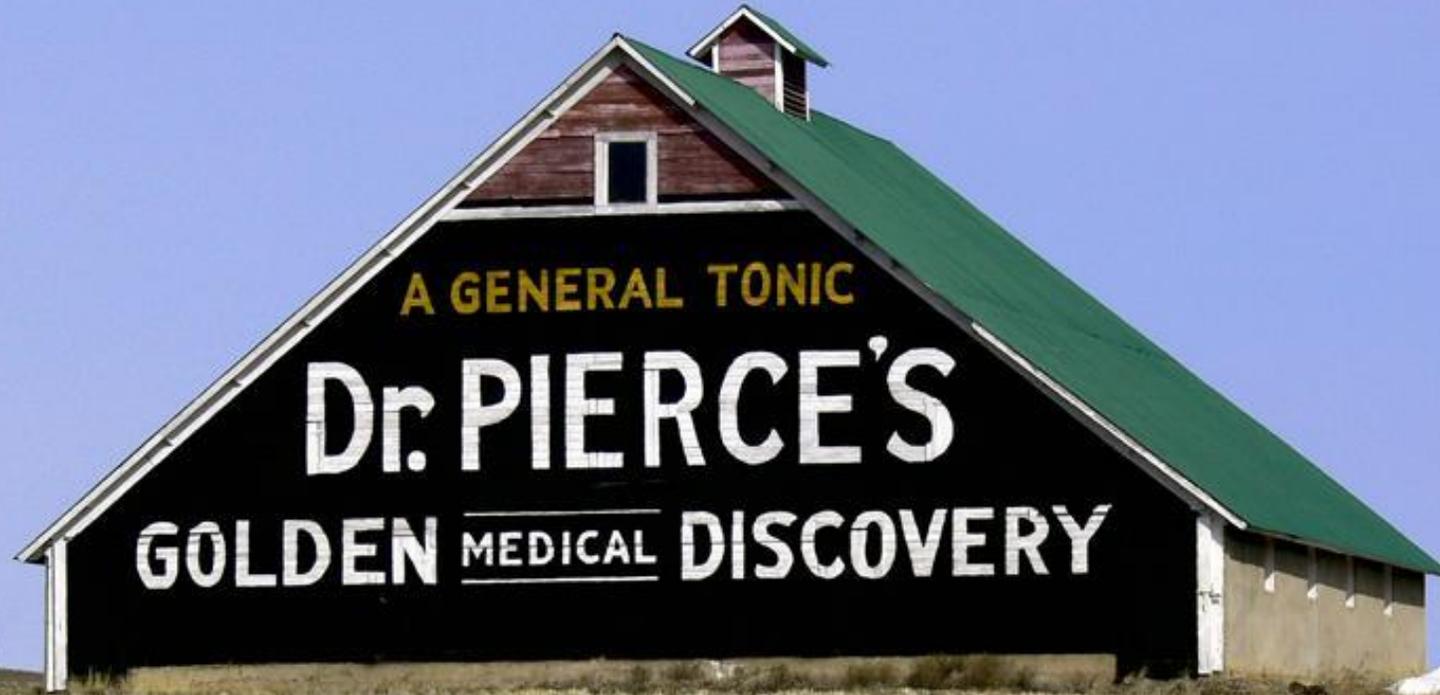
U.S. AIR FORCE



All images used have been paid for and are rights free. BadGiant and Giant Ideas www.giantideas.com The world's coolest agency!



Beware the Snake Oil Salesman!





Hinweis / Disclaimer

- Dies sind von mir geteilte Beobachtungen
- Ich habe *nicht* an diesen Projekten mitgearbeitet und repräsentiere die erwähnten Firmen nicht



Rebecca Lieb ✓

@lieblink

Folgen



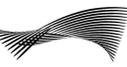
ex-Googler: "Managing content is hard. We tried it at Google & decided to focus on easier stuff, like self-driving cars."

 Tweet übersetzen

19:18 - 9. Okt. 2014

653 Retweets **400** „Gefällt mir“-Angaben





Unstrukturierter Content (Text)

- Nur von Menschen lesbar
- Niedriger Automatisierungsgrad

Strukturierter Content (logische Struktur)

- Von Menschen lesbar
- Von der Maschine verarbeitbar (Organisation, Display, ...), aber nicht „verstehbar“
- Mittlerer Automatisierungsgrad

„Smarter“ Content (semantische Informationen)

- Von Menschen lesbar
- Von der Maschine lesbar
- Kontextualisierung
- Hoher Automatisierungsgrad

Leistung	Zahl der Angebote
Chatbots	11
Text Mining	23
Data Mining	13
Textgenerierung (NLG)	12
Verstehen natürlicher Sprache (NLU)	15
Bildanalyse	11
Bildbearbeitung/Bilderstellung (GAN)	1
Videoanalyse	10
Videogenerierung	1

Angebot im Überblick

Kuratierung

- Nachfrage erkennen
- Themen erkennen
- Personen erkennen
- Daten korrelieren
- Den richtigen Kanal, den richtigen Verlag, das richtige Journal finden

(Personalisierte) Distribution

- Kollektionen / Produkte erstellen
- Produkte erstellen
- „Adaptives Lernen“
- Daten statt Formate verkaufen
- Informationen (Wissen?) statt Dokumente verkaufen
- Neue Kanäle & Formate (bspw. Quartz)

Content Erzeugung & QS

- Texte, Audios, Videos (semi-)automatisch generieren
- Daten sammeln und aggregieren
- Copy editing
- Statistikprüfung
- Bildmanipulationen erkennen
- Reviewer finden

Discovery

- Suchen (direkt und explorativ), v.a. “fachübergreifend”
- Sowohl für Nutzer als auch für Verlage intern (“Wie kann ich meine Backlist, meinen Contentpool besser monetarisieren?”)
- Empfehlungen
- Relationen / Zusammenhänge erkennen
- Produktbeschreibungen, Summaries, Profile generieren
- Verschlagwortung etc.

Beispiele: Textgenerierung

Neun europäische Presseagenturen
arbeiten bereits mit automatisch erstellten
Texten



Beispiel: Textgenerierung

- YSEOP, Arria, Narrative Science, Automated Insights, 2txt NLG, AX Semantics, boot.ai, Codiac, Liquid Newsroom, Retresco, textOmatic ...
- Generierung von Artikeln, Reports, Präsentationen basierend auf Datenbanken
- Lokale Sportberichte
- Finanz-/Börsenberichterstattung
- Wetterberichterstattung
- Produktinformationstexte



Was bringt es?

- Steigerung des Outputs
- Tiefere und breitere Berichterstattung
- Kostensenkung
- Größere Aktualität
- Beispiel Associated Press:
 - Ca 15-fache Steigerung des Outputs: 4400 statt 300 quarterly earnings reports
 - Ca 3 FTEs für andere Aufgaben „befreit“
- Beispiel Produktinformationen:
 - Personalisierung in Echtzeit
 - Discovery
- Beispiel Sportberichterstattung (Westline, Zwölfter, Yahoo Fantasy Football, Berliner Morgenpost ...):
 - Abdeckung unterer / ausländischer Ligen
 - Fantasy-Sports Artikel in Echtzeit



Beispiele: Quarterly Earnings Reports

- Automated Insights



Beispiel: Sport

- Spielberichtsgenerator von Retresco



Beispiel: Wetter

- Beispiel Stuttgarter Zeitung / Feinstaubalarm
- Technologie: AX Semantics

Beispiele:
Literatur-
wissenschaft



Just kidding ;-)

- <http://www.elsewhere.org/pomo/>
- Žižuku: <http://blog.talkingphilosophy.com/?p=219>

Beispiele: Verstehen
natürlicher Sprache (NLU)

Beispiele: Verstehen natürlicher Sprache (NLU)

- Bsp.: Notwendige Voraussetzung für Chatbots/automatische Dialogsysteme
- Bsp.: Analyse großer Textmengen, Aufbereitung zur besseren Navigation und Exploration durch konzeptionelle Erschließung von Wissen

Beispiele: Google Perspective

- Misst „Toxizität“ von Formulierungen
- <https://www.perspectiveapi.com/>
- Partner: The Guardian, The Economist, The New York Times, Wikipedia

Beispiele: Bildanalyse,
Bildbearbeitung,
Bildgenerierung

Beispiele: Bildanalyse

- Automatische Generierung von Metadaten
 - Lizenzierung
 - Barrierefreiheit (Bildbeschreibung)
 - Jugendschutz
 - Bildunterschriften
 - ...
- Seeing.ai (App): beschreibt die Welt (Personen, Texte, Objekte):
<https://www.microsoft.com/en-us/seeing-ai/>
- Captionbot: Generiert Bildbeschreibung:
<https://www.captionbot.ai/>

I think it's a bridge over a body of water.



I think it's a group of people on a stage in front of a crowd.



Beispiel: Bildgenerierung

- <https://blogs.microsoft.com/ai/drawing-ai/>
- “For now, the technology is imperfect. Close examination of images almost always reveals flaws, such as birds with blue beaks instead of black and fruit stands with mutant bananas. These flaws are a clear indication that a computer, not a human, created the images. Nevertheless, the quality of the AttnGAN images are a nearly three-fold improvement over the previous best-in-class GAN.”

Beispiel: Bildgenerierung

„Project Magenta“ (Google) zeichnet Doodles

- https://magenta.tensorflow.org/assets/sketch_rnn_demo/index.html

“Chat Painter” (Microsoft) zeichnet Bilder basierend auf einem Dialog

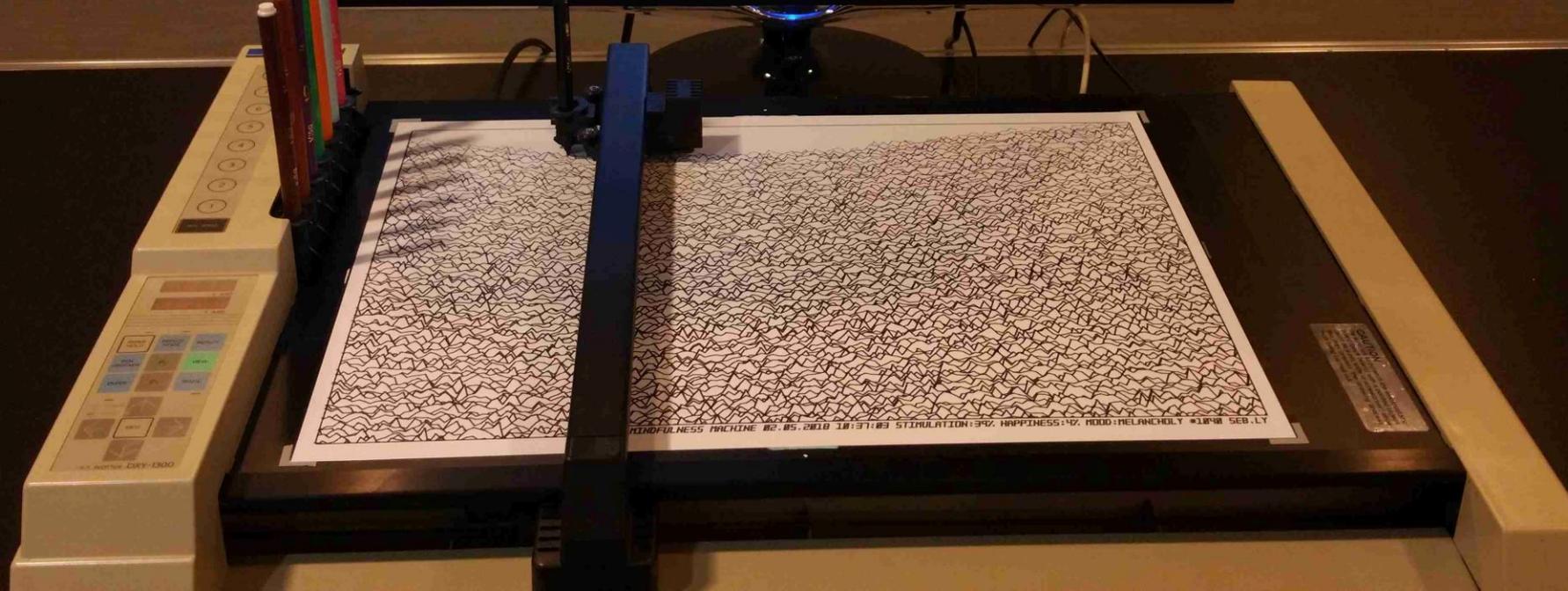
- <https://www.microsoft.com/en-us/research/blog/chatpainter-improving-text-image-generation-using-dialogue/>
- DA-GAN: Instance-level Image Translation by Deep Attention Generative Adversarial Networks (with Supplementary Materials): <https://scirate.com/arxiv/1802.06454>

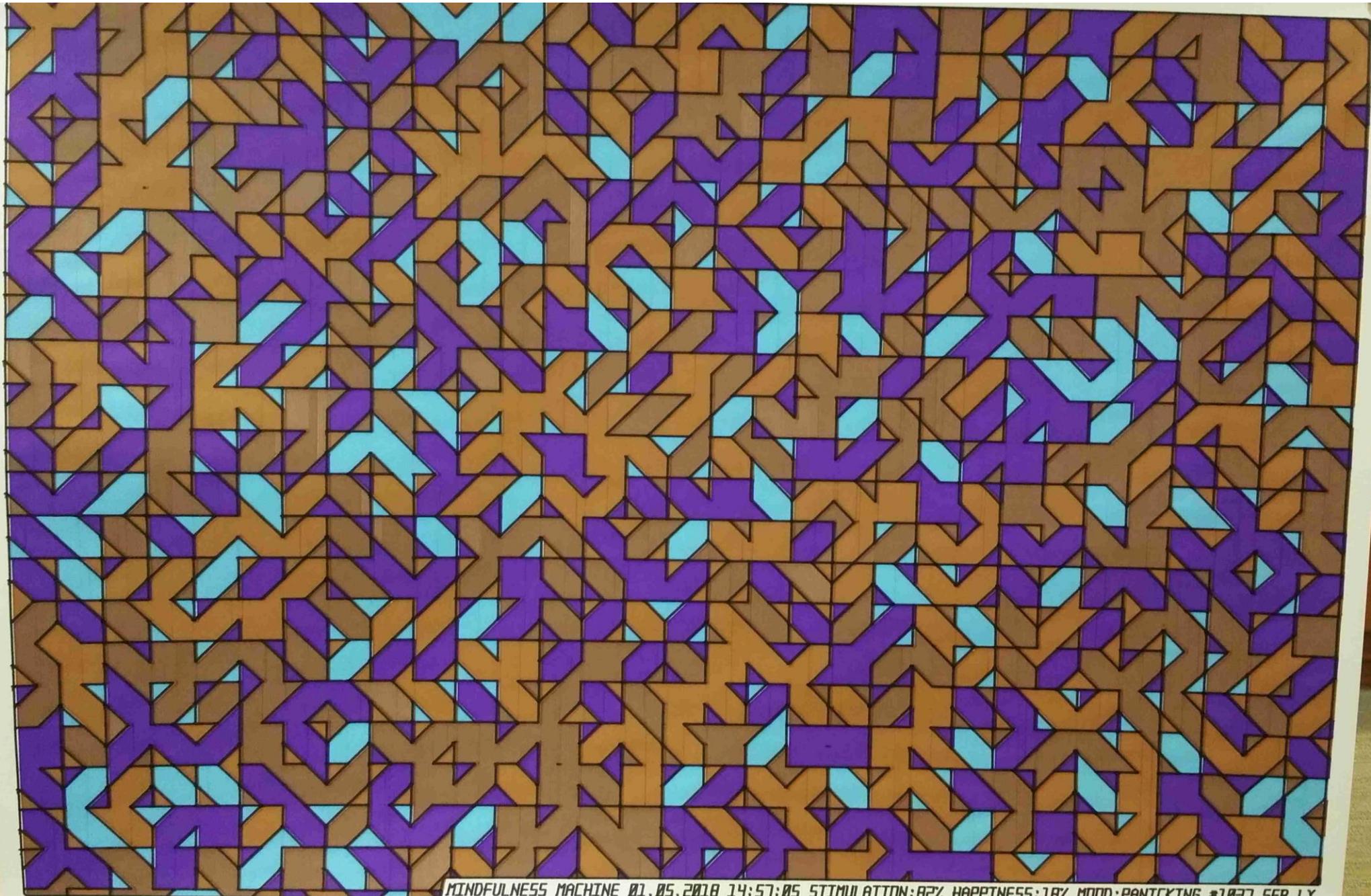


Beispiel: Kunst

„The Mindfulness Machine“

- Misst Stimmung der Umgebung und zeichnet basierend auf diesen Werten
- <https://futurism.com/images/meet-worlds-talented-robot-artists/>





MINDFULNESS MACHINE 01.05.2018 14:57:05 STIMULATION:82% HAPPINESS:18% MOOD: PANICKING #1037 5EB.LY

Beispiele: Videoanalyse & Videogenerierung

Beispiel: Videoanalyse

Durchsuchen, analysieren und taggen von Videos -
Beispiel Raypack AI:

- Face Recognition (automatische Erkennung von Promis etc.)
- Scene Recognition (wo befinden wir uns: Gebirge, Strand, Büro etc.)
- Object Detection (welche Objekte sind zu sehen?)
- Text to Speech Conversion (was wird gerade gesprochen?)
- In Echtzeit, pro Frame, returniert bspw. JSON Dateien, die dann für Suche verwendet werden können ...



Beispiel: Video- generierung

- “Automated video creation for professional storytellers” (Wibbitz)
 - Optimiert für Facebook, Instagram, Snapchat, YouTube, Twitter, OTT, etc.
 - Automatisierte text-to-video produziert rough-cut Video in Sekunden.
- eingesetzt u. a. bei Reuters, Bloomberg, NBC, Forbes



Beispiele: Qualitätssicherung



Beispiel: Qualitäts- sicherung

- Humboldt-Elsevier Advanced Data and Text (HEADT) Centre, Harvard Medical School, Elsevier: Automatisches Erkennen manipulierter Bilder in Forschungspublikationen (-> u.a. Photoshop Plugin ...)
 - Supervised Learning, basierend auf einem Korpus von „guten“ und „schlechten“ Beispielen
- Automatische Prüfung/Validierung von Statistiken: [StatReviewer](#)
- Penelope.ai prüft Manuskript, “lernt” mit zunehmender Anzahl geprüfter Manuskripte dazu

Beispiel: Copy Editing



Beispiel: Copy Editing Taylor & Francis

- Automatische Bewertung der sprachlichen Qualität von Manuskripten („Contextual Copyediting“)
- Entscheidet automatisch welchen Workflow-Pfad das Manuskript nimmt, wie viel Intervention/Bearbeitung notwendig ist
- Reduziert Zeit und Kosten

Beispiel: Discovery Tools

Beispiel: Discovery Tools

Unendlich viele Tools

- Meistens aber nicht klar, wie viel KI da wirklich drin steckt.
- Knowledge Graphs scheinen immer weitere Verbreitung als Navigationstool zu finden.
- Beispiele im Bereich Wissenschaft: [Sparrho](#), [Yewno](#), [Meta](#), [Kyndi](#), [Semantic Scholar](#), [Iris.ai](#), [Science Surveyor](#), ...



Beispiel: Sparrho

Personalisierte Discovery für Forscher/Wissenschaftler

- Vermischung aus Suche, Exploration und Kuratierung
- Sparrho checks 45,000+ journals and repositories hourly to provide easy-to-skim feeds for you to browse at your convenience. And there are thousands of expert-curated Pinboards that summarise the latest, most important research to give you a view into their specific areas of interest.
- Unsubscribe from all those journal eTOCs, clogging up your inbox, and keep your article discovery contextual by grouping keywords, journals and authors into custom Followed Searches—set up takes just seconds.
- Because Sparrho has been specifically designed to cater to the needs of scientists, we will always have all articles published on the journal website,—including figures and graphical abstracts where available—instead of having to rely on incomplete and slow-to-update RSS feeds.



Beispiel: Yewno

- Klassifikator mit hierarchischem Topic Modell
- Visualisierung als (navigierbarer) Graph



Beispiel:
Meta
(Chan
Zuckerberg
Foundation)

Ontologien + Machine Learning



Beispiel:
Nano (Springer
Nature)

- <https://nano.nature.com>
- Nutzt Unsilo, um zu ähnlichen Content zu finden, zu klassifizieren und Konzepte zu extrahieren

Beispiel: Science Direct (Elsevier)

„Topic Pages“ werden automatisch generiert

→ Orientierung, Navigation, Organisation,
Einordnung

- Herausforderungen:
 - Hoher Durchsatz, deshalb vollautomatisch
 - Topics erkennen
 - Definitionen erkennen
 - Disambiguierung





Beispiel: Science Direct (Elsevier)

- Benutzt über 15 verschiedene Taxonomien
- NLP + Heuristiken zur Textanalyse
- Feedback von menschlichen Subject Matter Experts + div. Metriken

Beispiel: Science Direct (Elsevier)

- Start July 2017
- August 2017: ca 270.000 visits - Januar 2018: ca 6.000.000 visits pro Monat
- 60% return rate
- Topic pages machen ca 8% des Science Direct Traffics aus (April 2018)
- Durchschnittl. Zahl der Views und Engagement ist 155% höher als bei anderen SD Content Seiten
- 92% der befragten Researcher sagen, dass ihnen Topic Pages bei der Erreichung ihrer Ziele helfen



Beispiele: Chatbots

- Einsatzmöglichkeit im Publishing: Erschließen und Ausspielen von Informationen/Content über Chatbots
- Erschließung neuer Zielgruppen und Kanäle
- ... dazu muss auch die Nutzeranfrage “verstanden” werden – das klappt noch nicht immer

Beispiel: Text & Data Mining



Beispiele: Text Mining

- Listening bei der Rheinischen Post, Technik von Convidera
 - Wer spricht wo im Web über mich? Was muss ich wissen, um Kampagnen oder Krisen zu managen? Was sind relevante Inhalte, über die meine Kunden kanalübergreifend diskutieren wollen?
 - Wie gut performen meine Prozesse und Kanäle?
- „Unsere Listening-Center-Lösung beantwortet in Echtzeit Fragen von ‚Wo wird gesprochen?‘ bis zu ‚Wer ist Meinungsmacher für mein Thema?‘“

Beispiele: Text Mining

Bestseller-Prognose (QualiFiction), wird derzeit mit Buchverlagen entwickelt, Start im Oktober/November 2018

- Abverkaufsvorhersage: Reduzierung des Fehlers in der Auflagenkalkulation um mindestens 10 % in den ersten 12 Monaten nach Bucherscheinen
- Stil-Analyse (Komplexität, poetische/sachliche Sprache)
- Sentiment-Analyse (Leichtigkeit/Schwere, Happy End ja/nein)
- Entitäten-Analyse (Beziehungen zwischen Personen, Zeit, Ort)
- Themen-Analyse (Krimi oder Romanze, Filterung nach Themen)
- „Auf Grundlage von künstlicher Intelligenz stellt unsere Software einen Zusammenhang her zwischen den Erkenntnissen aus bereits bestehenden Werken und deren Kennzahlen aus Abverkauf und Bestseller-Platzierung. Dieser Zusammenhang wird auf neue Manuskripte angewendet und eine Erfolgswahrscheinlichkeit ermittelt.“

Beispiele: Data Mining

- Vertrieboptimierung im Buchhandel (Readbox)
- Ziel: automatische Metadatenoptimierung, Preisfindung und -gestaltung, Zielgruppenmarketing
- Vorgehensweise:
 - Ermittlung von "Text-DNA" durch Textanalyse
 - Abgleich mit täglich abgefragten externen Marktdaten (Suchanfragen, Preis-/Absatz-Statistiken, Genre-, Zielgruppen- und andere Klassifizierungen, Umsatzanalysen ...)
 - Optimierung von Produkten, Meta-/Katalogdaten sowie Marketingaktivitäten (Conversions)



A thick, dark gray diagonal line runs from the bottom-left corner towards the top-right corner, separating the dark gray header area from the white main content area.

Vertieftes Beispiel: Backend Optimierung für eine A&I Datenbank



Beispiel 1: The Institute of Engineering and Technology

- Hochwertig kuratierte A&I Datenbank für Ingenieurwissenschaften, Informatik und Physik (+ verwandte Fächer)
- 40+ Jahre alt
- 17+ Millionen Abstracts
- 100kt++ Metadaten
- Hundert++ Personenjahre Investition in Qualität
- See also: Kohl, C. / Smith, D.: Concurrent 4F: A Long-Standing Promise Finally Fulfilled? Bots, Agents, etc. – Artificial Intelligence on the Rise. Session at the SSP Conference 2017. URL: <https://www.sspnet.org/events/past-events/annual-meeting-2017/2017-schedule/concurrent-4f-a-long-standing-promise-finally-fulfilled/>, last accessed 2.6.2017.

Vertieftes Beispiel: Tools für das Editorial

Beispiel 2: Springer Nature

- “Lecture Notes in Computer Science” (LNCS)
- Eine der wichtigsten Reihen im Bereich Informatik überhaupt
- Sehr hohes Volumen, ca 780 Bücher / Jahr
- Klassifikation ist essentiell (Auffindbarkeit, SEO, etc.), wird traditionell durch Lektoren manuell ausgeführt

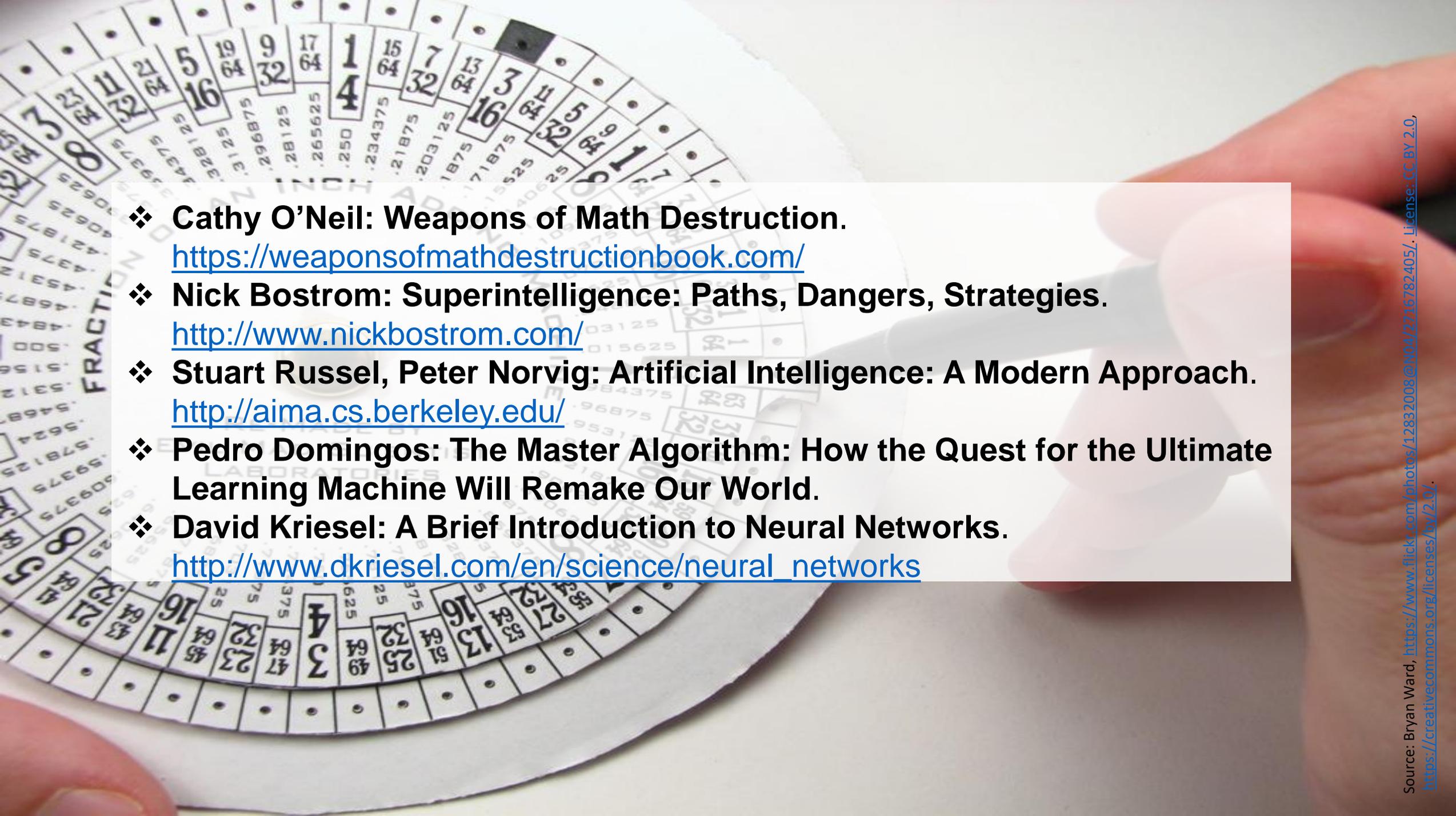


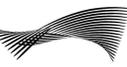
Smart Topic Miner (STM)

- <https://technologies.kmi.open.ac.uk/rexplore/smart-topic-miner/>
- Osborne F., Salatino A., Birukou A., Motta E. (2016) Automatic Classification of Springer Nature Proceedings with Smart Topic Miner. In: Groth P. et al. (eds) The Semantic Web – ISWC 2016. ISWC 2016. Lecture Notes in Computer Science, vol 9982. Springer, Cham.
- URL: <http://oro.open.ac.uk/id/eprint/46823>, last accessed 2.5.2018.

Smart Book Recommender (SBR)

- <http://rexplore.kmi.open.ac.uk/SBR-demo/>
- Osborne, Francesco; Thanapalasingam, Thiviyan; Salatino, Angelo; Birukou, Aliaksandr and Motta, Enrico (2017). Smart Book Recommender: A Semantic Recommendation Engine for Editorial Products. In: International Semantic Web Conference (ISWC) 2017, 21-25 Oct 2017, Vienna, Austria. URL: <http://oro.open.ac.uk/id/eprint/50892>, last accessed 2.5.2018.

- 
- ❖ **Cathy O'Neil: Weapons of Math Destruction.**
<https://weaponsofmathdestructionbook.com/>
 - ❖ **Nick Bostrom: Superintelligence: Paths, Dangers, Strategies.**
<http://www.nickbostrom.com/>
 - ❖ **Stuart Russel, Peter Norvig: Artificial Intelligence: A Modern Approach.**
<http://aima.cs.berkeley.edu/>
 - ❖ **Pedro Domingos: The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World.**
 - ❖ **David Kriesel: A Brief Introduction to Neural Networks.**
http://www.dkriesel.com/en/science/neural_networks



Best Practice

Immer vom Geschäftsmodell her denken, nicht von der Technik bzw. einzelnen Angeboten treiben lassen

Zuverlässigkeit des Anbieters ermitteln: Marktreife, Hintergrund, Referenzen; Referenzkunden kontaktieren

Proof-of-Concept oder Probetrieb vereinbaren (falls anwendbar)

Zu große/ambitionierte Projekte („moonshots“) vermeiden

Urheber- und Verwertungsrechte klären

Ist die Datenbasis wirklich repräsentativ? Ausgewogen?

Auf Datenschutz, Compliance, Sicherheit achten

Hajo Hoffmann & Christian Kohl

KI-Report 2018

**Marktüberblick
zum Einsatz von künstlicher Intelligenz
in Medien und Publishing**

Eine Übersicht zum Angebot im deutschsprachigen Raum



KI-Report 2018

Hajo Hoffmann, Christian Kohl: KI-Report 2018. Marktüberblick zum Einsatz von künstlicher Intelligenz in Medien und Publishing. Eine Übersicht zum Angebot im deutschsprachigen Raum

- Enthält detaillierte Anbieterprofile und damit auch Anregungen für das eigene Geschäft
- 102 Seiten, PDF mit Wasserzeichen, € 490 zzgl. ges. USt.
- Bestellung über studien@hspartner.de
- <http://hspartner.de/jsp292/site/hauptmenue/publikationen/studien/ki>

Christian Kohl

Beratung und Projektmanagement

Geisbergstr. 29

10777 Berlin

christian@kohl.consulting

www.kohl.consulting

Tel.: +49 (0)1577 1848842

