

KÜNSTLICHE INTELLIGENZ – HERAUSFORDERUNGEN FÜR DIE VERANTWORTUNGSZUSCHREIBUNG

PROF. DR. SUSANNE HAHN

HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF

Algorithmen schneiden besser ab als junge Ärzte



Maschinen lernen schnell

KI könnte Heilungschancen massiv verbessern

SEITE 2/2 **Müssen Radiologen um ihren Job fürchten?**

Das Ausgangsproblem

- * Wem ist Verantwortung zuzuschreiben, wenn ein Patient beispielsweise aufgrund einer Einschätzung durch einen Algorithmus nicht zu einer Präventivuntersuchung überwiesen worden war und er einen Herzinfarkt erlitten hat?
- * Wer trägt Verantwortung, wenn eine Patientin nach einer Einschätzung durch eine maschinelle Bilderkennung als krebserkrankt gilt und zu weiteren Eingriffen wie Gewebeentnahmen geschickt wird, ohne dass sich ein Tumor bzw. eine Tumorstufe findet.
- * Wer ist verantwortlich, wenn eine bösartige Veränderung nicht entdeckt wurde?
- * Wer ist verantwortlich, wenn aufgrund automatisierter Entscheidung ein Medikament verabreicht wurde, das zum Tod des Patienten geführt hat?
- * Wer trägt Verantwortung, wenn jemand mit 55 Jahren die Diagnose erhält, in der Zukunft wahrscheinlich an Alzheimer zu erkranken, er aber mit 89 an einer Lungenentzündung bei guter geistiger Gesundheit stirbt?

Gliederung

1. Methodische Vorbemerkung zur erforderlichen Sachkenntnis
2. Was ist maschinelles Lernen?
3. Die Zuschreibung von Verantwortung
4. Der Umgang mit Zurechnungslücken

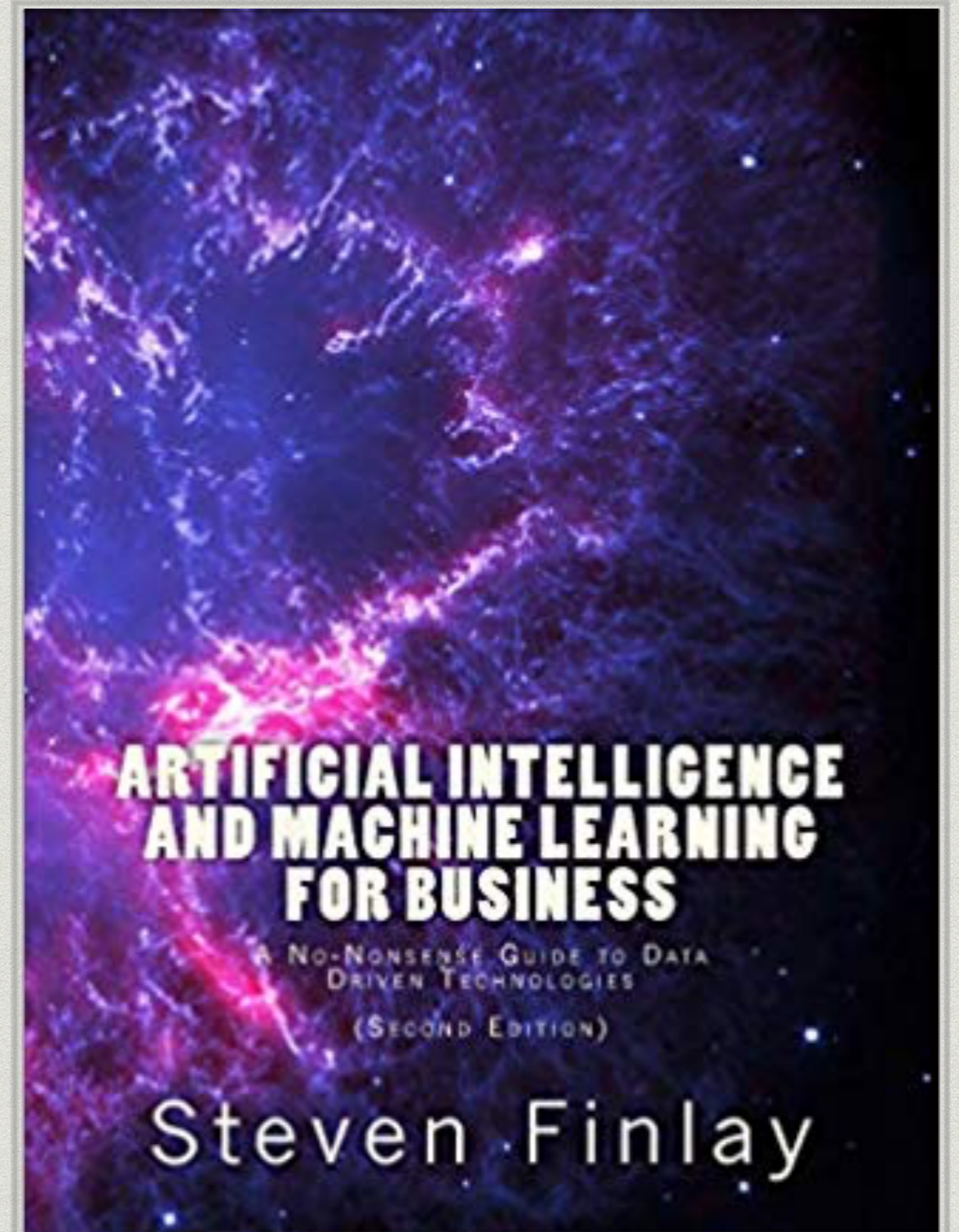
1. Methodische Vorbemerkung zur erforderlichen Sachkenntnis

Angewandte Philosophie:

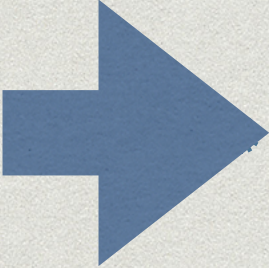
Verallgemeinernde Überlegungen werden auf spezifische Handlungsfelder (z.B. Organtransplantation, Gentherapie, Umwelttechnik, Einsatz maschinellen Lernens) bezogen oder aus ihnen gewonnen

FRAGE: Wie tief und umfassend muss man sich mit dem Wissensfundus und den Verfahren des Gegenstandsbereichs auseinandersetzen, um begriffliche Sortierungen und Einschätzungen vornehmen zu können?

Quelle für das Beispiel
und die Aufbereitung
maschinellen Lernens



Künstliche Intelligenz?

- * Vincent Müller im ersten Vortrag:
 - * Starke KI: Versprechen der klassischen KI eingelöst, mit Schmerzempfindung etc.
 - * Schwache KI: KI als Modell natürlicher Kognition
 - * Technische KI (fremde KI): KI als intelligente Problemlösung
-  Hier: Künstliche Intelligenz im dritten Sinn und zwar als **maschinelles Lernen**

2. Was ist maschinelles Lernen?

„Maschinelles Lernen ist die Verwendung von mathematischen Verfahren (Algorithmen) zur Analyse von Daten. Ziel ist es, nützliche Muster (Beziehungen oder Korrelationen) zwischen verschiedenen Daten zu entdecken. Wenn die Beziehungen einmal identifiziert sind, können diese dazu verwendet werden, Rückschlüsse auf das Verhalten neuer Fälle zu ziehen, wenn diese sich präsentieren.“

(Übersetzung S.H., S. 33)

Steven Finlay: Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies. iBooks. <https://itunes.apple.com/de/book/artificial-intelligence-machine-learning-for-business/id1338327760?mt=11>

Elemente in der **Anwendung** maschinellen Lernens

- * Dateninput
- * Datenaufbereitung
- * Voraussagemodelle
- * Entscheidungsregeln
- * „Output“

Maschinelles Lernen und Voraussagemodell

- * Es gibt unterschiedliche Voraussagemodelle und es gibt eine Vielzahl von Techniken maschinellen Lernens und von Algorithmen, um ein Modell zu gewinnen.
- * Gemeinsam ist allen Modellen, dass die Voraussage, die generiert wird, durch eine Zahl repräsentiert wird.
- * Die Modelle unterscheiden sich hinsichtlich der Nachvollziehbarkeit

Das Beispiel: Prognose von Herzerkrankungen

- * Beobachtungsdaten:

- * 500.000 medizinische Berichte von Personen ohne Herzerkrankung

- * Ergebnisdaten:

- * Gesundheitsverlauf dieser 500.000 Personen über 5 Jahre in Bezug darauf, ob diese eine Herzerkrankung entwickeln oder nicht - bei 30.000 Personen ist dies der Fall

- * Einsatz maschinellen Lernens:

- * Gegenüberstellung von **Beobachtungsdaten** und **Ergebnisdaten**
 - * Welche Parameter lassen in welchem Ausmaß auf eine Erkrankung schließen?

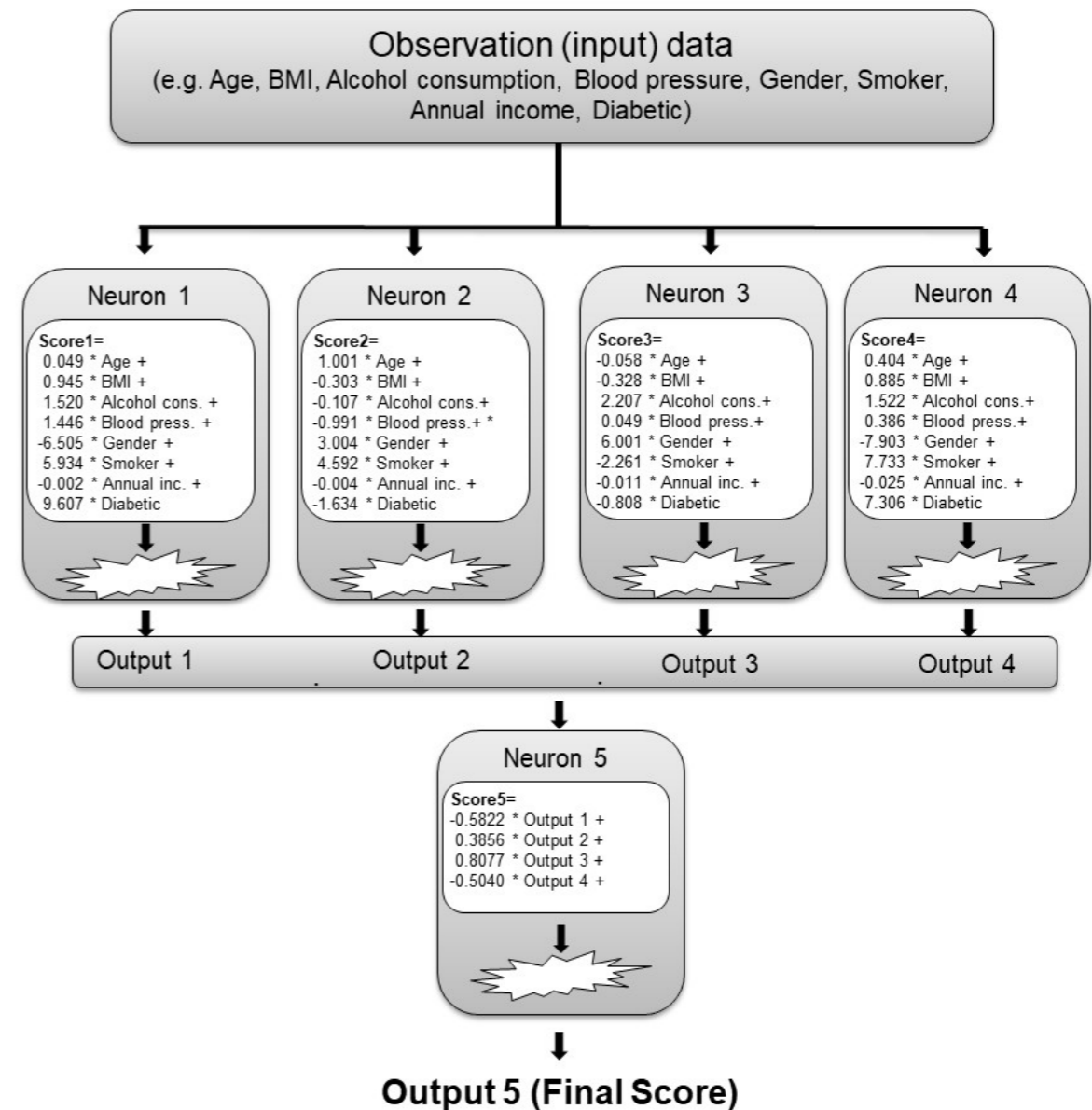
Figure 2. A scorecard for predicting heart disease.

Starting score (constant)	350		
Age (years)		Gross annual income (\$)	
<23	-57	< \$22,000	11
23 - 32	-26	\$22,001 - \$38,000	6
33 - 41	0	\$38,001 - \$60,000	0
42 - 48	7	\$60,001 - \$94,000	-3
49 - 57	15	\$94,001 - \$144,000	-5
58 - 64	24	>\$144,000	-6
65 - 71	31		
>71	65	Smoker ?	
		Yes	37
		No	0
BMI (weight in kg / {height in metres}²)		Diabetic ?	
<19	2	Yes	21
19 - 26	0	No	0
27 - 29	8		
30 - 32	14	Cholesterol level (mg per decilitre of blood)	
>32	29	Low (< 160 mg)	-2
		Normal (160 - 200 mg)	0
Gender		High (201 - 240 mg)	19
Male	2	Very high (>240 mg)	32
Female	-4		
Alcohol consumption (units/week)		Blood pressure	
0	4	Low (below 90/60)	3
1 - 12	0	Average (between 90/60 and 140/90)	0
13 - 24	5	High (above 140/90)	36
25 - 48	10		

Scorecard

Künstliches neuronales Netz

Figure 8. A Neural Network Model



Quelle: Steven Finlay. „Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies.“ iBooks. <https://itunes.apple.com/de/book/artificial-intelligence-machine-learning-for-business/id1338327760?mt=11>

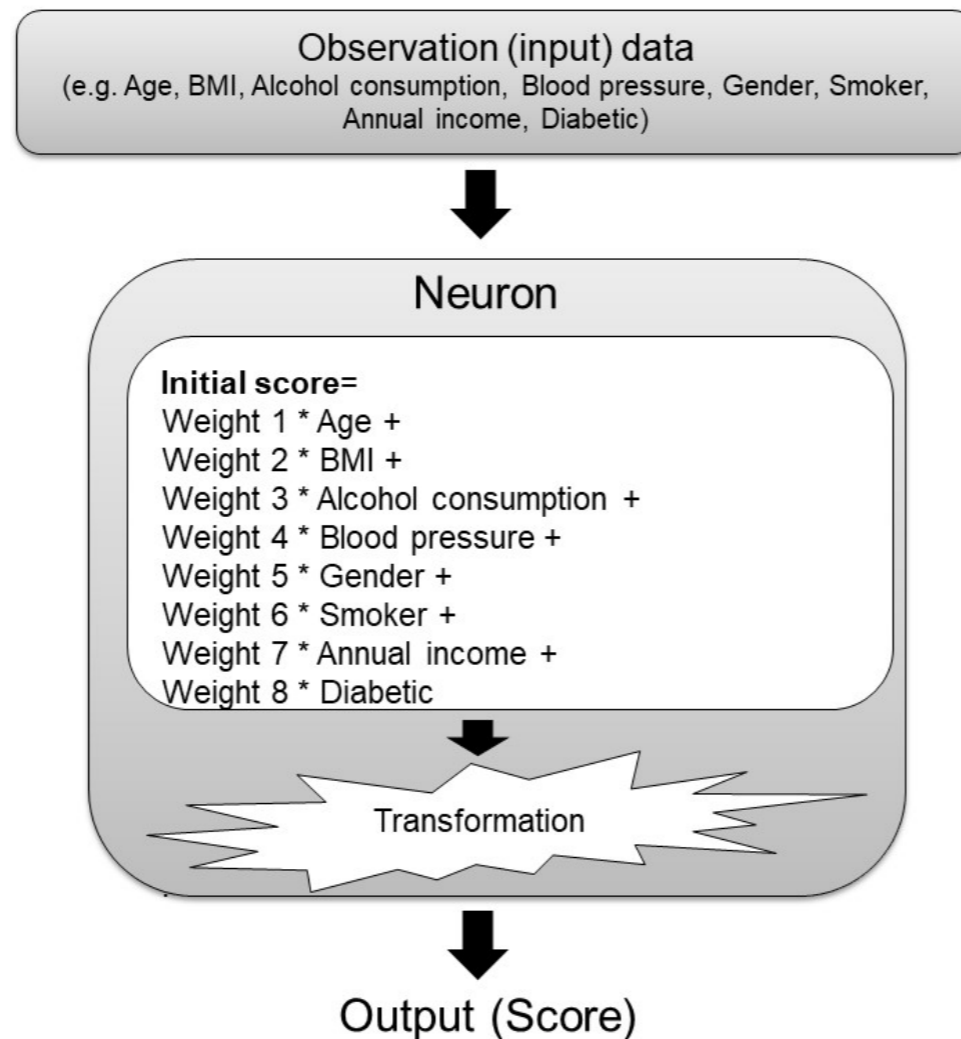
Figure 2. A scorecard for predicting heart disease.

Starting score (constant)	350		
Age (years)		Gross annual income (\$)	
<23	-57	< \$22,000	11
23 - 32	-26	\$22,001 - \$38,000	6
33 - 41	0	\$38,001 - \$60,000	0
42 - 48	7	\$60,001 - \$94,000	-3
49 - 57	15	\$94,001 - \$144,000	-5
58 - 64	24	>\$144,000	-6
65 - 71	31		
>71	65	Smoker ?	
		Yes	37
		No	0
BMI (weight in kg / {height in metres}²)		Diabetic ?	
<19	2	Yes	21
19 - 26	0	No	0
27 - 29	8		
30 - 32	14	Cholesterol level (mg per decilitre of blood)	
>32	29	Low (< 160 mg)	-2
		Normal (160 - 200 mg)	0
Gender		High (201 - 240 mg)	19
Male	2	Very high (>240 mg)	32
Female	-4		
Alcohol consumption (units/week)		Blood pressure	
0	4	Low (below 90/60)	3
1 - 12	0	Average (between 90/60 and 140/90)	0
13 - 24	5	High (above 140/90)	36
25 - 48	10		

Quelle: Steven Finlay. „Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies.“ iBooks. [https://itunes.apple.com/de/book/artificial-intelligence-machine-learning-for-business/](https://itunes.apple.com/de/book/artificial-intelligence-machine-learning-for-business/id1338327760?mt=11)

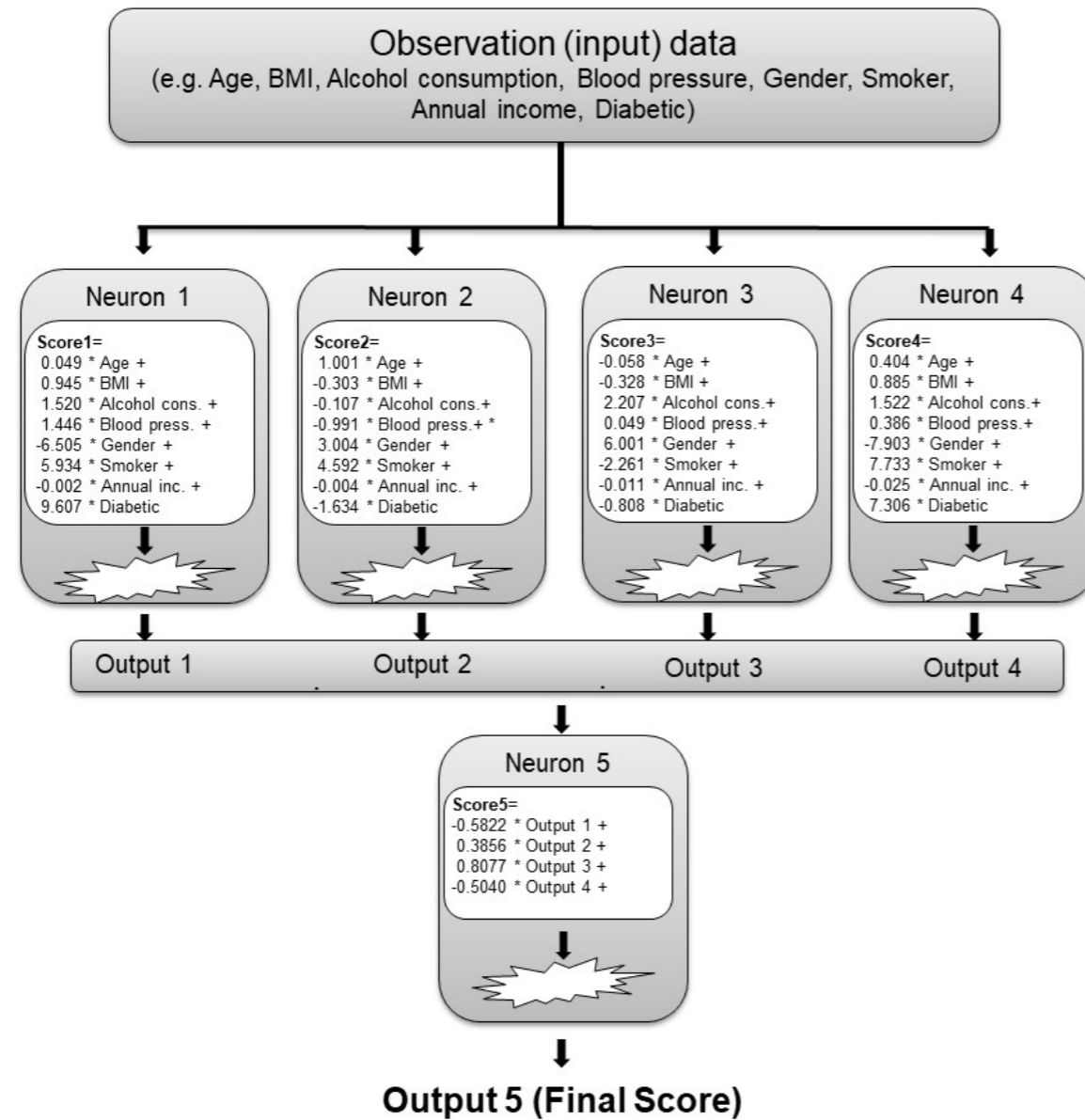
[id1338327760?mt=11](https://itunes.apple.com/de/book/artificial-intelligence-machine-learning-for-business/id1338327760?mt=11)

Figure 7. An Artificial Neuron



Quelle: Steven Finlay. „Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies.“ iBooks. <https://itunes.apple.com/de/book/artificial-intelligence-machine-learning-for-business/id1338327760?mt=11>

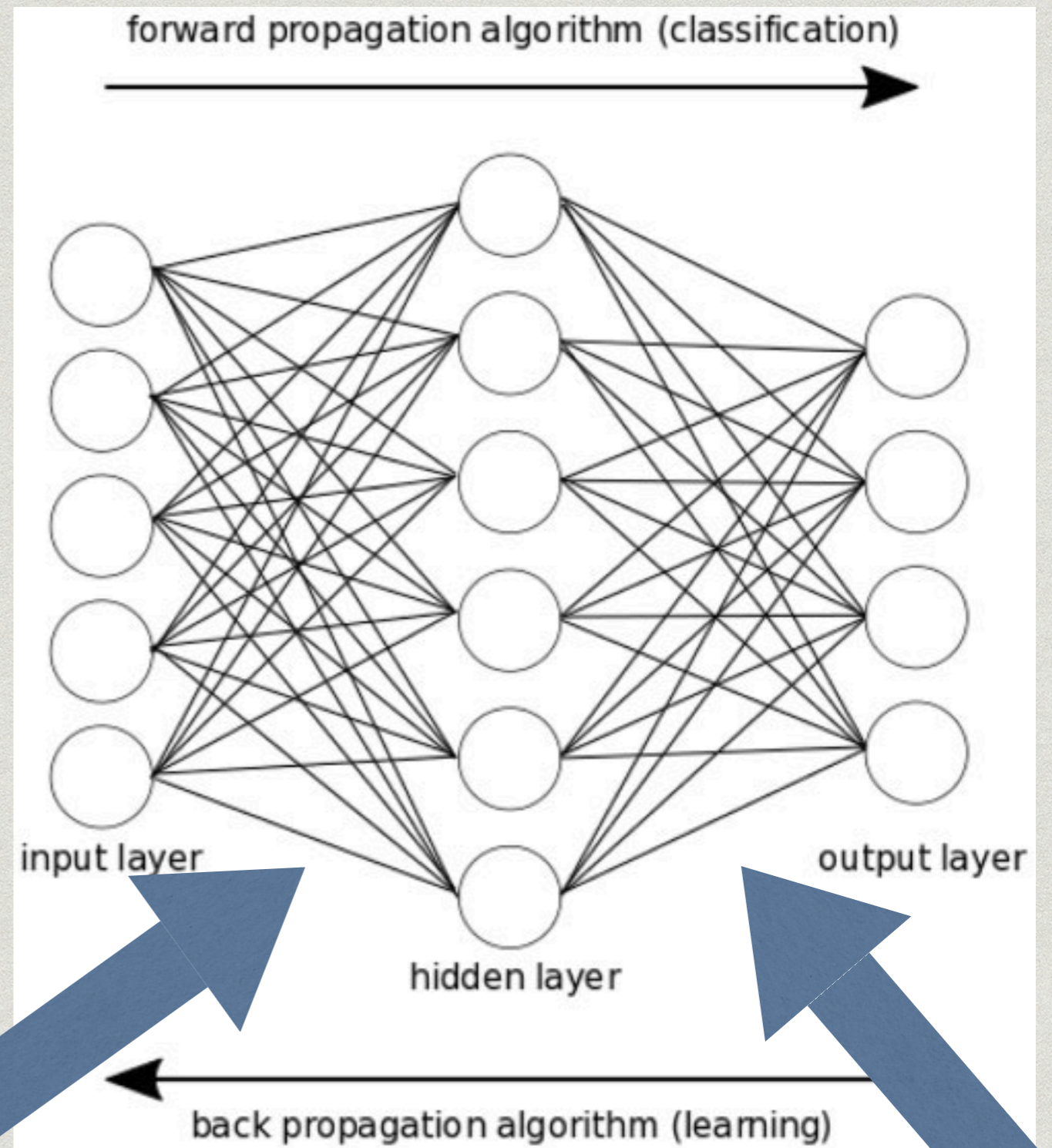
Figure 8. A Neural Network Model



Quelle: Steven Finlay. „Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies.“ iBooks. <https://itunes.apple.com/de/book/artificial-intelligence-machine-learning-for-business/id1338327760?mt=11>

Standarddarstellung eines neuronalen Netzes

Quelle: Jenna Burrell,
How the machine >thinks<:
Understanding opacity in machine learning
algorithms, in:
Big Data & Society, January-June 2016, 1-12



Was >lernt< der Algorithmus?

"Jeder dieser Verbindungslinien ist ein Wert oder eine Gewichtung zugeordnet. Die optimalen Werte für die Matrix der Gewichte sind das, was der Lernalgorithmus lernt. Was 'optimal' ist, wird durch den Satz von Gewichten definiert, die eine möglichst genaue Klassifizierung von Eingaben (...) erzeugen."

**Quelle: Jenna Burrell,
How the machine >thinks<:
Understanding opacity in machine learning algorithms, in:
Big Data & Society, January-June 2016, 1-12, Übersetzung S.H.**

Figure 3. A score distribution table.

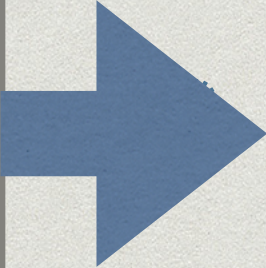
Group	Score range		Number of people	% of population	Number with heart disease after	% with heart disease after 5 yrs.
	From	To				
1	0	300	55,950	11.19%	40	0.07%
2	301	320	56,606	11.32%	68	0.12%
3	321	340	59,700	11.94%	129	0.22%
4	341	360	58,706	11.74%	216	0.37%
5	361	380	64,429	12.89%	403	0.63%
6	381	400	52,749	10.55%	575	1.09%
7	401	420	34,089	6.82%	600	1.76%
8	421	440	21,107	4.22%	632	2.99%
9	441	460	17,269	3.45%	878	5.09%
10	461	480	23,364	4.67%	2,020	8.65%
11	481	500	17,477	3.50%	2,553	14.61%
12	501	520	13,554	2.71%	3,366	24.84%
13	521	540	7,103	1.42%	3,463	48.76%
14	541	560	8,260	1.65%	6,587	79.74%
15	561	999	9,637	1.93%	8,469	87.88%
Total	Total		500,000		30,000	6.0%

Quelle: Steven Finlay. „Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies.“ iBooks. <https://itunes.apple.com/de/book/artificial-intelligence-machine-learning-for-business/id1338327760?mt=11>

Durch Einladung von 5 % der Bevölkerung (Risikoklassen 13-15) kann man 62% der potentiell Erkrankten identifizieren

Nachvollziehbarkeit?

- * Scorecard: ja
- * Neuronales Netz: In der Regel nein, man spricht von einer „Black Box“



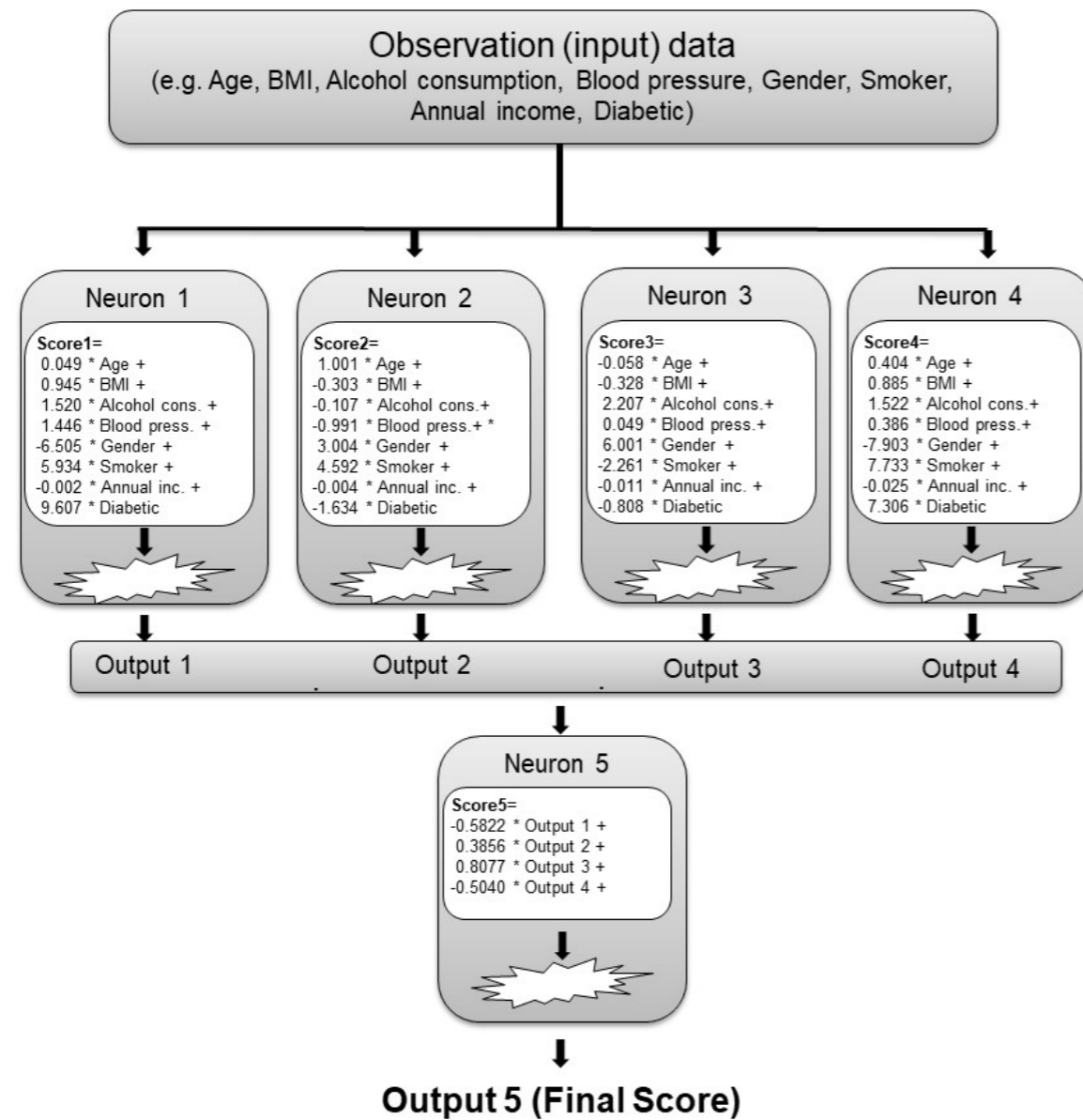
Wahl des Verfahrens maschinellen Lernens in der Hinsicht der Nachvollziehbarkeit nicht beliebig

Beispielberechnung Scorecard

Alter 45	7
Weiblich	-4
BMI 21	0
Alkohol 14 Einheiten pro Woche	5
Einkommen 58.000 \$	0
Nichtraucherin	0
Keine Diabetikerin	0
Cholesterol 170	0
Blutdruck 120/80	0

Score: $350 + 7 + (-4) + 5 = 358$
Risikoklasse 4 (p = 0,37%)

Figure 8. A Neural Network Model



Quelle: Steven Finlay. „Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies.“ iBooks. <https://itunes.apple.com/de/book/artificial-intelligence-machine-learning-for-business/id1338327760?mt=11>

Zwischenbilanz

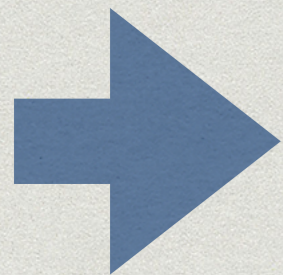
- * Maschinelles Lernen wird zur Mustererkennung und Voraussage eingesetzt und kann sehr effizient sein
- * Unterschiedliche Verfahren haben unterschiedliche Einsatzorte (für Bild- und Texterkennung sind neuronale Netze effizient)
- * Unterschiedliche Modelle gehen mit unterschiedlicher Nachvollziehbarkeit bzw. Erklärbarkeit einher

3. Zuschreibung von Verantwortung

- * Zu unterscheiden:

- * Wie wird Verantwortung faktisch zugeschrieben?

- * Wie wird Verantwortung korrekt zugeschrieben?



Was sind die Standards korrekter Verantwortungszuschreibung?

Korrekte Verantwortungszuschreibung

Ein Akteur kann von einer autorisierten Instanz für einen Schaden unter einer Norm **korrekt** verantwortlich gemacht werden, wenn der Akteur den Schaden mit seiner Handlung verursacht hat, er damit vorsätzlich gegen eine Pflicht verstoßen hat, es keine Rechtfertigungsgründe gibt und der Akteur schuldfähig ist.

RETROSPEKTIVE VERANTWORTLICHKEIT

Herr Krause kann von (einer autorisierten Instanz) für den Verletzungsschaden von Herrn Müller unter dem Verletzungsverbot korrekt retrospektiv verantwortlich gemacht werden genau dann, wenn	
Herr Krause hat mit dem Herunterstoßen des Blumentopfs den Verletzungsschaden verursacht	Kausalität
und Herr Krause ist verpflichtet, gemäß dem Verletzungsverbot zu agieren	Pflicht/Norm
und Herr Krause hat mit dem Herunterstoßen des Blumentopfs gegen das Verletzungsverbot agiert	Pflichtverletzung/ Normverstoß
und Herr Krause hat den Blumentopf absichtlich heruntergestoßen	Absicht
und Herr Krause wusste (musste wissen), dass er mit dem Herunterstoßen des Blumentopfs einen Verletzungsschaden herbeiführt	Wissen
und...	fehlende Rechtfertigungsgründe, Schuldfähigkeit

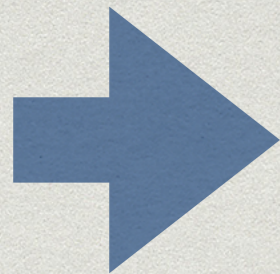
Ärztliches Handeln und durch maschinelles Lernen gestützte Diagnostik

Wer ist für mögliche Schäden verantwortlich?

- * Das klassische Konzept der Verantwortungszuschreibung ist nicht anwendbar
- * Es entstehen **Zurechnungslücken** für Schäden

Funktion der Verantwortungszuschreibung

- * Nicht nur retrospektiv Herstellung eines Ausgleichs für Schäden, sondern
- * Verhinderung zukünftiger Schäden, somit



Handlungssteuerung

- * **Wie lässt sich das Handeln so umorganisieren, dass unerwünschte Folgen erlaubten Handelns verhindert oder minimiert werden?**

4. Zum Umgang mit Zurechnungslücken

Ein Blick in die Vergangenheit:
Zum Umgang mit Zurechnungslücken
bei Dampfkesselunfällen

„Kesselgesetze“

(Nr. 649.) Bekanntmachung, betreffend allgemeine polizeiliche Bestimmungen über die Anlegung von Dampfkesseln. Vom 29. Mai 1871.

Auf Grund der Bestimmung im [§. 24. der Gewerbeordnung für den Norddeutschen Bund vom 21. Juni 1869.](#) hat der Bundesrath nachstehende
Allgemeine polizeiliche Bestimmungen über die Anlegung von Dampfkesseln
erlassen.

I. Bau der Dampfkessel.

§. 1. Kesselwandungen.

Die vom Feuer berührten Wandungen der Dampfkessel, der Feuerröhren und der Siederöhren dürfen nicht aus Gußeisen hergestellt werden, sofern deren lichte Weite bei cylindrischer Gestalt 25 Centimeter, bei Kugelgestalt 30 Centimeter übersteigt.

Die Verwendung von Messingblech ist nur für Feuerröhren, deren lichte Weite 10 Centimeter nicht übersteigt, gestattet.

§. 2. Feuerzüge.

Die um oder durch einen Dampfkessel gehenden Feuerzüge müssen an ihrer höchsten Stelle in einem Abstand von mindestens 10 Centimetern unter dem festgesetzten niedrigsten Wasserspiegel des Kessels liegen. Bei Dampfschiffskesseln von 1 bis 2 Meter Breite muß der Abstand mindestens 15 Centimeter, bei solchen von größerer Breite mindestens 25 Centimeter betragen.

Diese Bestimmungen finden keine Anwendung auf Dampfkessel, welche aus Siederöhren von weniger als 10 Centimeter Weite bestehen, sowie auf solche Feuerzüge, in welchen ein Erglühen des mit dem Dampfraum in Berührung stehenden Theiles der Wandungen nicht zu befürchten ist. Die Gefahr des Erglühens ist in der Regel als ausgeschlossen zu betrachten, wenn die vom Wasser bespülte Kesselfläche, welche von dem Feuer vor Erreichung der vom Dampf bespülten Kesselfläche bestrichen wird, bei natürlichem Luftzug mindestens zwanzigmal, bei künstlichem Luftzug mindestens vierzigmal so groß ist, als die Fläche des Feuerrostes.

II. Ausrüstung der Dampfkessel.

§. 3. Speisung.

An jedem Dampfkessel muß ein Speiseventil angebracht sein, welches bei Abstellung der Speisevorrichtung durch den Druck des Kesselwassers geschlossen wird.

Einführung des Gefährdungstatbestands

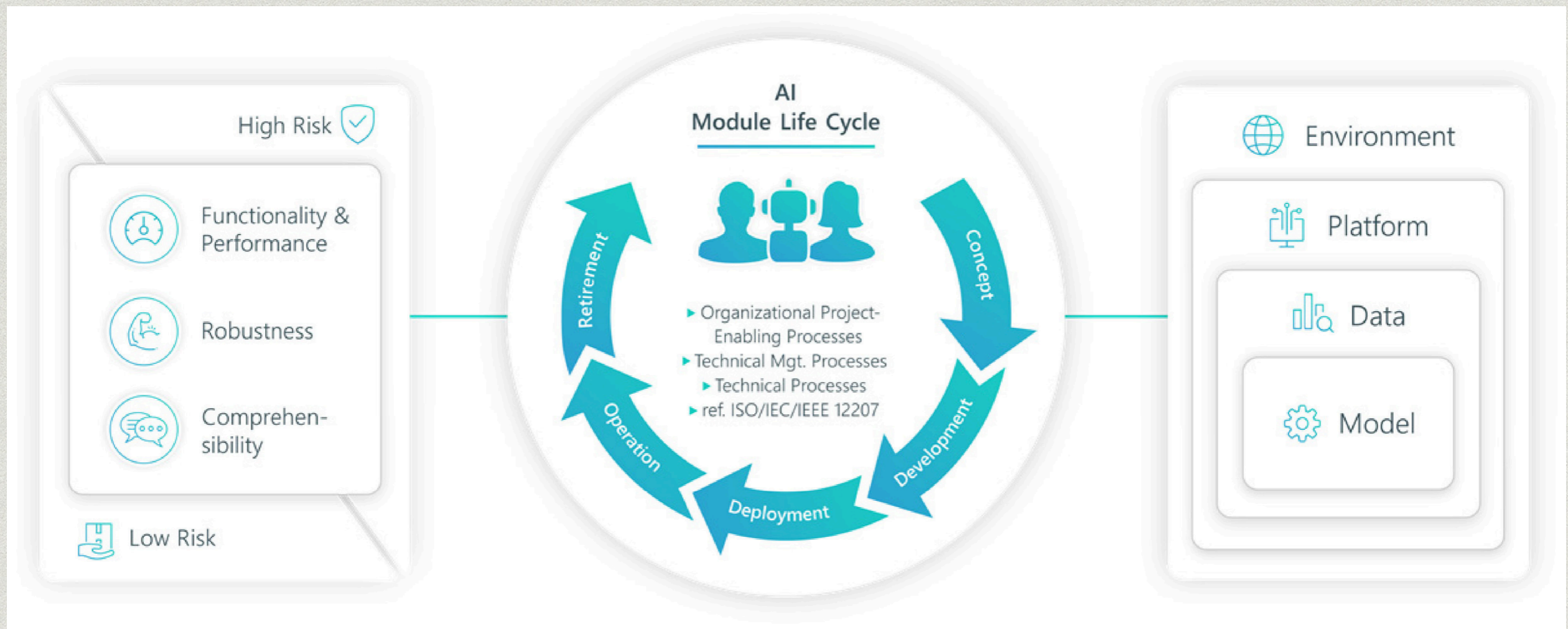
„Die Gesellschaft ist zum Ersatz verpflichtet für allen Schaden, welcher bei der Beförderung auf der Bahn, an den auf derselben beförderten Personen und Gütern, oder auch an anderen Personen und deren Sachen, entsteht und sie kann sich von dieser Verpflichtung nur durch den Beweis befreien, dass der Schaden entweder durch die die eigene Schuld des Beschädigten oder durch einen unabwendbaren äußeren Zufall bewirkt worden ist. Die gefährliche Natur der Unternehmung selbst ist als ein solcher, von dem Schadenersatz befreiender Zufall nicht zu betrachten.“ (Preußisches Eisenbahngesetz, 3.11.1838, § 25) nach: Olaf von Gadow, Die Zähmung des Automobils durch die Gefährdungshaftung, Berlin 2002, 68)

Antworten auf die Herausforderung durch
Zurechnungslücken –
Umorganisation des **erlaubten** Handelns durch:

- * Konstruktionsbedingungen
- * Rollenpflichten
- * Einführung eines Gefährdungstatbestands

Übertragbarkeit der Strategie auf
den Umgang mit Algorithmen?

Eine DIN-Richtlinie als Analogon zu den Konstruktionsvorschriften?



DIN SPEC 92001-1:2019-04

Lässt sich die DIN-Richtlinie als leistungsäquivalentes Gegenstück zu den Konstruktionsbedingungen für Kessel lesen?

- * Können Entwickler ihr Handeln so steuern, dass Schäden weitgehend vermieden werden, der Nutzen der Technik aber erhalten bleibt?
- * Welche Rolle spielt die Bedingung der Nachvollziehbarkeit?
- * Ist Nachvollziehbarkeit herzustellen?
- * Könnte an die Stelle der Nachvollziehbarkeit die Forderung nach der epistemischen Güte der Algorithmen treten? Worin besteht diese?

Automatisierte Vorhersagemodelle als Gefährdungstatbestände?

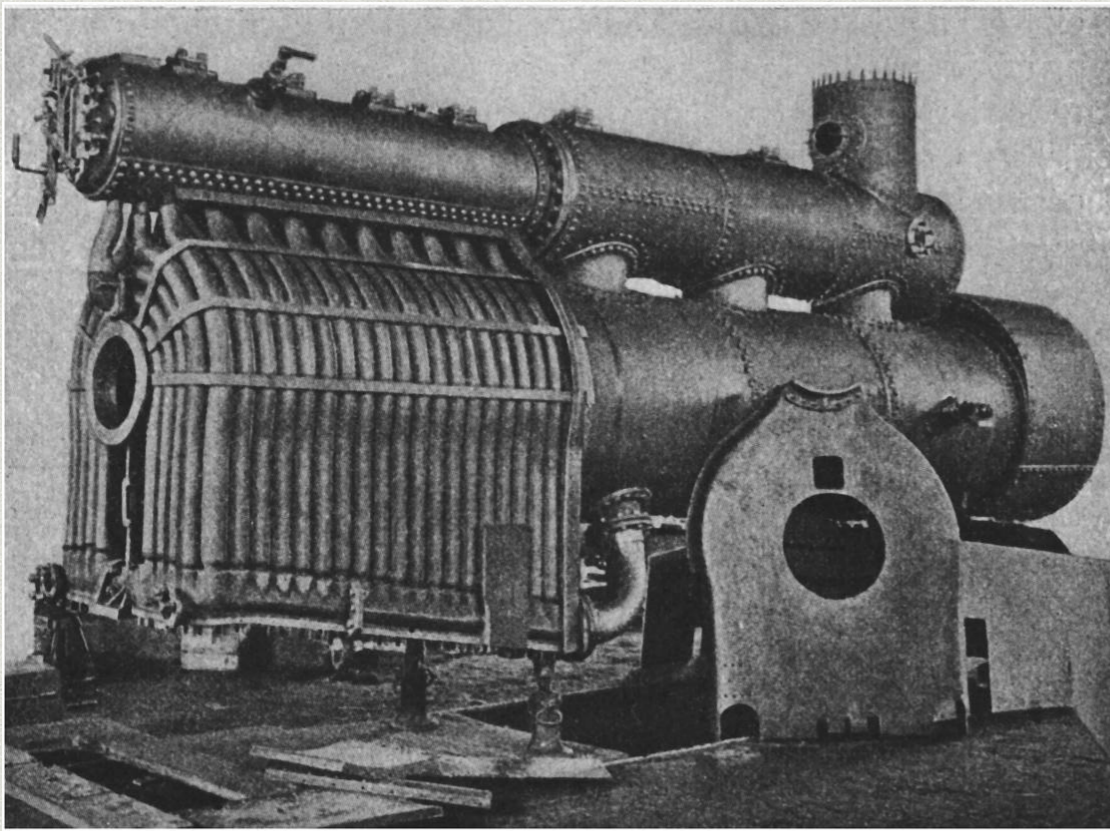


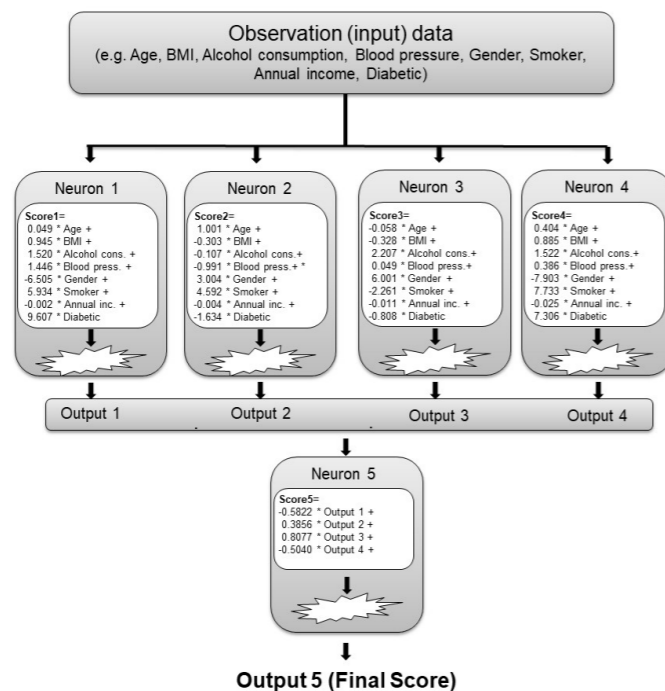
Abb. 208. Brotan-Kessel.

Gefährdung Dampfkessel



Gefährdung „Hund“

Figure 8. A Neural Network Model



Einsatz eines Algorithmus als Gefährdungstatbestand?

Gefährdungshaftung

- * An wen richtet sie sich? Welche Konsequenzen sind erwartbar?
- * Ist ein Know-How über die Handhabung der Technik vorausgesetzt?

Rollenpflichten

- * Rollenpflichten für wen?
 - * a) Dateningenieure und Mediziner
 - * b) Ärztinnen und Ärzte in der Behandlung?

Zwei Fragen als Bilanz:

- A. Sollen Handlungsbeschränkungen des eigentlich erlaubten und prinzipiell erwünschten Handelns eingeführt werden, um die unerwünschten Folgen, die Schäden, zu vermeiden?
- B. Welche Akteure können und sollen in ihrem Handeln wie angeleitet werden?

Zwei Fragen als Bilanz:

A. Sollen Handlungsbeschränkungen des eigentlich erlaubten und prinzipiell erwünschten Handelns eingeführt werden, um die unerwünschten Folgen, die Schäden, zu vermeiden?

Hierbei handelt es sich um eine Abwägungsentscheidung von ethischer Relevanz

- * Ziele
- * Interessen potentiell betroffener Individuen
- * Bestehende Rechte und Pflichten
- * Umsetzbarkeit

Zwei Fragen als Bilanz:

Die erste Frage ist mit der zweiten verknüpft:

B. Welche Akteure **können** und sollen in ihrem Handeln wie angeleitet werden?

Hierzu ist Auskunft von Informatikern, Datenwissenschaftlern und Fachwissenschaftlern erforderlich:

Lässt sich Transparenz herstellen?

Lassen sich andere Gütekriterien für Algorithmen zur Entscheidungsfindung formulieren?

Literaturempfehlungen zur Vortragsreihe

