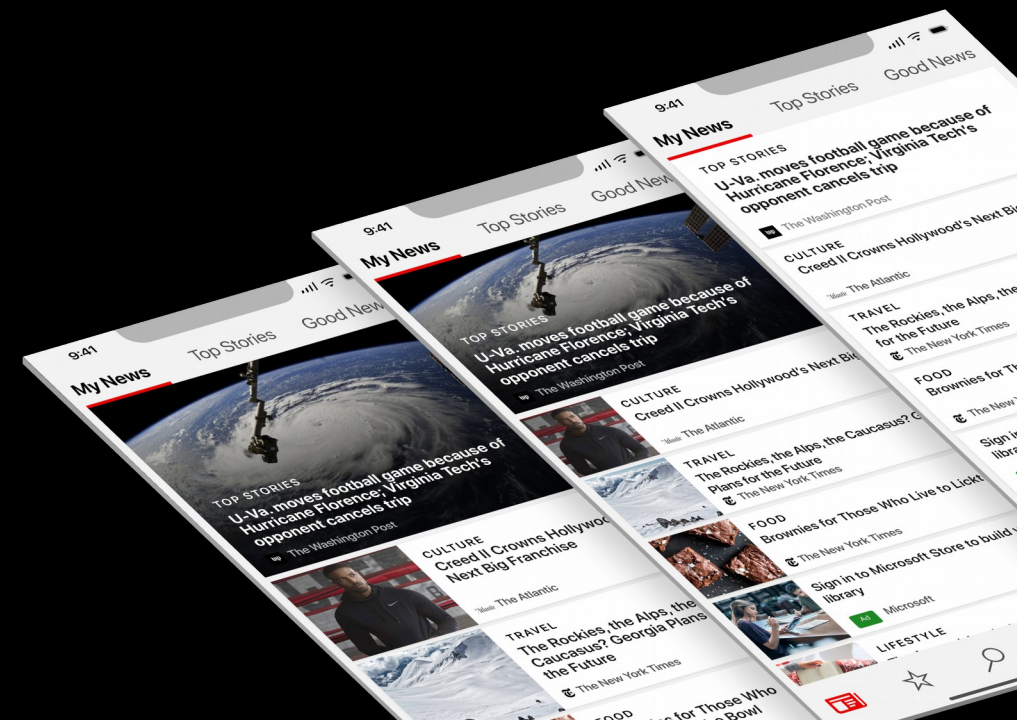


Gender und Fairness in personalisierten Medien

Ulrike Irmeler

Director, Product Management
Microsoft News Europe



Trusted news from
the world's best
journalists

“ It’s the best system out there. No propaganda. High-quality real news from thousands of organizations.

Microsoft is leading the industry on this.

”

Henry Blodget, CEO Business Insider tweet in response to the NYT Apple News story



Ein kleines Quiz...

Wie viele Medienbrands werden von Microsoft News in Deutschland, Österreich und der Schweiz kuratiert?

- a. 100
- b. 150
- c. Über 200

*145 aus Deutschland & Österreich und der dt. Schweiz
75 in französischer Sprache*



Microsoft News Medienpartner Deutschland, Schweiz, Österreich



Wie viele Redakteurinnen und Redakteure arbeiten in Europa an Microsoft News?

- a. 20
- b. 80
- c. 120

Microsoft News erreicht 20% der Europäerinnen und Europäer



Microsoft Newsroom in Be

- Seit August 2017
- Redaktion von 100+
- Multikulturell
- Hohe Diversität (über 50% Frauen, 20+ Nationalitäten, Altersspektrum 22-50+)

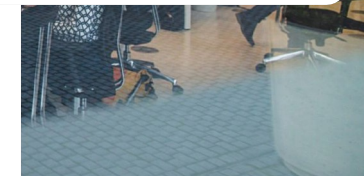
Digitalmedien Microsofts Nachrichtenzentrale

Zehn Länder, acht Sprachen, 55 Millionen Nutzer pro Monat: Der Berliner MSN-Newsroom wartet mit beeindruckenden Zahlen auf - und mit einem besonderen Konzept.

Text: Holger Schellkopf
26. Februar 2018



Versteckter Gigant



In Berlin inszeniert Microsoft mit seinem Newsroom eine ganz besondere Art des Arbeitens – und erreicht damit viele Menschen



TEXT: Holger Schellkopf
FOTOS: Stephan Framme für W&V

Wer nicht genau aufpasst, der geht einfach vorbei. Ist ja auch eher unheimlich, die Schaulustfront am Rande des Areals mit dem adretten Namen Heiligegeistkirche. Hinter den Schaulustern, die gar keine mehr sind, sitzen vergleichsweise junge Menschen vor ziemlich vielen Monitoren – hier in Berlin-Mitte auch nicht unbedingt exotisch. Direkt gegenüber, schon eine gute Spur zufälliger und nicht ganz zufällig, findet sich der Eingang zum Berliner Hauptquartier von C3. Für das Unternehmen gleichzeitig eine Art historische Stätte, denn der Heiligegeistkircheplatz war Heimat von Kirchner-Burkhardt und damit auch Bratschke, dessen, was heute unter C3 firmiert. Die lange Variante der 3 C, also Creative Code and Content legt die Spur zurück auf die andere Seite des Platzes, hin zur Schaulustfront.

Hinter der Front geht es nämlich um Kreativität, um Code, vor allem aber um Content. Seit etwas mehr als einem halben Jahr betreibt C3 mit seiner eigentlich in München ansässigen Tochter Wundermedia im Auftrag von Microsoft den MSN-Newsroom. Die Nachrichtenangebote für MSN, Bing und Edge, die Inhalte für diverse Apps – all das kommt aus dem Berliner Newsroom. Beispielt werden dabei nicht nur die deutschsprachigen Angebote, sondern so gut wie ganz Europa. Inhalte für zehn Länder, das Ganze in acht Sprachen – hier ist innerhalb von recht kurzer Zeit einer der am stärksten international geprägten Newsrooms entstanden, die man in Deutschland überhaupt finden kann.

Eine besondere Aufgabe, eine besondere Herausforderung, daraus macht auch Mike Lukanz, bei C3 als Director Process und Newsroom auch für das MSN-Projekt zuständig, keinen Hehl. „Für uns und natürlich auch für Microsoft war klar, dass ein Projekt in der Größenordnung, mit etwa 70 Mitarbeitern aus

W&V 9.2018 | 31



Wie viele europäische Länder werden von Microsoft News mit Inhalten versorgt?

a. 10

b. 21

c. 30

Deutschland, Schweiz, Österreich, Italien, Niederlande, Belgien, Ungarn, Tschechien, Russland, Polen, Türkei, Schweden, Dänemark, Norwegen, Finnland, Großbritannien, Irland, Frankreich Spanien, Portugal, Griechenland



Microsoft News in Zahlen (weltweit)



4500 Medien-Brands

170,000 Artikel, Videos, Bildgalerien, etc. pro Tag

Kuratiert von weltweit 800 Redakteurinnen und Redakteuren



Benutzerinnen und Benutzer in 140 Ländern

Redaktionelle Auswahl für 55 Sites

In 28 Sprachen



~500M (MAU) Monatlich aktive User

Seit 2015 \$750M in Werbeeinnahmen an Medienpartner
ausbezahlt

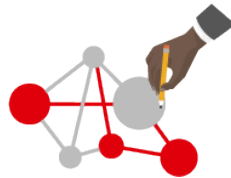
\$200M allein in 2018

Microsoft News Businessmodell



1. Lizenzierte Inhalte

- Artikel
- Videos
- Bilder
- Strukturierte Daten (z.B. Wetter, Sport)



2. Kuratieren der Inhalte

- KI gesteuerte Auswertung und Kategorisierung von Inhalten
- Content Management
- Mischung von redaktioneller und maschineller Programmierung



3. Monetarisierung durch Werbung

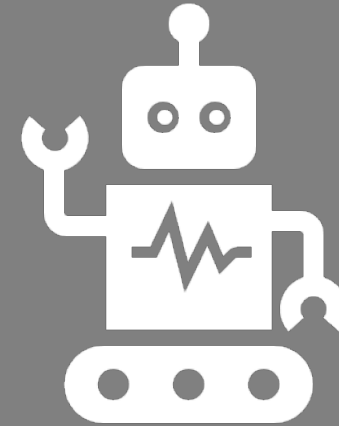
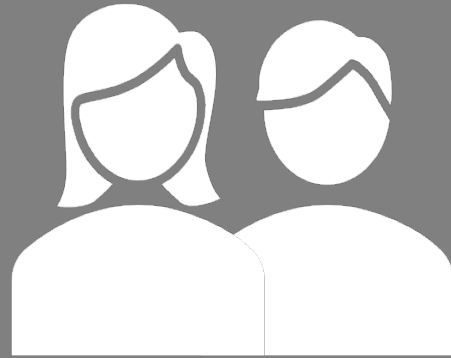
- Verschiedene Werbemittel
- Lokale Relevanz



Werbeeinnahmen fließen anteilig zurück an Medienpartner



Microsoft News Programmierungs-Strategie

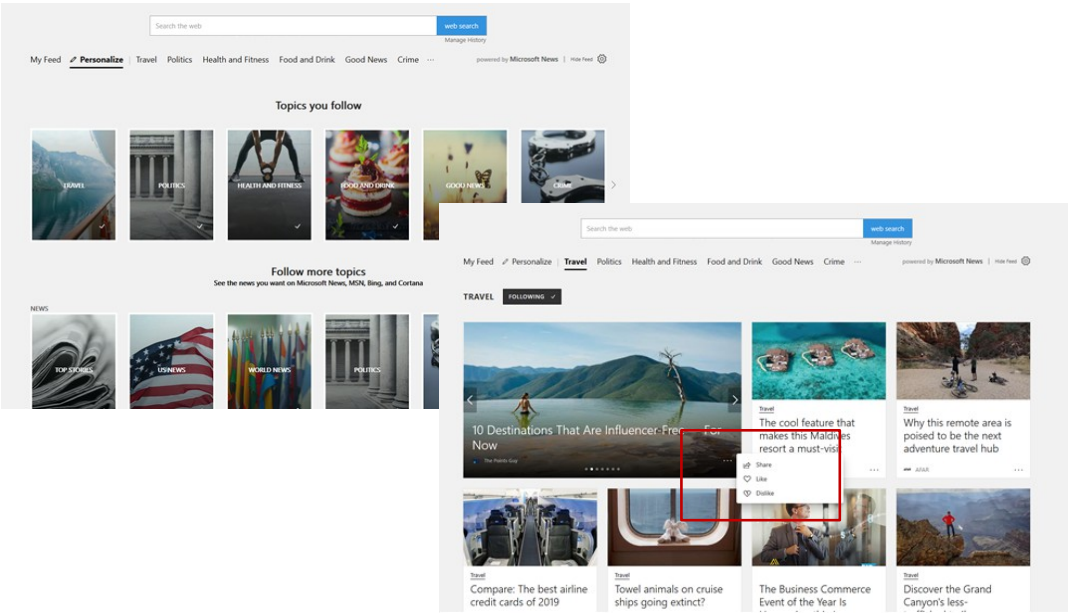


- Schlagzeilen, Eilmeldungen
- Balance und Spektrum von Medienbrands
- Zeitgebundene Berichterstattung
- Themenorientierte Programmierung
- Qualitätskontrolle

- Personalisierte, Interessen-basierte Programmierung
- Basiert auf
 - Signalen von Nutzerinnen und Nutzern
 - Text- u. Bildanalyse und -kategorisierung
 - Machine Learning

Signale - Explizit und Implizit

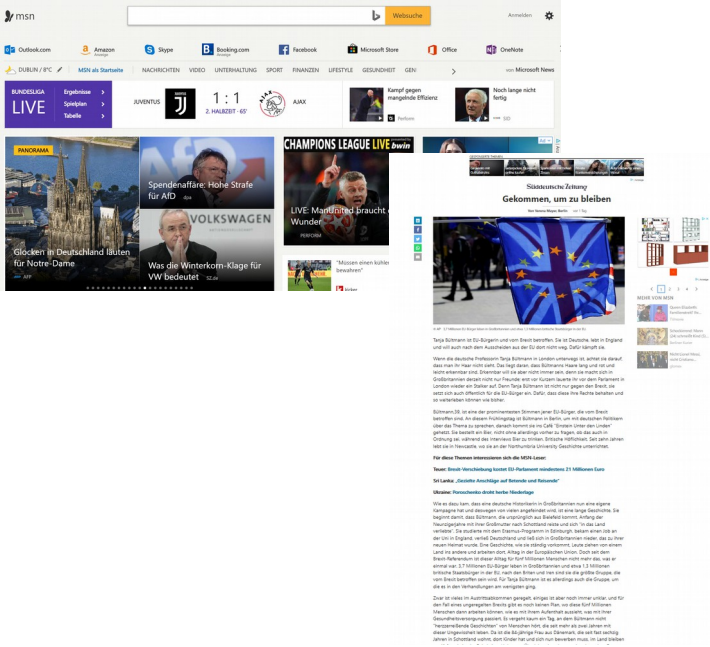
1



Topics & Interests
Like, Dislike, Sharing

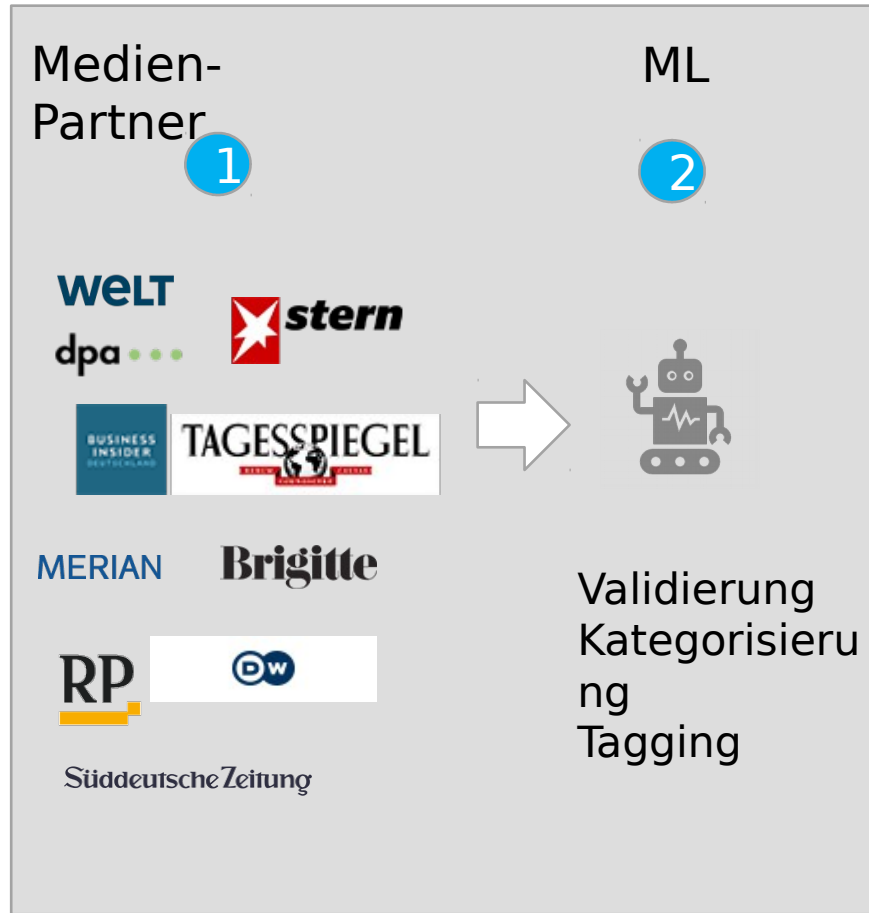
Microsoft News

2

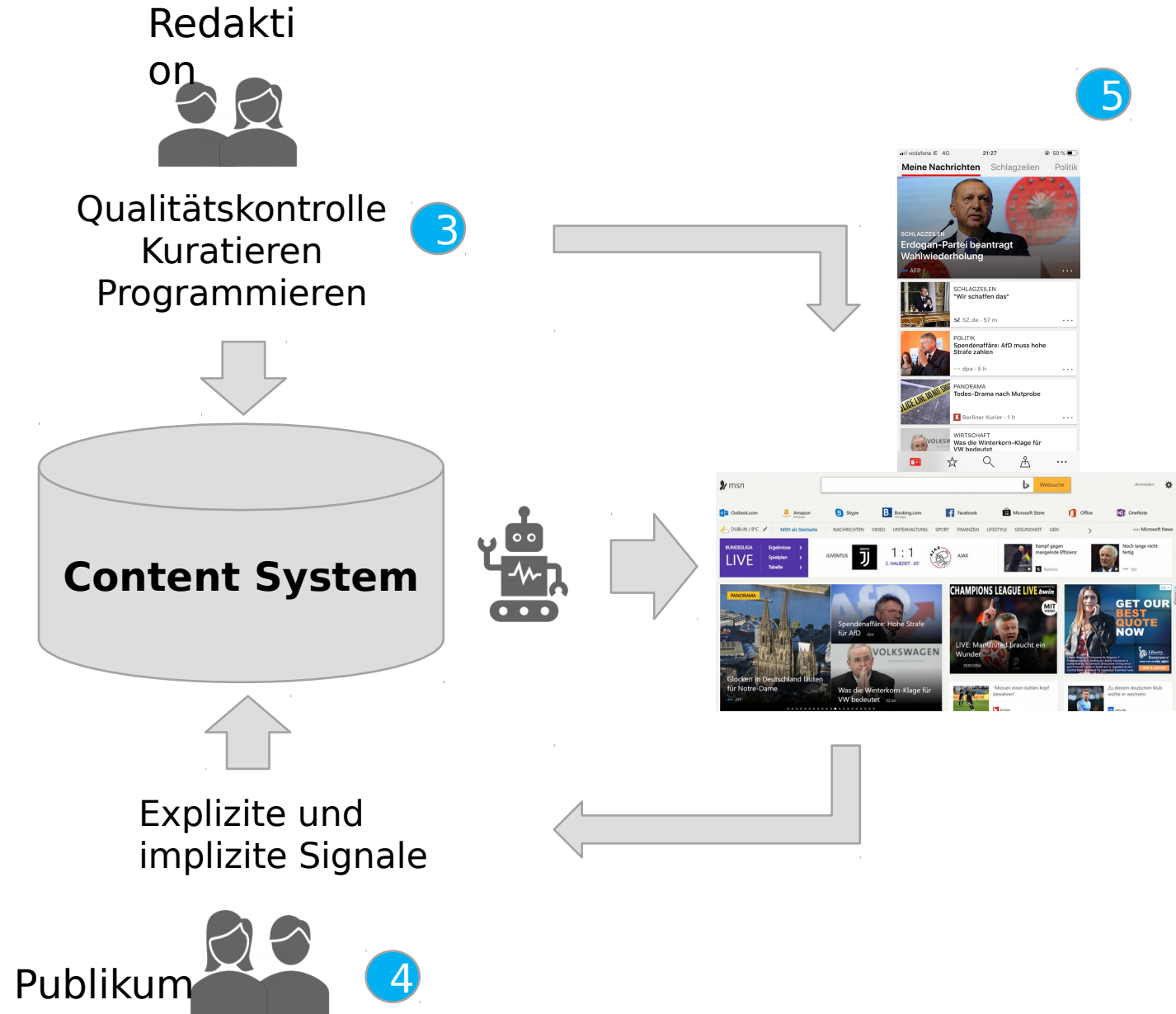


Clickverhalten (Kategorien, Themen)
Scrolling
Länge der Artikelansicht
Präferenzen des Inhaltstyps (Video, Artikel, Galerie)

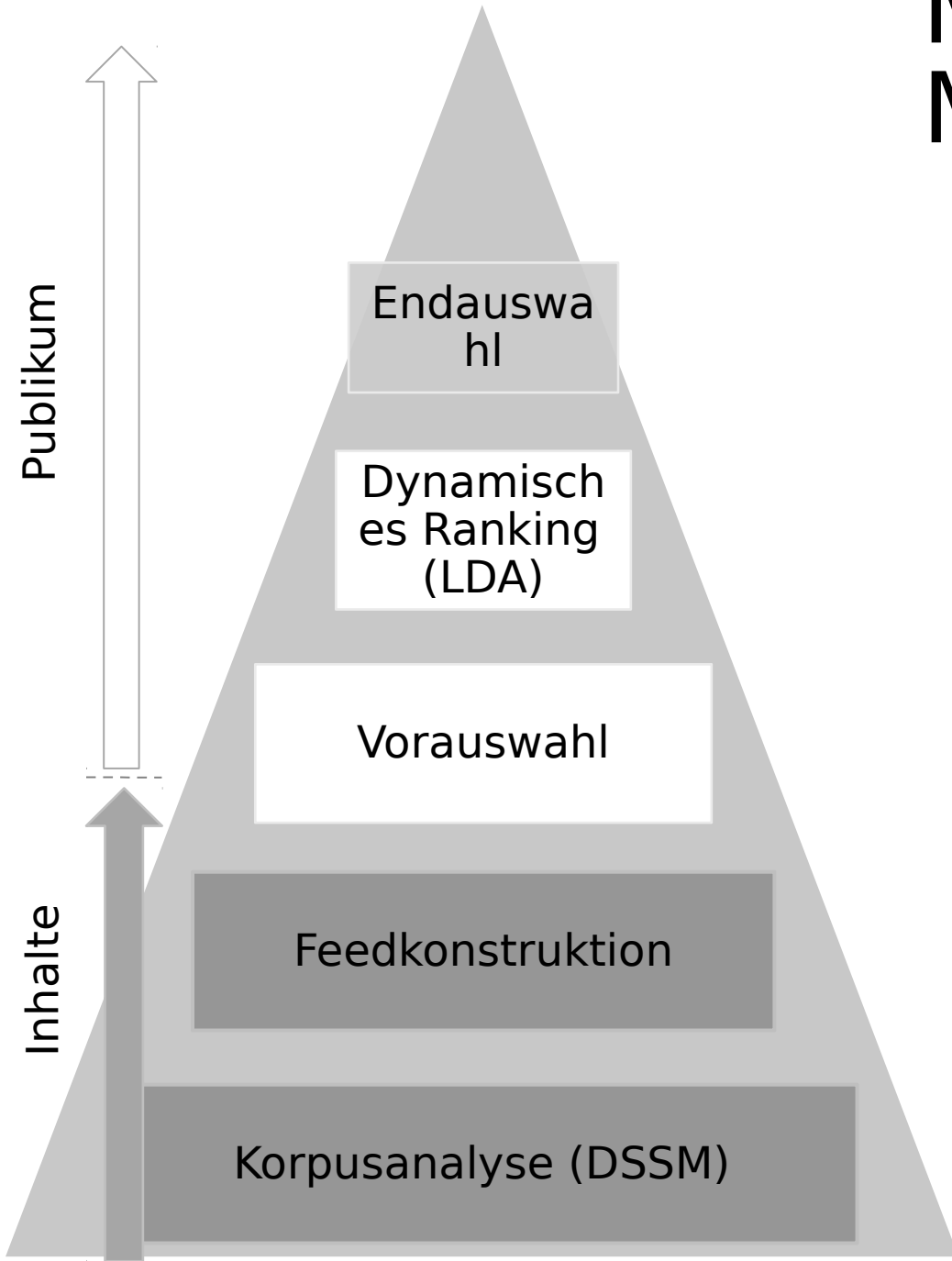
Menschliches und algorithmisches Programmieren von Inhalten



Microsoft News



Microsoft News und Machine Learning



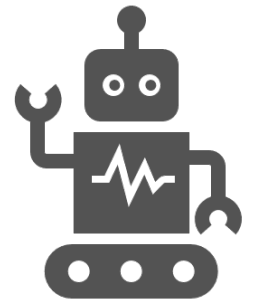
➔ Redaktionelle + personalisierte + experimentelle Auswahl

➔ Sortierte Auswahl

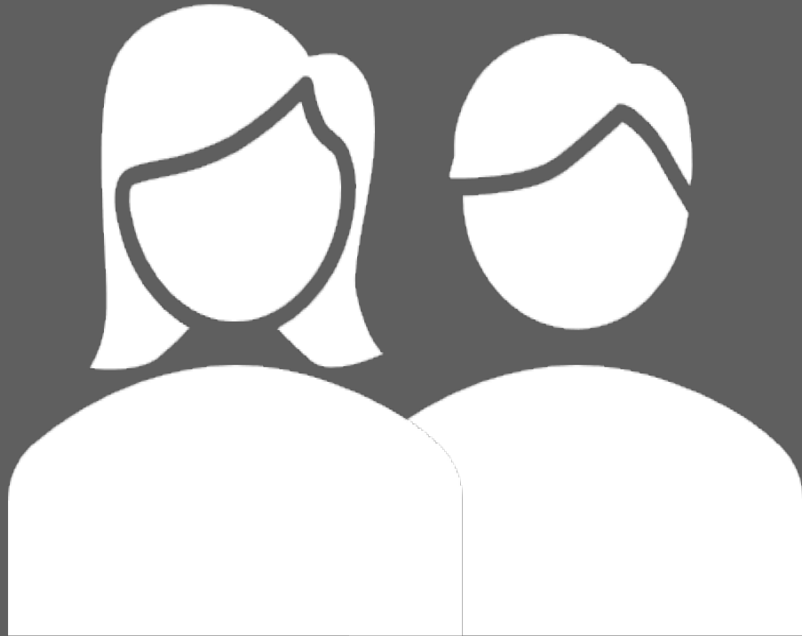
➔ Formierung eines personalisierten Korpus für einzelne Anwenderinnen und Anwender

➔ Redaktionelle Auswahl, Autokuration

➔ Annotation, Themenextrahierung, Locationtagging



Gender and Fairness in Medien und Machine Learning



Doppeltes Risiko

- Bias in Medien
- Bias-Verstärkung durch ML

Gender in Medien

2,353,652 articles

2,171,239 references to named persons

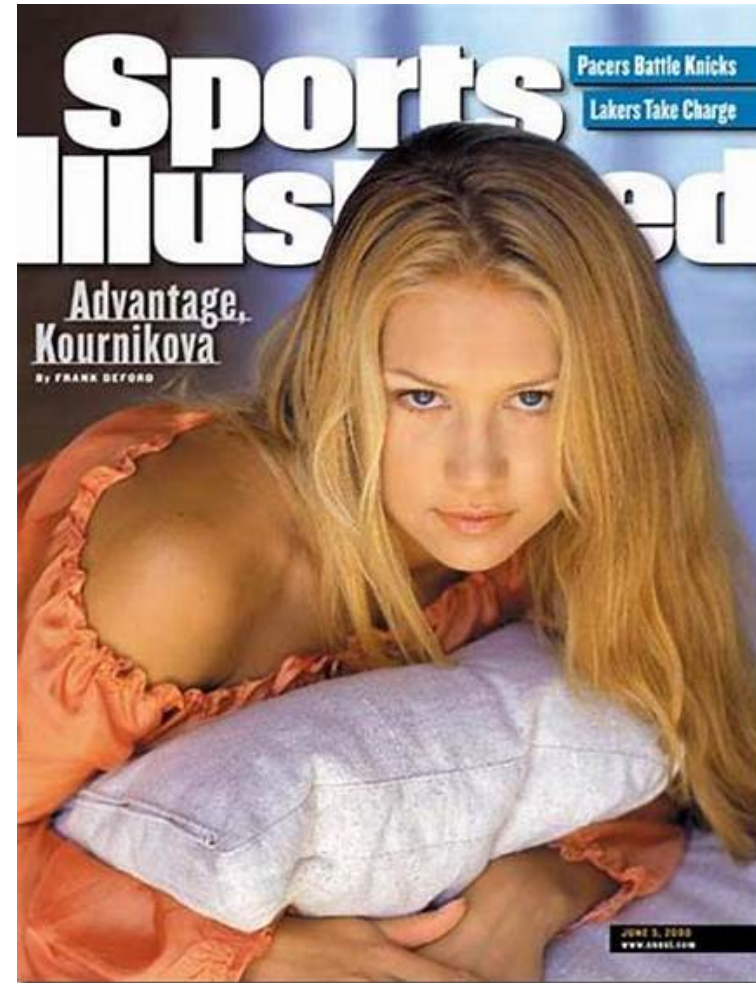
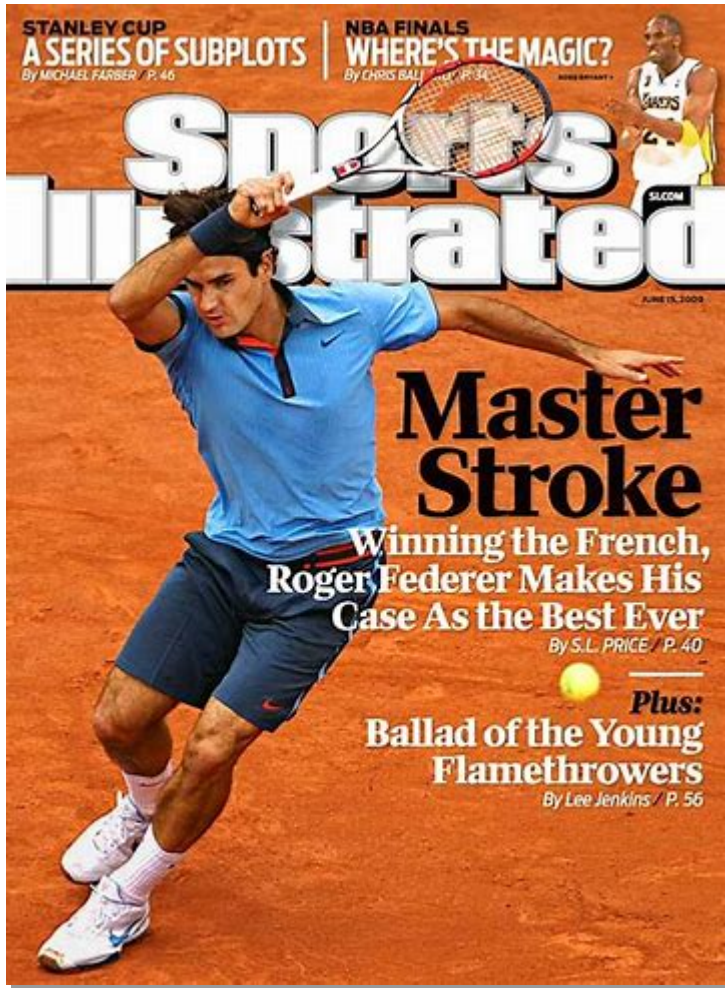
1,376,824 images resolving the gender of names and faces

“the news media are still very much male-dominated, with an overall probability of 77.0% that an entity mentioned in the text is male, or 69.6% that a face image is male”

“Fashion was found to be the only topic category where mentions of females in the text, or images of female faces, were more likely than those of their male counterparts, with the probability of an entity being male in Fashion equalling 45.9%, while the probability of a face image

Jia S, Lansdall-Welfare T, Sudhakar S, Carter C, Cristianini N (2016) **Women Are Seen More than Heard in Online Newspapers.** PLoS ONE 11(2): e0148434. doi:10.1371/journal.pone.0148434

Darstellung



GESPONSERTE THEMEN

Wo kann man Geld sicher anlegen

Sparkonto mit 5% Zinsen

Die 10 besten Krankenkassen

Die besten Girokonten

Kreuzfahrtschiffe ab Hamburg

Anzeige

USA

□ 1.5

□ 6.8

UK

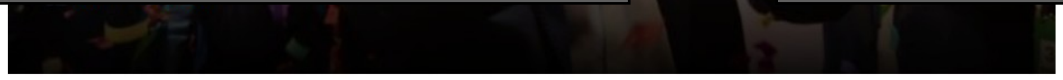
□ 1.4


□ 3.5


DEU


□ 1.1


□ 3.1




 In diesem Wald verschwinden Menschen

 Welt der Wunder


 Sri Lanka: Neue Bombenexplosion in der Nähe einer Kirche


 Unfassbare Begründung: Mann schmeißt Kind aus drittem Stock


 Berliner Kurier


 Diese Wespenarten sind bei uns heimisch

 Das Haus

 Beziehungen zur Ukraine

 Handelsblatt

 Neue Opferzahlen in Sri Lanka

 Euronews

 Der europäische Weg der Digitalisierung

 t3n Magazin

 der Ukraine

 SZ SZ.de

 Ranger posieren mit Gorillas im Nationalpark

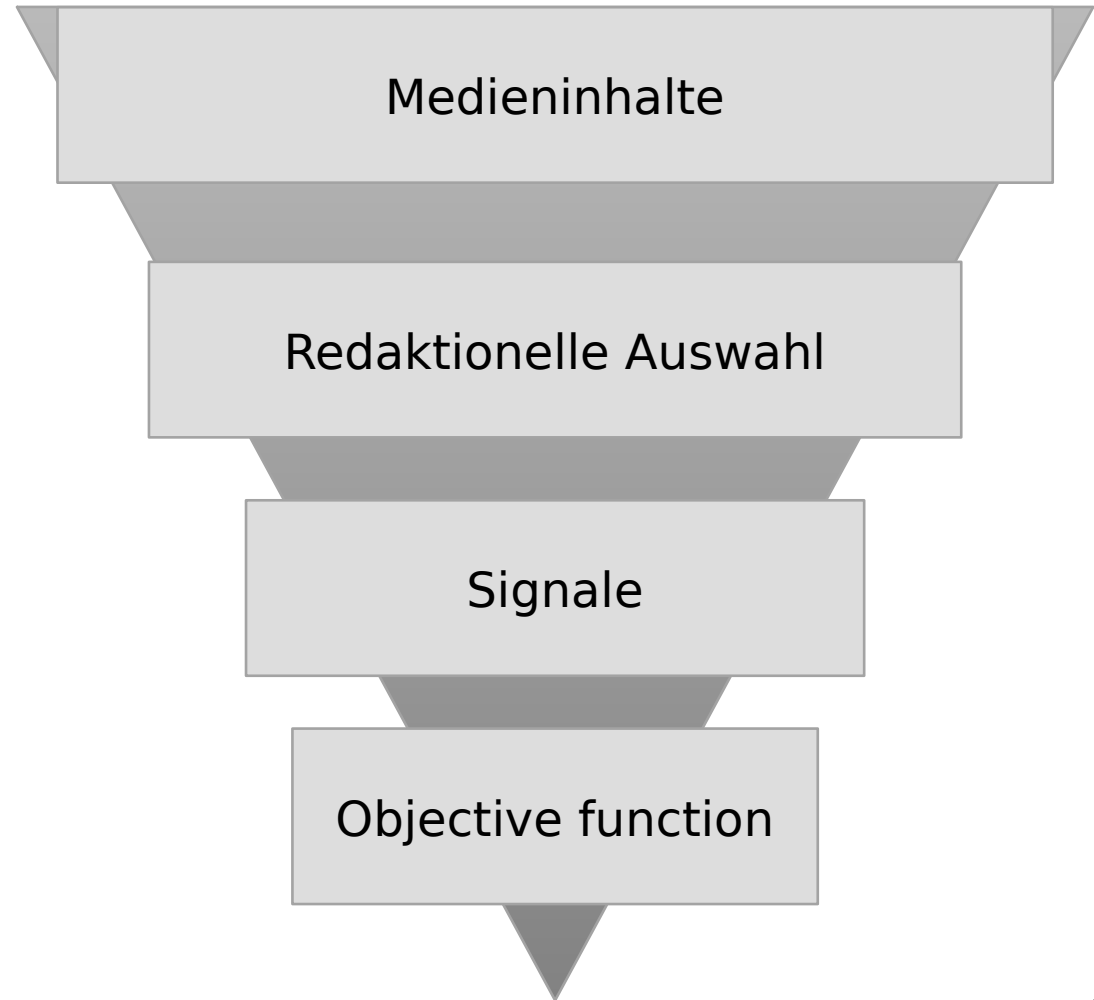
 stern

 Verdacht auf Korruption und Geldwäsche

 WELT

Gender Bias und Fairness

- Medieninhalte enthalten de facto Bias
- Redakteure sind nicht frei von Bias
- Signale basieren auf der demographischen Mischung unseres Publikums (tendenziell männlich, tendenziell älter)
- Auf welche messbaren Daten optimieren wir unser ML-Modell?



Types of Harm

	Allocation	Quality of Service	Stereotyping	Denigration	Over- or Under-Repr.
Hiring system does not rank women as highly as men for technical jobs	x		x		x
Gender classification software misclassifies darker-skin women		x			
Machine translation system exhibits male/female gender stereotypes			x	x	
Photo management program labels image of black people as "gorillas"		x		x	
Image searches for "CEO" yield only photos of white men on first page			x		x

1. Verteilung
2. Servicequalität
3. Stereotypisierung
4. Herabwürdigung
5. Über- oder Unterrepräsentation

<https://note.microsoft.com/MSR-Webinar-Machine-Learning-and-Fairness-Registration-LIVE/>

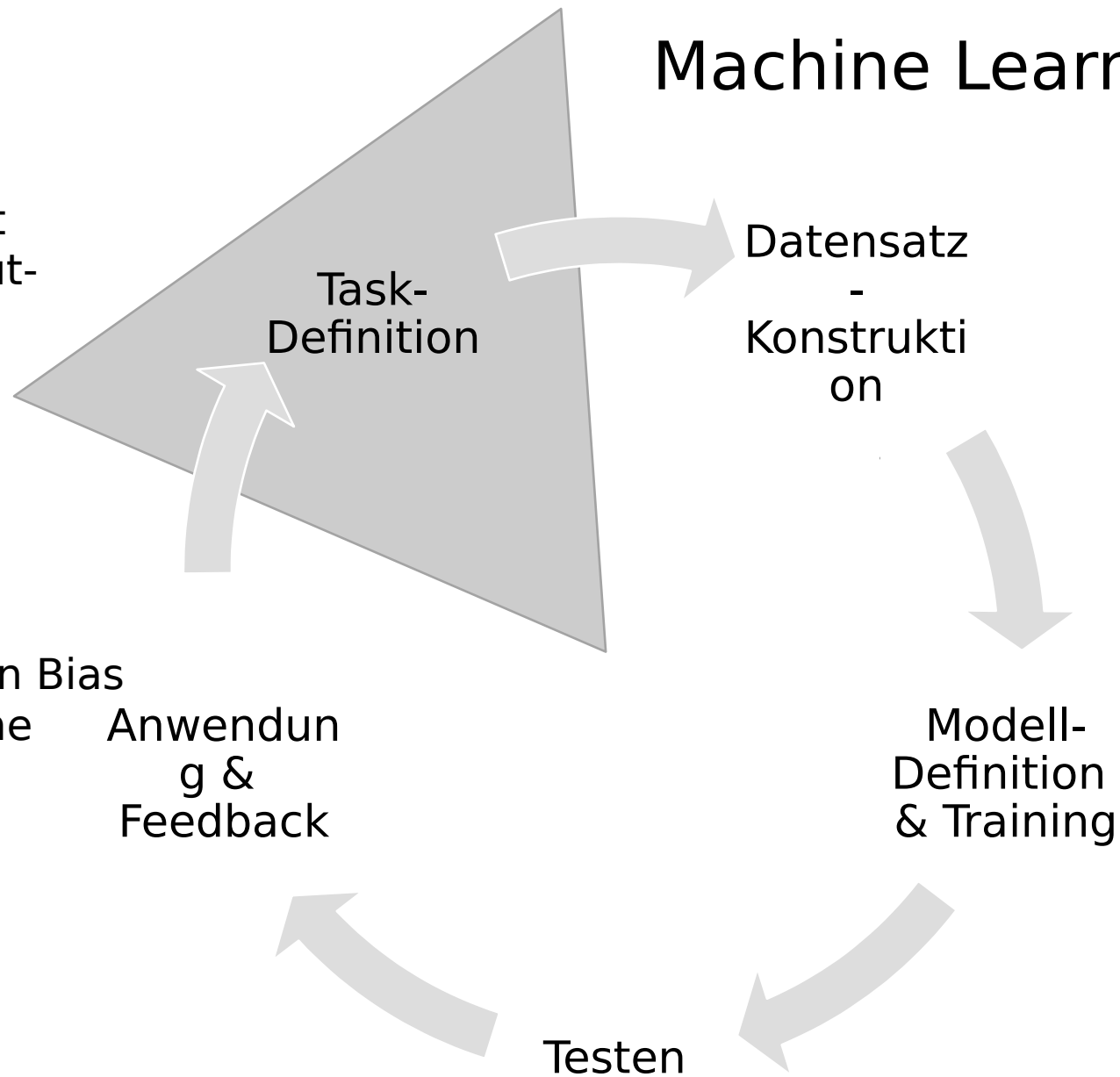
Microsoft AETHER Committee (AI and Ethics in Engineering and Research)
Hanna Wallach, Jennifer Wortman Vaughan - Types of Harm - Microsoft Research Talk

Machine Learning-Prozess

Komplexe Tasks mit verschiedenen Input-Quellen
(Analyse, Auswahl, Ranking)

Risiken:

- Filterblasen
- Verstärkung des demographischen Bias
- Qualitätsabnahme



Machine Learning-Prozess

Komplexer Datensatz

Risiken:

- Wort- und Bildererkennung
- BenutzerInnen-Signale
- Business-Logik
- Redaktionelles Input

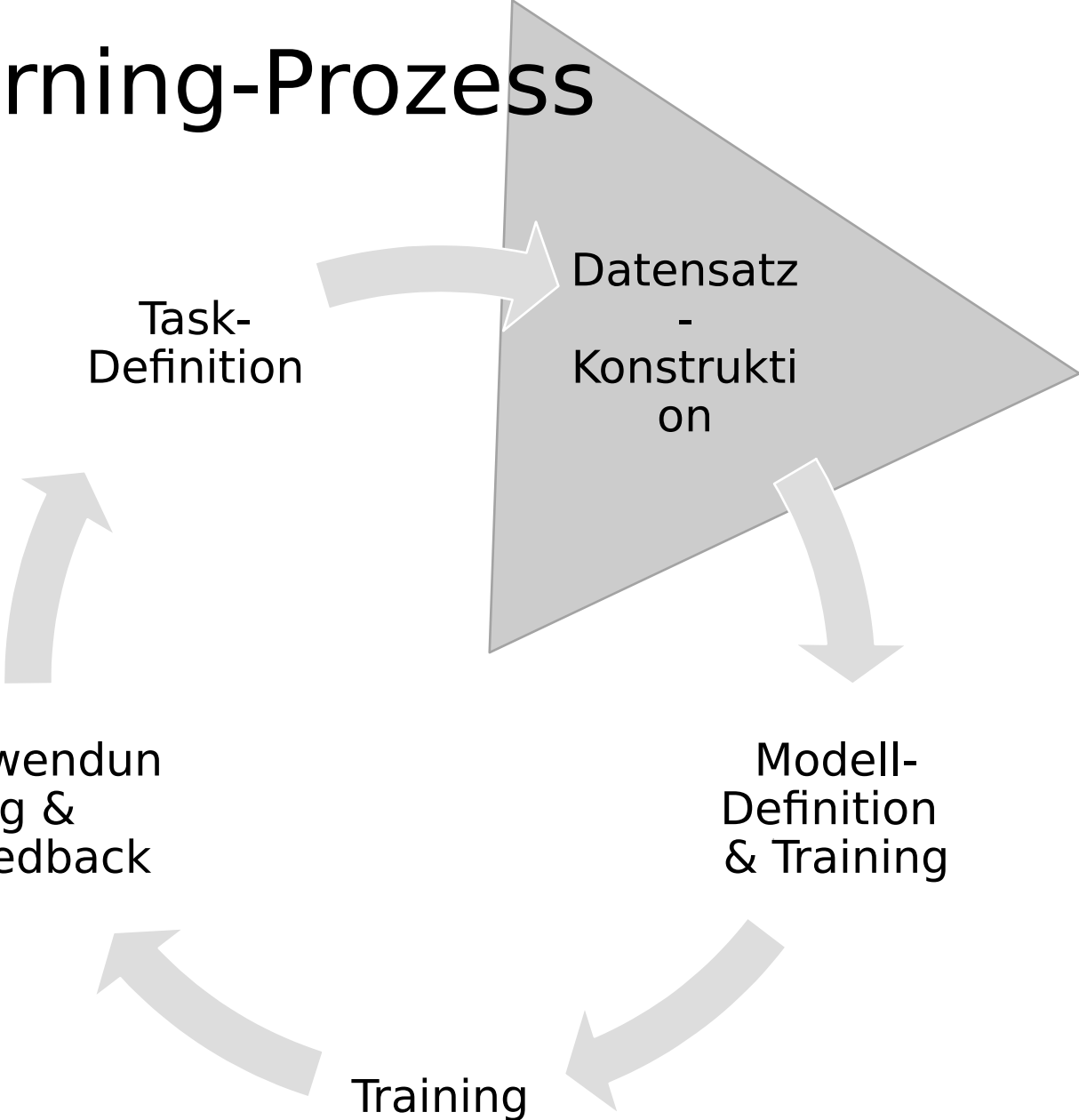
Anwendung & Feedback

Task-Definition

Datensatz - Konstruktion

Modell-Definition & Training

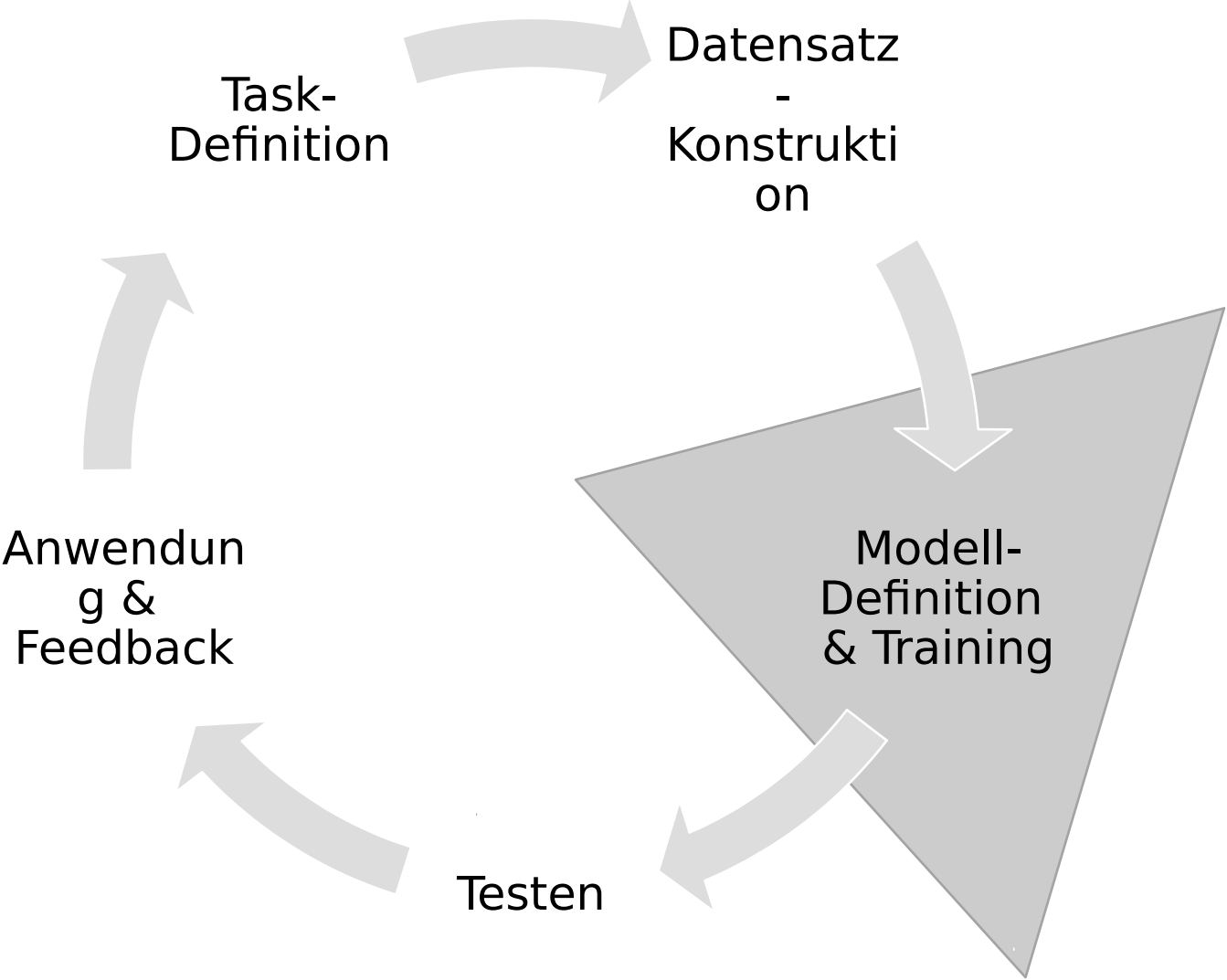
Training



Machine Learning-Prozess

Mehrere Modelle und Algorithmen

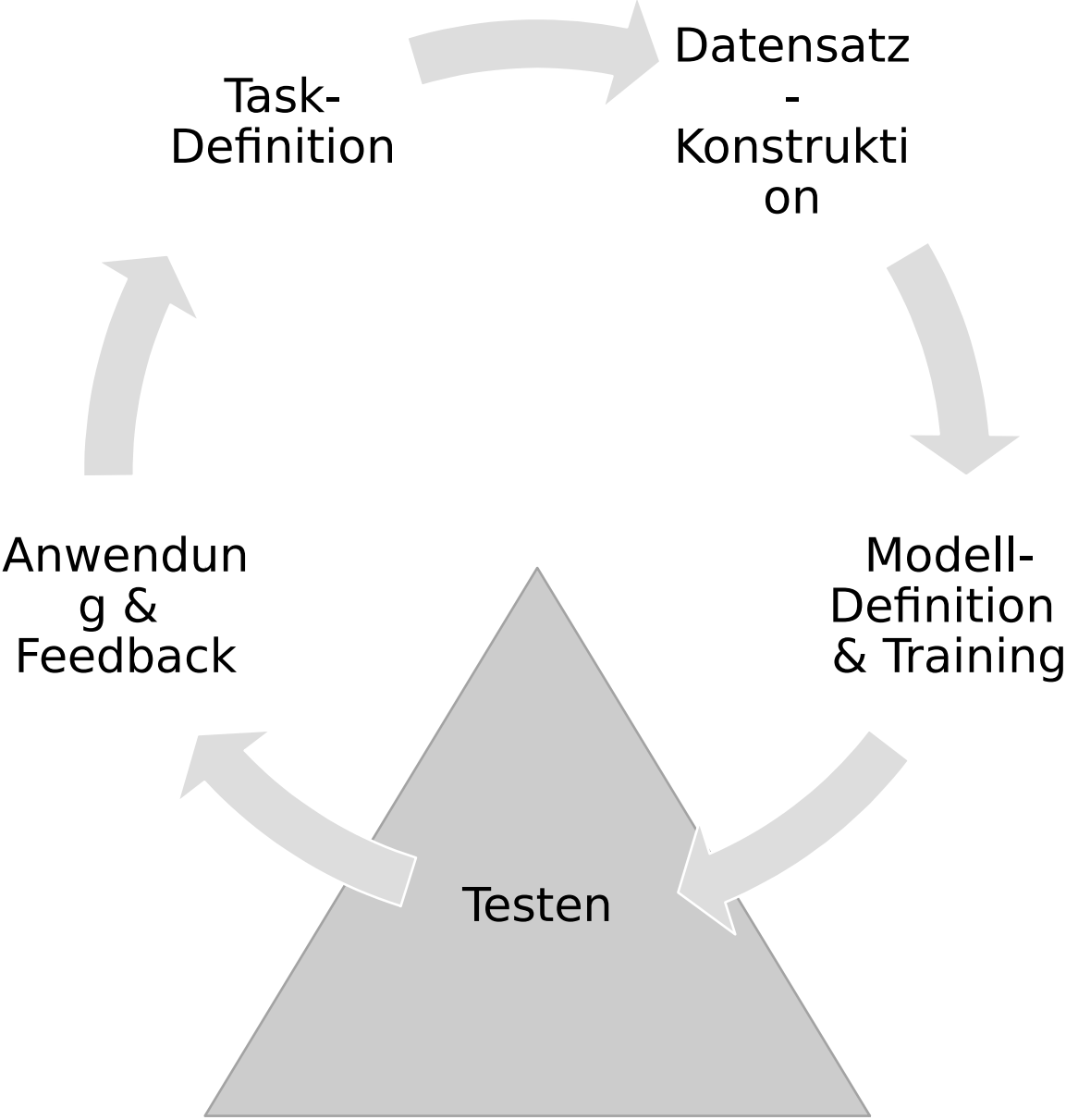
Schwierige Definition der objective function
- auf welche messbare Größe optimieren wir?



Machine Learning-Prozess

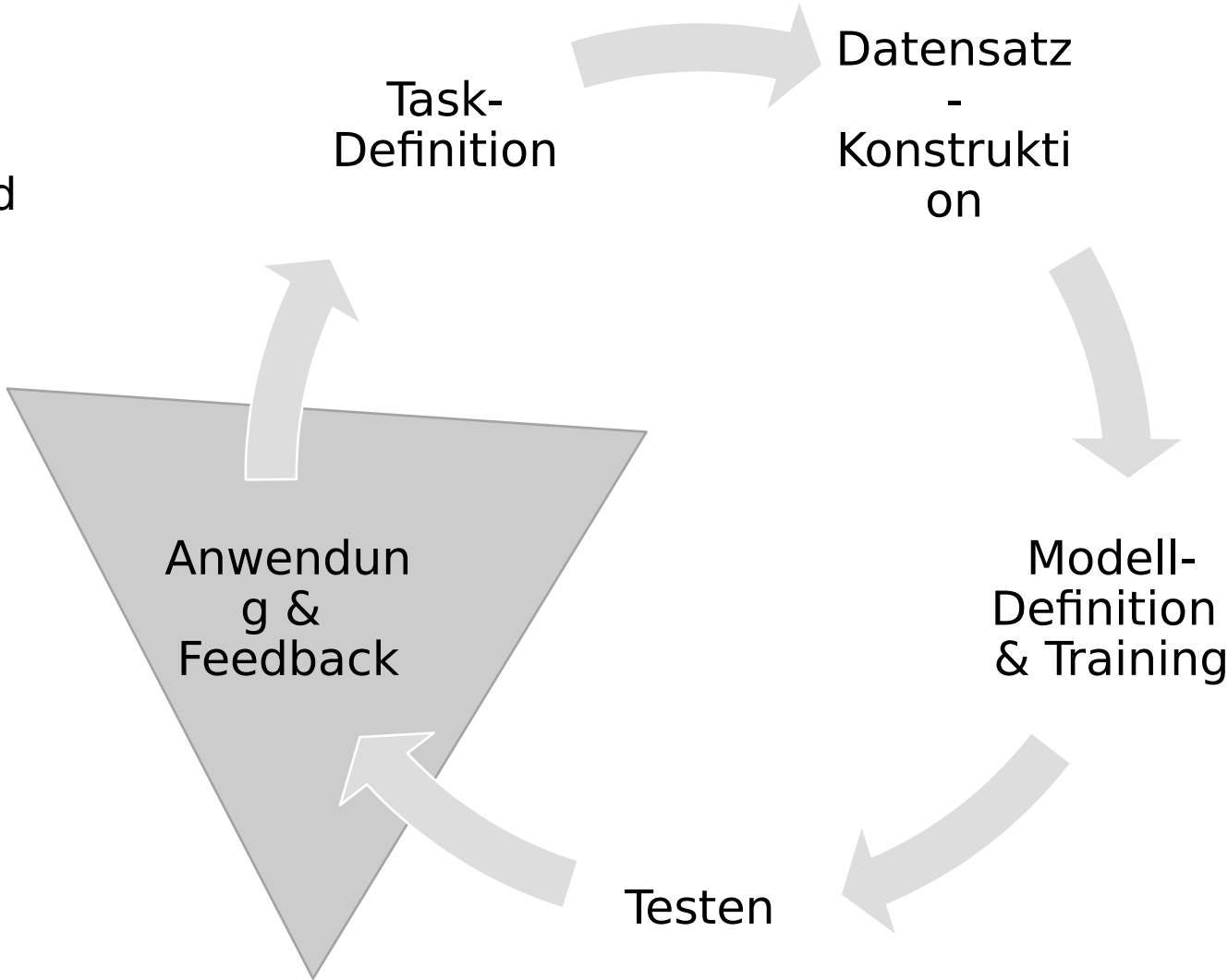
Komplexe Testfälle:
- Profil-basiert
- Datenanalyse

Redaktionelle
Qualitätskontrolle



Machine Learning-Prozess

Kontinuierliche
Beobachtung, und
evtl redaktionelle
Intervention



Analyse potentieller Schäden und Lösungsansätze

Schaden	Risiken	Ansatz
Verteilung	Verteilung der Autorenschaft	Analyse der Autorenschaft von Artikeln und Bildern Zusammenarbeit mit Medien-Partnern Redaktionelle Tools
Servicequalität	Bestimmte Gruppen fühlen sich weniger angesprochen	Analyse der Publikumszufriedenheit nach Segmenten Analyse der Qualitätskontrolldaten
Stereotypisierung	Stereotype Darstellungen	Richtlinien für Redakteure und Werbung Intervention für Kategorisierung und Wort-bzw. Bild-Assoziationen
Herabwürdigung	Herabwürdigende Darstellung	Richtlinien für Redakteure und Werbung Algorithmische Analyse Redaktionelle Stichproben
Über- oder Unterrepräsentation	Ungleichgewichtige Darstellung und Berichterstattung	Algorithmische Analyse der programmierten Sites (Bildererkennung, Artikel) Redaktionelle Tools

Zusammenfassung

- Bias ist in Medienprodukten unvermeidbar, aber wir können ihn reduzieren
- Ein Lösungsansatz muss sowohl auf der menschlichen als auch der ML-Seite stattfinden