

# SATW & APROPOS\_Stiftung Risiko-Dialog

*u*<sup>b</sup>

## Künstliche Intelligenz in unserem Alltag

b  
UNIVERSITÄT  
BERN

**Dr. Matthias Stürmer, Forschungsstelle Digitale Nachhaltigkeit**

21. November 2019



# Matthias Stürmer

- Seit 2013 Leiter der **Forschungsstelle Digitale Nachhaltigkeit** am Institut für Informatik der **Universität Bern**, dazu Dozentur für Digitale Transformation am Institut für Informatik und Dozentur für Digitale Nachhaltigkeit am Institut für Wirtschaftsinformatik
- 2010 bis 2013 bei **EY (Ernst & Young)** als Senior Consultant bzw. Manager im IT Advisory: Beratung zu Open Source Software, Open Government Data und Social Media
- 2009 bis 2010 Business Development und Projektleiter bei **Liip AG**
- 2006 bis 2009 Assistent an der **ETH Zürich** bei Prof. Georg von Krogh am Lehrstuhl für Strategisches Management und Innovation (D-MTEC)
- 2000 bis 2005 Studium Betriebswirtschaft und Informatik an der **Universität Bern**, Lizenziatsarbeit zu Open Source Community Building
- Seit 2017 Mitglied Arbeitsgruppe Smart Capital Region der **Hauptstadtregion Schweiz**
- Seit 2016 Präsident Verein **Digital Impact Network**
- Seit 2012 Vorstandsmitglied und Mitgründer Verein **Opendata.ch**
- Seit 2009 Geschäftsleiter **Parlamentarische Gruppe Digitale Nachhaltigkeit**
- Seit 2006 Vorstandsmitglied und heute Vizepräsident Verein **CH Open**
- 2011 bis 2019 **Stadtrat von Bern**



**Dr. Matthias Stürmer**

Leiter Forschungsstelle  
Digitale Nachhaltigkeit

Universität Bern  
Institut für Informatik  
Schützenmattstrasse 14  
CH-3012 Bern

Telefon direkt: +41 31 631 38 09

Mobile: +41 76 368 81 65

Tel: +41 31 631 38 79 (Sekretariat)

Twitter: @maemst

matthias.stuermer@inf.unibe.ch

[www.digitale-nachhaltigkeit.unibe.ch](http://www.digitale-nachhaltigkeit.unibe.ch)

# Forschungsstelle Digitale Nachhaltigkeit



# Forschungsstelle Digitale Nachhaltigkeit

- Gegründet 2014 am Institut für Wirtschaftsinformatik (IWI) der Universität Bern, heute am **Institut für Informatik (INF)**, Team von über 20 Mitarbeitenden
- **Forschung** (u.a. laufendes SNF-Projekt im NFP 73) zu digitaler Nachhaltigkeit, Open Source Software, Blockchain, Open Data und Linked Data, Open Government und Smart City, künstlicher Intelligenz sowie öffentliche Beschaffungen
- **Vorlesungen** zu Open Data und Requirements Engineering am IWI, digitale Nachhaltigkeit am INF, Betreuung von über 100 abgeschlossener Bachelor- und Masterarbeiten, Dozenturen am IWI und am INF, Dozent in mehreren CAS und MAS
- **Weiterbildungen** zu IT-Beschaffung und nachhaltiger Beschaffung, ab 2020 CAS «Öffentliche Beschaffungen» und CAS «Technologie-Beschaffungen»
- **Dienstleistungen** (Beratung, Gutachten, Studien, Prototyping) für Hochschulen, Bundesverwaltung, Kantone, Städte, Unternehmen und Verbände

# Künstliche Intelligenz

- **Technologische** Fragestellungen
- **Wirtschaftliche** Fragenstellungen
- **Rechtliche** Fragestellungen
- **Gesellschaftliche** Fragestellungen



# Begrifflichkeiten

## Künstliche Intelligenz (KI) / Artificial Intelligence (AI)

- Entwicklung von Ideen und Visionen
- Schaffung der konzeptionellen und mathematischen Grundlagen
- Aber zu wenig Rechenleistung und Daten  
→ «KI-Winter»

## Maschinelles Lernen (ML) / Machine Learning

- Realisierung von ersten KI-Anwendungen
- Basiert auf Mustererkennung
- Überwachtes (supervised) Lernen der Algorithmen durch Experten-Wissen

## Deep Learning

- Selbstlernende Programme basierend auf neuronalen Netzen
- Simulation des menschlichen Gehirns
- Grosse Datenmengen (Big Data) und grosse Rechenleistung (Server) nötig

1950

1980

2010

# Anwendungsgebiete von KI

- **Prognosen:** Erkennen von Betrug
- **Expertensysteme:** Empfehlungen im Gesundheitsbereich
- **Texterkennung:** Optical Character Recognition (OCR) beim Scannen von Dokumenten
- **Textverständnis:** Natural Language Processing (NLP) für Übersetzungen, Chatbots
- **Bilderkennung:** Facial Recognition, Objekterkennung, Bildoptimierungen, autonomes Fahren
- **Tonanalyse:** Spracherkennung, Musikanalyse
- etc.

# Bisher ethische Fragen eher hypothetisch



Start

Beurteilen

Klassik

Designen

Durchsuchen

Über

Feedback

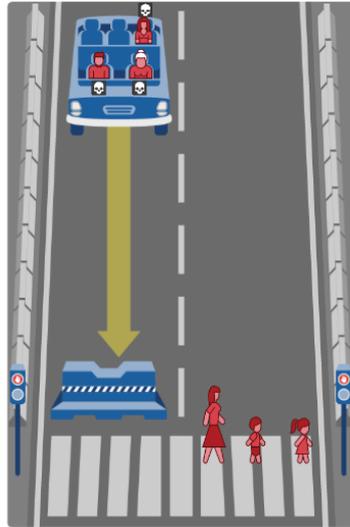


## Was soll das selbstfahrende Auto machen?

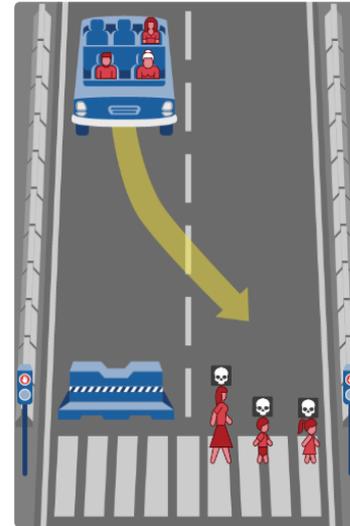
1 / 13

Das selbstfahrende Auto mit plötzlichem Bremsversagen wird in diesem Fall geradeaus weiterfahren und in eine Betonbarriere prallen. Das führt zu Tot:

- 1 Ältere Frau
- 1 Mann
- 1 Frau



Beschreibung ausblenden



Beschreibung ausblenden

Das selbstfahrende Auto mit plötzlichem Bremsversagen wird in diesem Fall ausweichen und über einen Zebrastreifen auf der gegenüberliegenden Spur fahren. Das führt zu

Tot:

- 1 Frau
- 1 Junge
- 1 Mädchen

Beachte, dass die betroffenen Fußgänger die Straße unrechtmäßig bei rot überqueren

# KI-Beispiel Gesichtserkennung

## 1. Gesichter finden

HOG version of our image

HOG face pattern generated from lots of face images

Face pattern is pretty similar to this region of our image—we found a face!

The diagram illustrates the process of finding faces using Histogram of Oriented Gradients (HOG). On the left, a grid of small arrows represents the HOG face pattern generated from many face images. This pattern is compared against a 'HOG version of our image' (a grayscale image of a face). A yellow box highlights a region where the face pattern is similar to the image's HOG pattern, indicating a face has been found. This process is shown for two different images: one of a man in a blue cap and another of a man in a suit.

## 2. Gesicht erkennen (68 face landmarks)

The diagram shows a face image with 68 numbered landmarks (dots) placed at key facial features such as the eyes, nose, mouth, and jawline. A yellow box highlights the face region. This step is used for facial recognition and expression analysis.

## 3. Gesicht kodieren (128 measurements)

Input Image

128 Measurements Generated from Image

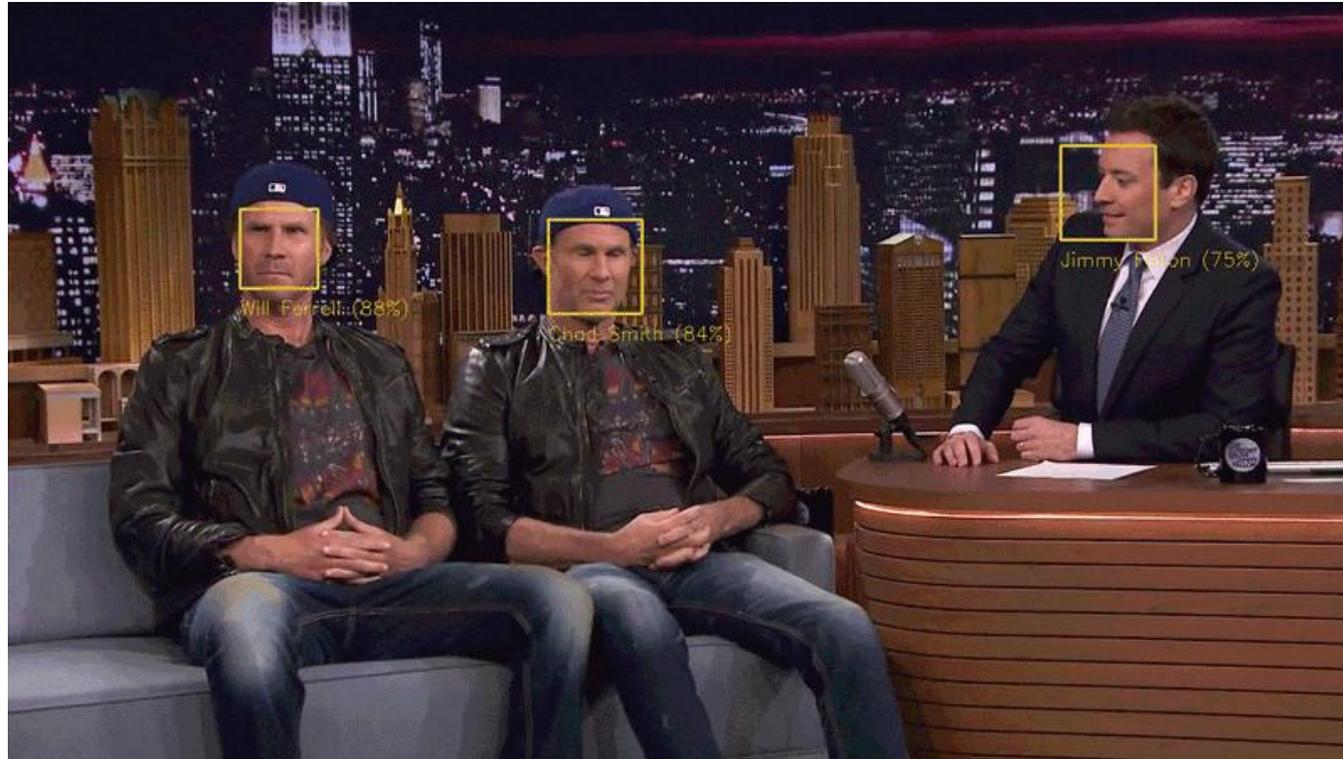
The diagram shows an input face image being processed into a set of 128 numerical measurements. These measurements are listed in a grid below the image. The measurements represent various facial features and are used for face coding.

0.097486204899098	0.26322306383984	-0.12816887678293	0.83200441886424
0.1232628741129	0.230209119127216	0.1721621171962	0.02097082819627
0.23060438718725	0.011981477231189	0.1081380568365	-0.005216327841189
0.208552908064631	0.265552908064631	0.0731306011544	-0.1188511010111
-0.091486843481871	0.1226282872932	-0.2026282872932	0.005957919268889
-0.208491711660984	0.230209119127216	0.1826282872932	0.242374212343289
-0.14131525158882	0.14114824748518	-0.001381584841149	-0.033346127700711
-0.0485490209289	0.151102616870287	-0.151102616870287	0.07811816508817
-0.12687176924778	-0.10588545213066	-0.12728654841171	-0.07628616251173
0.061487712527474	-0.17428152471171	-0.26536522527209	0.15268461713058
-0.0467416971574	0.06178185124811	0.1474654276508	0.06418422309588
-0.1213060143147	-0.21109199174051	0.041109122790382	-0.08972747920258
0.161887481608495	0.113452870131979	0.02130201651852	0.0064904208684223
0.081898940702915	0.1837202846114	-0.06728623243152	-0.023286187481432
0.1084110504782	0.0048048200741915	0.0048048200741915	0.0048048200741915
-0.104146227412783	0.0048048200741915	0.21180312235491	-0.0058238210049
0.103484916231687	0.0048048200741915	0.0048048200741915	-0.0048048200741915
-0.12116668761002	-0.007277733588489	-0.008801291458799	-0.0343627773279
0.08204545121813	-0.007277733588489	-0.0702864420814	-0.04301765570587
-0.087945111095905	-0.1147843287904	-0.08962143173013	-0.11385110780088
-0.021407811989238	-0.1484119890971	0.0782217817748	-0.1198889712387
-0.01208880441656	0.049325442830866	0.132278387468	0.07200327432166
-0.011413153868917	-0.051018297181381	-0.051018297181381	0.0005011638276258
0.000876334888338	-0.00281278720878	-0.1542746848889	-0.0148206833883
0.08139257120027	0.0048048200741915	0.0048048200741915	-0.04378648912023
0.0258210274802051	-0.1144570280325	0.07189795441475	0.01206206480002
-0.05729334680293	0.01488388967261	0.05238154733777	0.0127744465407939
0.02828031081488	-0.0017082804888	0.08883388967261	0.08883388967261
-0.008097913188324	0.037020358488952	0.1192047814589	0.11898788878918
0.02020554188409	-0.0178813183078	0.1882388967261	0.0182388967261
0.004932370829002	-0.094388014247117	-0.1196248071254	-0.10281407731889
0.05198708226261	-0.100483162777	-0.0409725826216	-0.082041338696178

## 4. Gesichter trainieren

The diagram shows a grid of face images used for training. The images are labeled with file names such as '138738862-G00Y-45690...', '42202\_pro.png', '50127.3088.png', 'chad-smith.png', 'chad-smith\_5214color.png', 'chad.png', '28-jimmy-fallon-w529A529.png', '27-jimmy-fallon-finger.w12...', '66627-14323921698-jimmy...', 'artic-019A4H4B00005...', 'CDM\_612991.png', 'maxrscdefaut.png', 'chad\_smith\_2010\_05\_06.png', 'chadsmith45.png', 'drummerch\_david\_63047...', 'fallon101115\_2\_250.png', 'jimmyfallon42-600x800.png', and 'jimmy-fallon-1024.png'.

# KI-Beispiel Gesichtserkennung



# Grosse Mengen an Testdaten

## Labeled Faces in the Wild



### Menu

- LFW Home
  - Explore
  - Download
  - Train/Test
  - Results
  - Information
  - Errata
  - Reference
  - Resources
  - Contact
  - Support
  - Changes
- Part Labels
- UMass Vision

### Labeled Faces in the Wild Home



#### NEW RESULTS PAGE:

WE HAVE RECENTLY UPDATED AND CHANGED THE FORMAT AND CONTENT OF OUR [RESULTS PAGE](#). PLEASE REFER TO THE [NEW TECHNICAL REPORT](#) FOR DETAILS OF THE CHANGES.

#### DISCLAIMER:

Labeled Faces in the Wild is a public benchmark for face verification, also known as pair matching. No matter what the performance of an algorithm on LFW, it should not be used to conclude that an algorithm is suitable for any commercial purpose. There are many reasons for this. Here is a non-exhaustive list:

# Gesichtserkennung in China

South China Morning Post

SIGN IN/UP



Start-ups

## Shenzhen AI start-up Intellifusion helps city police identify jaywalkers and banned drivers

- In the first 10 months after installation of the equipment, as many as 13,930 jaywalking offenders were named and shamed



Li Tao

Published: 6:00am, 4 May, 2019

### TOP PICKS

Tech

DJI is making an app for people to identify drones

14 Nov 2019



Tech

This university helped groom some of Singapore's biggest start-up founders

30 Oct 2019



Tech

Chinese taxpayers are helping to underwrite tech start-up risks

13 Nov 2019



# Gesichtserkennung

## Chancen:

- Erhöhung der Sicherheit im öffentlichen Raum
- Nutzen bei Fahndungen und bei Bekämpfung von Kinderpornografie
- Bessere Marketing-Möglichkeiten

## Risiken:

- Überwachung und Kontrolle durch den Staat
- Verletzung der Privatsphäre und des Datenschutzes
- Bei Hacking-Attacken Diebstahl von heiklen Daten

# KI-Beispiel DeepFake



MOTHERBOARD  
TECH BY VICE

## AI-Assisted Fake Porn Is Here and We're All Fucked

Someone used an algorithm to paste the face of 'Wonder Woman' star Gal Gadot onto a porn video, and the implications are terrifying.

By [Samantha Cole](#)

Dec 11 2017, 8:18pm [Share](#) [Tweet](#) [Snap](#)



IMAGE: SCREENSHOT FROM SENDVIZ

There's a video of Gal Gadot having sex with her stepbrother on the internet. But it's not really Gadot's body, and it's barely her own face. It's an approximation, face-swapped to look like she's performing in an existing

### Living portraits



# KI-Beispiel DeepNude



Digital > Online-Sicherheit > DeepNude: App erzeugt gefälschte Nacktbilder aus Fotos von Frauen



screenshot: motherboard

## Die DeepNude-App zieht Frauen aus – ohne Einwilligung der Betroffenen



© 02.07.19, 10:26 © 02.07.19, 22:54



### How it works

We have developed an advanced artificial intelligence algorithm that allows you to remove clothes from photos.



Step 1: Select a picture



Step 2: Crop the picture



Step 3: Nude it!

# Das Ende von DeepNude



Here is the brief history, and the end of DeepNude. We created this project for user's entertainment a few months ago. We thought we were selling a few sales every month in a controlled manner. Honestly, the app is not that great, it only works with particular photos. We never thought it would become viral and we would not be able to control the traffic. We greatly underestimated the request.

Despite the safety measures adopted (watermarks) if 500,000 people use it, the probability that people will misuse it is too high. We don't want to make money this way. Surely some copies of DeepNude will be shared on the web, but we don't want to be the ones who sell it. Downloading the software from other sources or sharing it by any other means would be against the terms of our website. From now on, DeepNude will not release other versions and does not grant anyone its use. Not even the licenses to activate the Premium version.

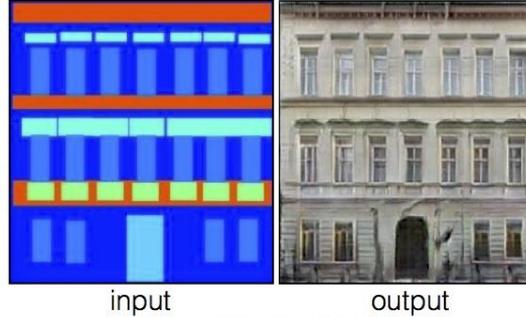
People who have not yet upgraded will receive a refund.  
The world is not yet ready for DeepNude.

# Basiert auf Open Source Bibliothek pix2pix

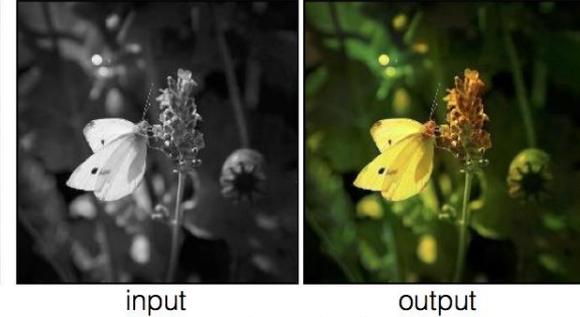
Labels to Street Scene



Labels to Facade



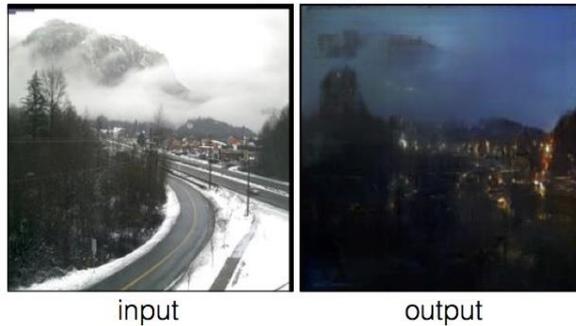
BW to Color



Aerial to Map



Day to Night



Edges to Photo



# Anwendungen von pix2pix

## #edges2cats



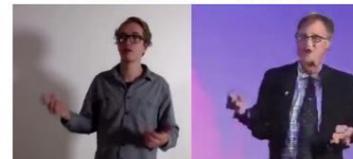
Christopher Hesse trained our model on converting edge maps to photos of cats, and included this in his [interactive demo](#). Apparently, this is what the Internet wanted most, and #edges2cats briefly [went viral](#). The above cats were designed by Vitaly Vidmirov (@vvid).

## Alternative Face



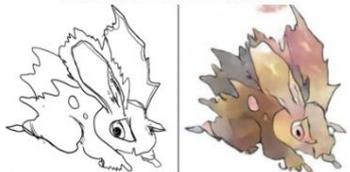
Mario Klingemann used our code to translate the appearance of French singer Francoise Hardy onto Kellyanne Conway's infamous "alternative facts" interview. Interesting articles about it can be read [here](#) and [here](#).

## Person-to-Person



Brannon Dorsey recorded himself mimicking frames from a video of Ray Kurzweil giving a talk. He then used this data to train a Dorsey→Kurzweil translator, allowing him to become a kind of puppeteer in control of Kurzweil's appearance.

## Interactive Anime



Bertrand Gondouin trained our method to translate sketches→Pokemon, resulting in an interactive drawing tool.

## Background masking



Kaihu Chen performed a [number of interesting experiments](#) using our method, including getting it to mask out the background of a portrait as shown above.

## Color palette completion



Colormind adapted our code to predict a complete 5-color palette given a subset of the palette as input. This application stretches the definition of what counts as "image-to-image translation" in an exciting way: if you can visualize your

# Bildoptimierungen bzw. Fake-Videos

## Chancen:

- Forschungserfolge
- Humorvolle Beiträge
- Verbesserung der Bildqualität
- Förderung von Produktivität, Kreativität und Innovationen

## Risiken:

- Fake News
- Betrug
- Verletzung von Persönlichkeitsrechten
- Verstärkung von Sexismus
- Erpressung, Depression, Suizid

# Ethischer Quellcode

- **KI-Tools** basieren auf frei verfügbaren Open Source Komponenten
- **Open Source Lizenzen** erlauben uneingeschränkten Einsatz der Software
- Darum «**Just World License**» bzw. «Do No Harm License» geschaffen
- **Verbot zur Verwendung der Software** für Menschenhandel, Glücksspiele, Tabakindustrie, Atomenergie, Kriegsführung, Waffenherstellung etc.



## 6 myths about “ethical” open source licenses

April 17th 2018

TWEET THIS



Last week we released the [Just World License](#) (JWL), our ethical open source license, and explained [why it's needed](#).

Whenever the idea of an ethical license is raised in the open source community, common objections come up. In thinking about these, it's worth thinking about the early days of the GPL (GNU Public License) and the

# Ethische KI-Richtlinien

- «**Asimolar AI Principles**»:  
Liste von 23 Zielen wie KI-Forschende ihr Wissen einsetzen sollen
- **Einige dieser Prinzipien:**
  - Nützliche und wohltätige KI erschaffen
  - Forschungsgelder für sinnvolle KI nutzen
  - Austausch zwischen KI-Forschenden und politischen Entscheidungsträgern
  - Transparenz bei Fehlfunktion
  - Entscheidungsfindende Prozesse müssen nachvollziehbar sein



AI Biotech Nuclear Climate Podcasts

## DIE KI-LEITSÄTZE VON ASIOMAR

Diese Leitsätze wurden auf der Asiomar Konferenz 2017 (Videos hier) durch den hier beschriebenen Prozess beschlossen.



Click here to see this page in other languages: [English](#) [Chinese](#) [Japanese](#) [Korean](#) [Russian](#)

Künstliche Intelligenz (KI, engl. artificial intelligence, AI) hat uns bereits nützliche Anwendungen geliefert, die jeden Tag von Menschen auf der ganzen Welt gebraucht werden. Ihre stetige Weiterentwicklung, gelenkt von den hier beschriebenen Grundsätzen, wird den Menschen in den kommenden Jahrzehnten und Jahrhunderten spektakuläre Möglichkeiten bieten, sich zu helfen und zu verbessern.

### Forschungsthemen

1) **Forschungsziel:** Das Ziel von KI-Forschung sollte lauten, keine ungerichtete, sondern nützliche und wohltätige Intelligenz zu erschaffen.

2) **Forschungsgelder:** Investitionen in KI sollten immer auch solcher Forschung zugutekommen, die ihre wohltätige Nutzung sichert. Dazu gehört die Betrachtung schwieriger Fragen in Bereichen der Computerwissenschaft, Wirtschaft, Barrierefreiheit, Ethik und Sozialwissenschaft.

## Fazit

1. **Software-Anwendungen sind wie Messer:** Diese können für gute und für schlechte Zwecke verwendet werden – manchmal ist Unterscheidung Ansichtssache bzw. eine Frage der Position
2. **Programmiererinnen und Programmierer tragen ethische Verantwortung** für ihre Werke und müssen sich deshalb Gedanken über deren Konsequenzen machen (wie auch Ärzte, Physiker, Chemiker etc.)
3. **Neue Software-Lizenzen und ethische Richtlinien** schaffen Klarheit was gut und was schlecht ist und sollten deshalb vermehrt in der Praxis angewendet werden

