

**Technologie im Gespräch**  
**Discussing Technology**

**Hannes Androsch, Wolfgang Knoll,  
Anton Plimon (Hg. Eds.)**

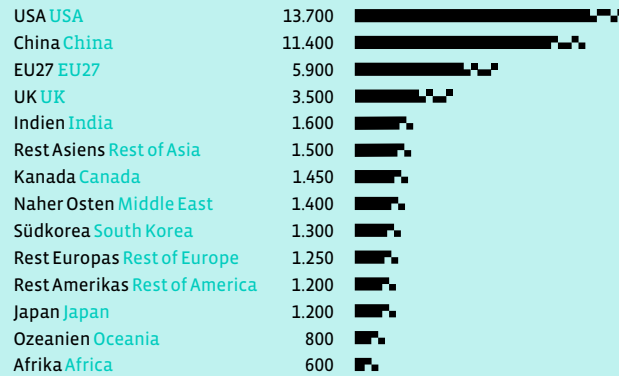
**KI in der Praxis**  
**Applying AI**

Jahrbuch zu den Alpbacher Technologiegesprächen 2022  
Alpbach Technology Symposium Yearbook 2022

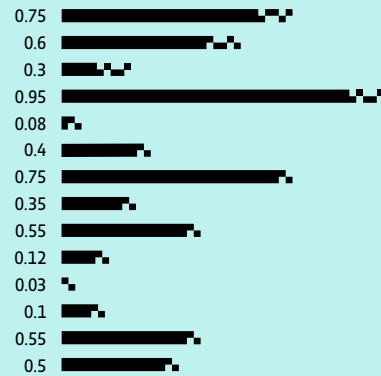
# Künstliche Intelligenz Artificial Intelligence

## 1 Zahl der Organisationen, die sich mit KI beschäftigen Number of organizations dealing with AI

Zahl der KI-Player  
Number of AI-Players

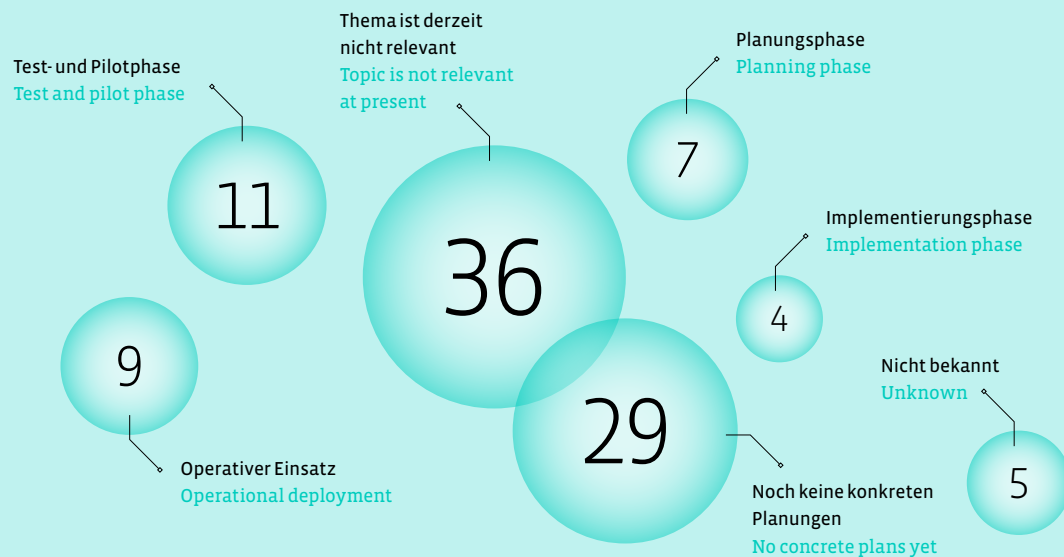


Zahl der KI-Player bezogen auf das BIP  
Number of AI-Players based on the GDP



## 2 Wie weit sind Unternehmen beim Einsatz von KI? How far along are companies in the use of AI?

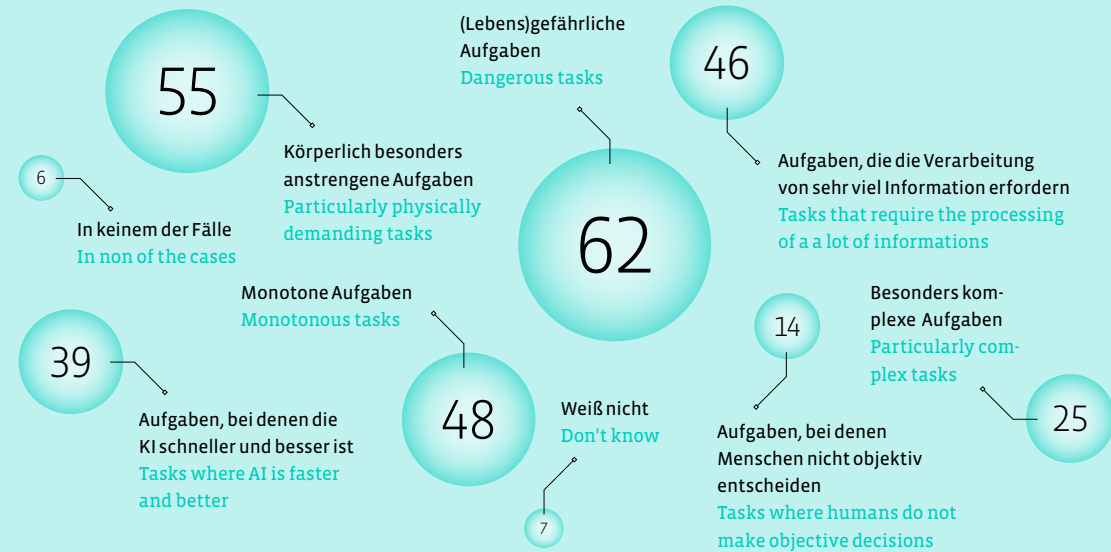
Befragung von 355 österreichischen Unternehmen  
Survey with 355 austrian companies



Angaben in Prozent der Befragten  
Data in percent of respondents

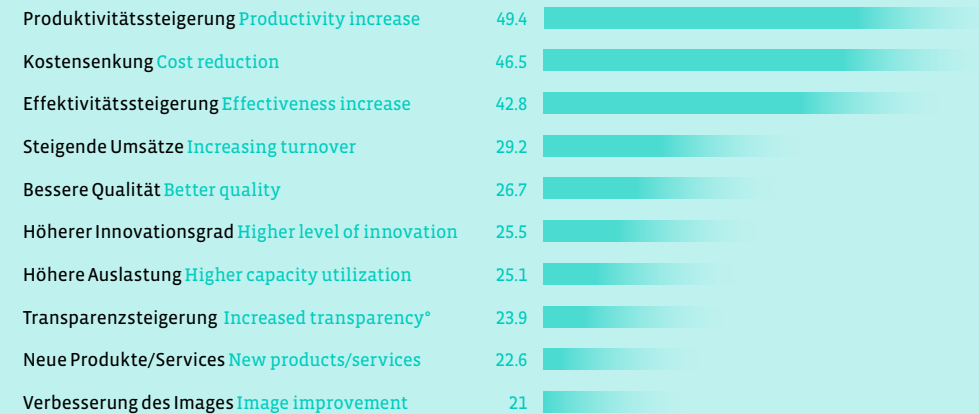
## 3 Bei welchen Aufgaben kann der Einsatz von KI anstelle des Menschen sinnvoll sein? For which tasks can the use of AI instead of humans be useful?

Befragung von 1.000 Deutschen zwischen 16 und 75 Jahren  
Survey with 1.000 Germans aged 16 to 75 (Ipsos online panel)



## 4 Woran messen Sie den Erfolg Ihrer Machine-Learning-Projekte? How do you measure the success of your machine learning projects?

Befragung von Entscheidern und IT-Fachleuten aus 360 Unternehmen  
Survey of decision-makers and IT professionals from 360 companies



### Quellen Sources

- 1 AI Watch Index 2021/RC Technical Report  
AI Watch Index 2021/RC Technical Report
- 2 Media Innovation Report 2019, nextMedia. Hamburg  
Media Innovation Report 2019, nextMedia. Hamburg
- 3 Statista/TÜV Verband  
Statista/TÜV association
- 4 Studie Machine Learning (INFORM DataLab)  
Study Machine Learning (INFORM DataLab)

Angaben in Prozent der Befragten  
Data in percent of respondents

## Inhalt Contents

- Hannes Androsch/Martin Kugler  
4 Wir werden die Krisen überwinden – und Technologie hilft uns dabei!  
*We will overcome the crises – and technology will help us!*

### Allgemeines zu KI *General aspects of AI*

- 16 Methoden der Künstlichen Intelligenz  
*Methods of Artificial Intelligence*
- 28 Metaversum: Wo Cyberspace und KI auf die reale Welt treffen  
*Metaverse: Where Cyberspace and AI Meet the Real World*
- 38 Interview Helga Nowotny  
„KI wird definitiv bleiben“  
*“AI is definitely here to stay”*

### Praktische Anwendungen von KI *Practical applications of AI*

- 54 Interview Andreas Kugi  
„Ohne Domänen-Wissen hat man kaum Chancen auf Erfolg“  
*“Without specific domain knowledge, there is little chance of success”*
- 68 Intelligente Helfer im Alltagsleben  
*Intelligent tools in everyday life*
- 74 Sprachalgorithmen: Wenn Computer Geschichten schreiben  
*Language Algorithms: When computers write stories*
- 80 Sicherheit & Fake News  
*Security & Fake News*
- 86 Wie KI Ärzt:innen hilft  
*How AI is Helping Doctors*

ARTTEC zeigt Schnittstellen von Kunst, Technologie und Wissenschaft

ARTTEC highlights interfaces between art, technology, and science.



Diesjähriger Kulturpartner ist das Angewandte Interdisciplinary Lab (AIL).

This year's culture partner is Angewandte Interdisciplinary Lab (AIL).



- 94 Automatisierung: Optimale Prozesse und (teil)autonome Maschinen  
*Automation: Optimal processes and (semi-)autonomous machines*

- 102 High-Tech Stadtplanung  
*High-Tech Urban Planning*

- 106 Wissenschaft bekommt neue Einsichten  
*New Scientific Insights*

- 112 Interview Ross King  
„KI ist kein Wundermittel“  
*“AI is not a magic bullet”*

### Diskussionen über KI *Discussing AI*

- 126 Aktuelle Debatten über KI  
*Current Debates concerning AI*

- 136 Green IT: Wie KI mit Energie und Emissionen zusammenhängt  
*Green IT: How AI relates to energy consumption and emissions*

- 144 Interview Katja Schechtner  
„Die Frage ist: Macht ein KI-System etwas besser als der Mensch?“  
*“The question is: Does an AI system do something better than humans?”*

### Positionen der Kunst *Positions of Art*

- 152 Interview Process-Studio for Art and Design (Martin Grödl und Moritz Resl)  
„KI ist ein potentes Werkzeug“  
*“AI is a potent tool”*

- 162 Angewandte Interdisciplinary Lab (AIL)

- 176 Alpbacher Technologiegespräche: „The New Europe“  
*Alpbach Technology Symposium: “The New Europe”*



Hannes Androsch und Martin Kugler

## Wir werden die Krisen überwinden – und Techno- logie hilft uns dabei!

**Die multiple Krise, die wir derzeit erleben, schürt Sorgen, Ängste und Nöte. Sie kreierte aber auch Möglichkeiten, die wir ergreifen müssen. Einen wichtigen Schlüssel dazu bieten Technologien wie etwa Künstliche Intelligenz, mit deren Hilfe wir den Krisenerscheinungen im wahrsten Sinn des Wortes entwachsen können: Die Technologien, die wir heute schaffen, sichern den Wohlstand in der Zukunft.**

Die Anhäufung von Schocks, Krisen und Umbrüchen, die wir gerade erleben, sucht in der jüngeren Geschichte ihresgleichen. Ein aktueller Bericht des Stockholmer Friedensforschungsinstituts SIPRI hält im Detail fest, dass wir in einer neuen Ära von komplexen und oft unvorhersagbaren Risiken leben. Dabei überlagern sich mehrere Krisen und verstärken einander gegenseitig: Corona-Krise, Klimawandel, Biodiversitätsschwund, Energiekrise, der Krieg in der Ukraine, Inflationskrise, Rezessionsgefahr sowie viele andere geopolitische Spannungen und weltweite Verwerfungen, die das Leben von immer mehr Menschen immer stärker beeinträchtigen. Wir spüren die Auswirkungen der vermässelten Energiewende genauso wie die Sanktionen und Embargos, die Konsequenzen der zerrissenen Lieferketten genauso wie die hektischen Versuche, ausbleibende Lieferungen fossiler Energieträger teuer zu ersetzen.

Die Preisexplosion bei Energie und bei Rohstoffen bringt alle Volkswirtschaften der Erde durcheinander und vermindert das Wirtschaftswachstum deutlich. Ärmere Länder sind zudem mit einer ernsthaften Bedrohung der Ernährungssicherheit ihrer Bevölkerung konfrontiert. Das UN-Welternährungsprogramm (WFP) zählte im Frühsommer 345 Millionen Menschen in 82 Ländern, die akut Hunger leiden – mehr als doppelt so viele wie vor Corona. Von globaler Bedeutung ist auch die Verschiebung der „Terms of Trade“, der Austauschbeziehungen zwischen verschiedenen Staaten und Regionen: Die Exporteur:innen bekommen einen größeren Anteil am Kuchen, und zwar auf Kosten der Importeur:innen. Die einen sind Wohlstandsgewinner:innen, die anderen Wohlstandsverlierer:innen. Diese Umverteilung verändert die Machtverhältnisse und verschiebt die Grenzen zwischen den Machtblöcken. Und diese Transformation der wirtschaftlichen und sozialen Verhältnisse ist wiederum ein gefährlicherer Zündstoff für die politische Stabilität, ja für die Demokratie als solche.

Hannes Androsch and Martin Kugler

## We will overcome the crises – and technology will help us!

**The multiple crisis that we are currently going through is fueling worries, fears, and difficulties. But it is also creating opportunities that we must take advantage of. Technologies such as Artificial Intelligence provide an important key in this area, with the help of which we can literally outgrow the crisis: the technologies that we are creating today will secure our prosperity in the future.**

The accumulation of shocks, crises, and upheavals that we are currently experiencing is unparalleled in recent history. A recent report by the Stockholm Peace Research Institute (SIPRI) provides details showing that we are living in a new era of complex and often unpredictable risks. Several crises are overlapping and reinforcing each other: the COVID crisis, climate change, loss of biodiversity, the energy crisis, the war in Ukraine, inflation crisis, risk of recession and many other geopolitical tensions are affecting the lives of an ever-increasing number of people. We are feeling the effects of the screwed up energy transition as well as the sanctions and embargoes; the consequences of disrupted supply chains; and also the hectic attempts to compensate for the lack of supplies of fossil fuels at a high cost.

The price explosion that is affecting energy and raw materials is turning all of the global economies upside down and is significantly reducing economic growth. Poorer countries are in addition facing serious threats to the secure food supply for their populations. In early summer, the UN World Food Program (WFP) counted 345 million people in 82 countries suffering from acute hunger – more than twice as many as before COVID. The shift in the “terms of trade”, the exchange relationships among different countries and regions, is also of global importance: exporters are gaining a larger slice of the cake, at the expense of importers. Some people are gaining in wealth, while others are losing out. This redistribution is changing the balance of power and shifting the boundaries between the power blocs. And this transformation in the economic and social conditions is in turn putting political stability, and even democracy, at risk.

## **Aus der Krise herauswachsen**

All diese Krisenerscheinungen werden uns noch länger begleiten, als uns lieb ist. Uns stehen keine rosigen Zeiten bevor – und niemand weiß, welche weiteren Konsequenzen die komplexe Melange der verschiedenen Krisenursachen noch haben werden. Aber auch die schwierigsten Zeiten können und müssen überwunden werden. Und jede Krise kreiert auch Möglichkeiten: Bei allen Sorgen und Nöten muss man auch im Blick haben, wo die Chancen liegen, die es jetzt zu ergreifen gilt. Einen wichtigen Schlüssel dazu bieten Technologien. Wohlstandsverluste kann man längerfristig nur durch effizienzverbessernde, produktivitätssteigernde und innovationsbeschleunigende Maßnahmen wettmachen – man muss ihnen im wahrsten Sinn des Wortes entwachsen. Wir müssen jetzt in der Krise das schaffen, was den Wohlstand in der Zukunft sichert.

Künstliche Intelligenz, der dieses Jahrbuch gewidmet ist, ist eine dieser großen Chancen: Diese Technologie hat das Potenzial, die Wirtschaft effizienter, unser Arbeiten produktiver, unseren Umgang mit der Umwelt schonender und unser Zusammenleben gedeihlicher zu machen. Dass mit dieser Technologie auch Gefahren verknüpft sind, liegt auf der Hand. So wie bei jeder Technologie: Ein Messer ist ein äußerst nützliches Werkzeug, aber man kann damit auch andere Menschen verletzen oder töten. Ebenso ist das Auto eine große Errungenschaft, doch das bedeutet nicht, dass man unbekümmert mit 200 km/h durch die Gegend rasen muss und Menschen und Umwelt gefährdet.

## **Metaverse & Co: Digitalisierung treibt die Entwicklung**

Digitalisierung ist einer der zentralen Treiber für Innovationen und für die Transformation von Wirtschaft und Gesellschaft. Dabei muss uns klar sein, dass damit ein zusätzlicher Strombedarf einhergeht – was im Übrigen auch für Wasserstoff gilt. Ein Ausdruck der rasanten Entwicklung digitaler Technologien ist das „Metaverse“ – die Idee, die reale und virtuelle Welt innig miteinander zu verschränken. Sämtliche Elemente des Alltags sollen miteinander vernetzt werden, sodass die User:innen sie mit einem Klick, einem Wischen oder einer anderen Aktion abrufen können. Egal ob es darum geht, Anrufe zu tätigen, Geld abzuheben, Freizeit zu planen, Urlaube zu buchen, Erlebnisse zu teilen oder Zeit mit Freund:innen zu verbringen: Das alles soll im Metaverse möglich sein. Jede Handlung in der virtuellen Welt ist dabei mit einer Handlung in der realen Welt verknüpft. Dadurch bekommen digitale Techniken, wie etwa KI oder Vernetzung, einen noch stärkeren Einfluss auf unser Leben und Arbeiten. Diese vom Technologiekonzern Meta (Facebook) und anderen forcierte Technologie – oder genauer: dieses Bündel von Technologien, die kombiniert werden – bietet immense Möglichkeiten, unsere Art der Kommunikation, des Austauschs von Ideen, des Handelns oder unsere Kreativität zu transformieren. Von Beobachter:innen wird sie als „Zukunft des Internet“, als „nächster Schritt nach der Mobilkommunikation“ oder als „nächste große Arbeitsplattform“ gesehen. Manche sehen indes einen bloßen Hype, der bald wieder vorbei sein wird. Wieder andere machen darauf aufmerksam, dass die großen Technologiekonzerne dadurch noch mehr Macht bekommen. Das „Metaverse“ trägt jedenfalls die Gefahr in sich, noch weiter in eine passive Konsumhaltung abzurutschen – wovon nur einer profitieren wird: nämlich der/die Anbieter:in, der/die dann immer mehr Macht über die Menschen bekommt. Die unglaubliche Potenz einer Handvoll Technologieunter-

## **Rising from the crisis**

All of these crises will remain with us for longer than we would like: we don't have rosy times ahead of us – and nobody knows what further consequences the complex mixture of the various issues that caused the crises will have. But we can and must get through these difficult times. And each crisis also creates opportunities: with all the worries and hardships, you also have to keep an eye on where the opportunities are that need to be seized now. Technologies are an important key feature for this. Loss of prosperity can only be made up for in the long term by measures that improve efficiency, increase productivity, and accelerate innovation – this loss needs to be outgrown in the truest sense of the word. During the crisis now, we need to create something that will ensure prosperity in the future.

Artificial Intelligence, which this yearbook is dedicated to, is one of these great opportunities: this technology has the potential to make the economy more efficient, our work more productive, improve how we treat the environment, and ensure that we coexist in a more prosperous manner. This technology also obviously involves risks. It's the same for any technology: a knife is an extremely useful tool, but it can also be used to hurt or kill other people. Likewise, the car is a great achievement – but that doesn't mean that you have to be racing at 200 km/h and put people and environment at risk.

## **Metaverse & Co.: Digitization is driving development**

Digitization is one of the key drivers behind innovation and for the transformation of the economy and society. At the same time, we must be aware that this will require additional electricity – which, by the way, also applies to hydrogen. One expression of the rapid development of digital technologies is the “metaverse” – the idea of intimately intertwining the real world and the virtual world. All elements of everyday life are to be networked with each other so that users can access them with one click, swipe or other action. Whether it's making phone calls, withdrawing money, planning leisure time, booking vacations, sharing experiences, or spending time with friends: the intention is for all this to be possible within the metaverse. Any action we take in the virtual world is linked to an action in the real world. As a result, digital technologies such as AI and networking will have an even greater impact on our life and work. This technology is being pushed by the technology group Meta (Facebook) and others – or more precisely this bundle of combined technologies – offers huge opportunities to transform our way of communicating, exchanging ideas, doing business, and our creativity. Observers see it as the “future of the Internet”, the “next step after mobile communications” or the “next major working platform”. Some see it as mere hype that will soon be over. There are others who draw attention to the fact that this is giving the big technology companies even more power: in any case, the “metaverse” is at risk of slipping even further towards passive consumerism – from which only one will benefit: the provider, who will then gain more and more power over people. The incredible potency of a handful of technology companies will then become even more of a threat to our social system and our political constitu-



nehmen wird dann noch mehr zu einer Gefahr für unser Gesellschaftssystem und unsere politische Verfasstheit. Geballte wirtschaftliche Kraft kann keine akzeptable Legitimation für die Gestaltung von Politik sein. Der Schlusssatz von Abraham Lincolns „Gettysburg Address“ gilt heute genauso wie vor 159 Jahren: „... dass die Regierung des Volkes, durch das Volk und für das Volk, nicht von der Erde verschwinden möge.“ Dieser Wunsch ist zunehmend durch post-faktische autokratische Systeme gefährdet. So sterben Demokratien einen stillen Tod in Dunkelheit.

Alle Schritte in Richtung „Metaverse“ verschärfen jedenfalls alle derzeitigen Problemfelder der Digitalisierung und des Internets – von Datenschutz und Privacy-Aspekten über Cyberwar, Cybercrime und Sicherheitsbedenken bis hin zu (immer der Entwicklung hinterherhinkenden) rechtlichen und ethischen Herausforderungen.

### **Die Entwicklung mitgestalten**

Die Frage ist, wie wir zu einer akzeptablen neuen Ordnung kommen. Wie können wir das disruptive Potenzial von Künstlicher Intelligenz zu unserem Besten nutzen? Wie kommen wir Menschen zu einer gedeihlichen Koexistenz mit künstlich intelligenten Systemen, die nicht bedrohlich auf uns wirken, sondern uns bei unseren großen Herausforderungen helfen? Wie stellen wir sicher, dass Automatisierung auch diesmal mehr Jobs schafft, als durch die Technik ersetzt werden? Wie können wir trotz des Datenhungers von KI-Systemen eine angemessene Privacy bewahren? Wie entgehen wir den Extremen eines totalitären Überwachungsstaats auf der einen Seite und eines oligopolistischen Überwachungskapitalismus auf der anderen? Wie können wir verhindern, dass übermächtige Algorithmen unser Leben dominieren und autonome Killerroboter zu tödlichen Gefahren werden? Wie können wir sicherstellen, dass eine Verknüpfung zwischen KI und synthetischer Biologie keine nicht mehr rückgängig zu machenden Folgen zeitigt?

Eine Voraussetzung für die Lösung der vielen ethischen, rechtlichen und ökonomischen Fragen im Zusammenhang mit KI ist, dass wir endlich eine digitale Aufholjagd starten. Nur zeitgemäße Bildung, Aus- und Weiterbildung, genügend Mittel für Forschung, Technologieentwicklung und Umsetzung sowie die Entwicklung eines Innovationsgeistes können uns auf die Gewinnerseite bringen. Das betrifft zum einen das Verständnis der Menschen für die neuen Technologien, deren Funktionsweise, Chancen und Gefahren. Angst vor etwas zu haben, ist immer ein schlechter Ratgeber. Viel besser ist es, die Entwicklung, ausgestattet mit dem nötigen Rüstzeug, selbst mitzugestalten.

### **Fakten statt Wunschdenken**

Zum anderen geht es auch um die Zukunft Europas. Derzeit wird der „alte Kontinent“ zwischen den Machtblöcken USA und China zerrieben, die bei der Anwendung von Cyberphysics meilenweit voraus sind. Europa hat allerdings auch seine Stärken: So arbeiten europäische Wissenschaftler:innen in der Grundlagenforschung und bei der Methodenentwicklung an der Weltspitze mit. Und wie ein aktueller Bericht der EU-Kommission festhält, hat Europa in einigen Bereichen, wie zum Beispiel bei autonomen Robotern, sogar einen klaren Wettbewerbsvorteil gegenüber der Konkurrenz in Ost und West. Diese Stärken gilt es auszubauen. Denn die Technologien, die heute entwickelt werden, werden maßgeblich die globalen Verhältnisse der Zukunft mitbestimmen. Wer heute bei der Entwicklung von Zukunftstechnolo-

tion. Concentrated economic power cannot provide justification for the shaping of politics! The closing sentence of Abraham Lincoln’s Gettysburg Address is as true today as it was 159 years ago: “[...] that government of the people, by the people, for the people, shall not perish from the earth.” This desire is increasingly threatened by post-factual autocratic systems. Thus democracies die a silent death in obscurity.

In any case, all steps towards the “metaverse” exacerbate all current problem areas regarding digitization and the Internet – from data protection and privacy aspects to cyberwar, cybercrime and security concerns and legal and ethical challenges (which are always lagging behind developments).

### **Shaping developments**

The question is how we can arrive at an acceptable new order. How can we use the disruptive potential of Artificial Intelligence to serve our best interests? How do we humans reach a thriving coexistence with Artificial Intelligence systems that will not appear threatening to us, but instead help us with our major challenges? How do we ensure that automation will once again create more jobs than are replaced by this technology? How can we ensure an appropriate level of privacy despite the greed of AI systems for data? How do we escape the extremes of a totalitarian surveillance state on the one hand and oligopolistic surveillance capitalism on the other? How can we prevent overpowering algorithms from dominating our lives and autonomous killer robots from becoming deadly threats? How can we ensure that a link between AI and synthetic biology does not have irreversible consequences?

One prerequisite for solving the many ethical, legal, and economic questions relating to AI is that we must finally start a digital race to catch up. Only state-of-the-art education, training and further education, sufficient funding for research, technology development and implementation, as well as developing an innovative spirit and mindset can put us on the winning side. On the one hand, this affects people’s understanding of the new technologies, their functioning, opportunities and risks. Being afraid of something does not help you make the right decisions. It is much better to help shape the development yourself, equipped with the necessary tools.

### **Facts instead of wishful thinking**

On the other hand, it’s also about the future of Europe. The “old continent” is currently under threat of being crushed between the power blocs of the U.S. and China who are miles ahead when it comes to deploying Cyberphysics. Europe, however, does have its strengths: European scientists are among the world’s best in basic research and in developing methods. And as a current report by the EU Commission states, in some areas, such as autonomous robots, Europe even has a clear competitive advantage over the competition in the East and the West. These strengths need to be expanded. This is because the technologies that are being developed today will have a decisive influence on the global conditions of the future. Those who are ahead in the development of future technologies today will make their mark on the world in the future. This applies both from an economic, as well as a political and military point of view.

gien die Nase vorn hat, wird in Zukunft der Welt seinen Stempel aufdrücken. Das gilt sowohl in wirtschaftlicher als auch in politischer und militärischer Hinsicht. Innovationen werden zunehmend zu einem wichtigen machtpolitischen Instrument. Wir brauchen Technologien auch, um unsere großen längerfristigen Herausforderungen – von Klimaschutz und Dekarbonisierung über Umweltschutz, Welternährung und Wasserversorgung bis hin zum demografischen Wandel – bewältigen zu können. Entscheidend dabei ist, dass wir uns nicht nur hehre Ziele setzen, sondern auch wissen, wie wir diese erreichen können. Statt Wunschdenken und ideologischen Luftschlössern brauchen wir einen klaren Plan, wie wir vom heutigen bedrohlichen Zustand zum erwünschten Zielzustand gelangen. In vielen Bereichen sollten wir uns weniger von Illusionsideologien als von Fakten leiten lassen.

### **Rahmen für KI-Anwendungen nötig**

Weltweit fehlt es derzeit an angemessenen Spielregeln für Digitalisierung und Künstliche Intelligenz. Wenn wir nicht in die Falle von Goethes „Zauberlehrling“ tapen wollen – dass uns Technologien über den Kopf wachsen –, brauchen wir so etwas wie eine „Magna Charta“ der Technologie, eine Art digitale Straßenverkehrsordnung für die neuen virtuellen Welten, die zunehmend unser Leben prägen. Diese neuen Regularien können nur gemeinsam entwickelt werden – jegliche Form von Partikularismus, Antiglobalisierung oder nationalistischem Protektionismus ist auch in dieser Hinsicht eine dramatische Fehlentwicklung. Es bedarf vielmehr internationaler, zumindest aber europaweiter Regelungen, denn kein einzelnes Land kann hier allein erfolgreich sein. Dass die EU derzeit an der weltweit ersten umfassenden Regulierung von KI arbeitet, ist in dieser Hinsicht ein Lichtblick – unter der Voraussetzung, dass daraus kein überbordendes und lebensfremdes Reglement wird und dass andere Teile der Welt in der Folge mitziehen. Für Europa ist das eine Chance, um in der Systemrivalität, die die Weltpolitik beherrscht, europäische Werte stärker zu verankern.

Ein stabiler Rahmen für KI ist jedenfalls für die Bereiche, die in diesem Jahrbuch angesprochen werden, eine wichtige Voraussetzung. Hier geht es um die Anwendung von KI-Methoden in bestimmten Domänen und für sehr spezifische Fragestellungen – im Gegensatz zu Systemen, die in vielen Bereichen einsetzbar sind („horizontal AI“) und in Richtung einer „Artificial General Intelligence“ (AGI) gehen. Darunter versteht man die hypothetische Intelligenz eines Computerprogramms, das die Fähigkeit besitzt, jede intellektuelle Aufgabe zu verstehen oder zu lernen, die ein Mensch ausführen kann. Solche Systeme sind reine Zukunftsmusik: Es gibt sie derzeit nicht und wie viele Fachleute meinen, wird das auch noch sehr lange Zeit so bleiben. Ähnliches gilt für komplementäre Ansätze wie etwa „synthetische Intelligenz“: Dabei gehen manche Forscher:innen davon aus, dass die heutigen KI-Methoden Intelligenz nur simulieren würden – gesucht werden stattdessen andere technische Verfahren, die nicht nur intelligent zu sein scheinen, sondern wirklich intelligent sind. Mit solchen Ansätzen befindet man sich inmitten komplexer philosophischer Debatten.

Innovations are increasingly becoming an important tool of political power. We also need technologies to be able to cope with our major long-term challenges – from climate protection and decarbonization via environmental protection, global nutrition and water supply to demographic changes. It is crucial that we do not just set ourselves noble goals, but also know how to achieve them. Instead of wishful thinking and building ideological castles in the air, we need a clear plan about how to move from today's threatening state to the desired target state. There are many areas in which we should be guided less by illusionary ideologies and more by facts.

### **Framework required for AI applications**

Around the globe, appropriate rules for digitization and Artificial Intelligence are currently missing. If we don't want to fall into the trap of Goethe's "sorcerer's apprentice" – i.e., that technology outgrows us – we need something like a "Magna Carta" for technology, something like digital road traffic regulations for the new virtual worlds that are increasingly shaping our digital lives. We can only develop these new regulations together – any form of particularism, anti-globalization or nationalist protectionism is a dramatically misguided development also in this respect. Instead, internationally, or at least on a pan-European basis, regulations are needed, because no single country can be successful in this area on its own. The fact that the EU is currently working on the world's first comprehensive regulation of AI is a ray of hope in this respect – provided that it does not result in excessive and unrealistic regulations and that other parts of the world follow the EU. For Europe, this is an opportunity to anchor European values more firmly in the system rivalry that dominates world politics.

In any case, a stable framework for AI is an important prerequisite for the areas addressed in this yearbook. This relates to the application of AI methods in certain areas regarding very specific questions. These systems can only be used within each of their narrowly limited frameworks. They can, however, result in massive progress in comparison with the status quo. In the technical jargon, this is known as "vertical AI" – in contrast to systems that can be used in many different areas ("horizontal AI") and that are heading in the direction of "Artificial General Intelligence" (AGI). By this is meant the hypothetical intelligence of a computer program that has the ability to understand or learn any intellectual task that a human can perform. These kind of systems are the stuff of the future: they don't currently exist yet, and according to many experts, they may not exist for a very long time to come. The same is true for complementary approaches such as "synthetic intelligence": some researchers assume that today's AI methods only simulate intelligence – instead, they are looking for other technical processes that not only appear to be intelligent, but are really intelligent. With such approaches, we are in the midst of complex philosophical debates.

## Fahrplan durch dieses Jahrbuch

Demgegenüber widmet sich dieses Jahrbuch zu den heurigen Alpbacher Technologiegesprächen, die vom AIT Austrian Institute of Technology und ORF Radio Ö1 ausgerichtet werden, jenen Anwendungen von KI, die bereits heute unser Leben und Arbeiten durchdringen. Nach einer Einführung in die technischen Grundlagen von KI wird anhand zahlreicher Beispiele gezeigt, wo wir bei der Anwendung von KI in verschiedenen Domänen derzeit stehen. Das reicht von nützlichen Tools für unser Alltagsleben über die Konzipierung künftiger autonomer Maschinen, die eng mit dem Menschen zusammenarbeiten, und Sprachalgorithmen bis hin zum Einsatz im Gesundheitswesen und in der Stadtplanung. Bei all diesen Beispielen, die unter anderem aus den Laboren des AIT und anderer heimischer Forschungsstätten stammen, wird deutlich, wie die unterschiedlichsten KI-Methoden sinnvoll eingesetzt werden können, wo ihre Grenzen liegen und in welchen Bereichen man besser auf ihren Einsatz verzichten sollte.

Zur Sprache kommen überdies die Regulierung und Governance der neuen Technologien, die ökologischen Auswirkungen von KI sowie ihre Rolle in den Künsten. Ergänzt und vertieft werden diese Themenbereiche durch ausführliche Interviews mit führenden Forscherpersönlichkeiten, wie etwa Helga Nowotny, der „Grande Dame“ der österreichischen Wissenschaftsforschung, oder Andreas Kugi, Automatisierungsforscher an der TU Wien und am AIT, der tiefe Einblicke gibt, wie KI die Werkzeugkiste der Forschung erweitert. ✖

---

**Hannes Androsch**, geboren 1938 in Wien, war in seiner politischen Tätigkeit (SPÖ) Abgeordneter zum Nationalrat (1966–1970), Bundesminister für Finanzen (1970–1981) und Vizekanzler (1976–1981). Danach war er Generaldirektor des Creditanstalt-Bankvereins (1981–1988) und Vorsitzender der Oesterreichischen Kontrollbank AG (1985–1986). 1989 gründete er die AIC Androsch International Management Consulting GmbH und begann 1994 den Aufbau einer industriellen Beteiligungsgruppe (Austria Technologie & Systemtechnik AG, Österreichische Salinen AG u. a.). 2004 errichtete er die „Stiftung Hannes Androsch bei der

Österreichischen Akademie der Wissenschaften“ und ist dort seit 2005 Mitglied des Senats. 2007 bis 2021 war er überdies Aufsichtsratsvorsitzender des AIT Austrian Institute of Technology, bis 2020 Vorsitzender des RFTE Rats für Forschung und Technologieentwicklung und bis Juni 2016 Aufsichtsratsvorsitzender der FIMBAG Finanzmarkt-beteiligungsgesellschaft des Bundes. Er erhielt Ehrendokorate und ist Ehrensensator verschiedener österreichischer und internationaler Universitäten, u. a. der Montanuniversität Leoben und der Universität New Orleans, USA.

## A roadmap through this yearbook

In contrast, this yearbook accompanying the Alpbach Technology Symposium, hosted by the AIT Austrian Institute of Technology and ORF Radio Ö1, is dedicated precisely to those applications of AI that are already permeating our lives and work. After an introduction to the technical basics of AI, numerous examples will show where we currently stand in the application of AI in various domains. These range from useful tools for our everyday lives, the design of future autonomous machines that collaborate with humans, and language algorithms, to applications in healthcare and urban planning. All these examples, which come from the laboratories of the AIT and other domestic research institutes, among others, make it clear how the diverse AI methods can be used sensibly, where their limits lie and in which areas it is better not to use them.

The regulation and governance of the new technologies, the ecological effects of AI and its role in the arts are also discussed. These topics are complemented by in-depth interviews with leading researchers, such as Helga Nowotny, the grande dame of science studies in Austria, or Andreas Kugi, automation researcher at TU Wien and AIT, who provides deep insights into how AI is expanding the toolbox of research. ✖

---

**Hannes Androsch**, born in Vienna in 1938, was during his political career (SPÖ) Member of the National Assembly (1966–1970), Federal Minister of Finance (1970–1981), and Vice Chancellor (1976–1981). After this, he served as Director General of Creditanstalt-Bankverein (1981–1988) and as Chairman of Oesterreichische Kontrollbank AG (1985–1986). In 1989, he founded AIC Androsch International Management Consulting GmbH, and in 1994 initiated the establishment of an industrial investment group (Austria Technologie & Systemtechnik AG, Österreichische Salinen AG, etc.). In 2004, he founded the “Hannes Androsch Foundation at the Austrian Academy of Sci-

ences,” where he has been a member of the senate since 2005. From 2007 to 2021, he was Chairman of the Supervisory Board of AIT Austrian Institute of Technology, until 2020 Chairman of RFTE Council for Research and Technological Development, until 2016 Chairman of the Supervisory Board of FIMBAG Finanzmarkt-beteiligungsgesellschaft des Bundes. He has received honorary doctorates from and is an honorary senator of various Austrian and international universities, including the Montanuniversität Leoben and the University of New Orleans, USA.



Allgemeines

zu KI

General

aspects of AI

# Methoden der Künstlichen Intelligenz

**Im landläufigen Sprachgebrauch meint man mit Künstlicher Intelligenz (KI) häufig nur maschinelle Lernverfahren. Doch es gibt viel mehr KI-Methoden, die jeweils ihre Vorzüge und Schwächen haben. Ein Überblick.**

Es gibt keine allgemein anerkannte Definition von „Künstlicher Intelligenz“ (KI) – nicht zuletzt deshalb, weil es auch keine einheitliche Definition von „Intelligenz“ gibt. Wenn man einmal Begriffe wie emotionale oder soziale Intelligenz beiseitelässt, hat Intelligenz auf jeden Fall etwas mit kognitiver Leistungsfähigkeit zu tun. Es geht beispielsweise um das Lösen von Problemen, um das Kombinieren unterschiedlicher Faktoren, um Lernen oder um logisches Schlussfolgern. Wobei jedes Forschungsgebiet seinen eigenen Betrachtungswinkel hat.

Wenn von KI die Rede ist, gibt es mindestens ebenso viele Ansichten. Im Online-Lexikon „Wikipedia“ findet sich etwa folgender Eintrag: „Meist bezeichnet künstliche Intelligenz den Versuch, bestimmte Entscheidungsstrukturen des Menschen nachzubilden, indem zum Beispiel ein Computer so gebaut und programmiert wird, dass er relativ eigenständig Probleme bearbeiten kann.“ Manche Expert:innen rücken die Fähigkeit, Probleme zu lösen oder komplexe Ziele zu erreichen, in den Vordergrund. Andere betrachten ein System dann als intelligent, wenn es Fähigkeiten hat, die eigentlich nur dem Menschen zukommen. Wieder andere Expert:innen sehen das logische Denken oder das Lernen als zentral an. Manche betonen Aspekte wie Verstehen oder Entscheidungsfindung. Und wieder andere meinen kurz und knapp: Künstliche Intelligenz ist Intelligenz, die nicht natürlich ist.

Eine hilfreiche Unterscheidung hat vor knapp 50 Jahren der US-Philosoph John R. Searle etabliert: Er sprach auf der einen Seite von „starker KI“, wenn die technische Nachbildung der allgemeinen menschlichen Intelligenz, die ein universelles Werkzeug zum Lösen von Problemen in praktisch allen Teilbereichen darstellt, gemeint ist. Heute ist dafür auch der Begriff „Artificial General Intelligence“ (AGI) gebräuchlich. In diesem Zusammenhang werden viele heiß umstrittene Fragen diskutiert – etwa, ob Computer Menschen irgendwann überflügeln werden („Singularität“), ob Maschinen

# Methods of Artificial Intelligence

**In common language, Artificial Intelligence (AI) often just refers to machine learning methods. But there are many more AI methods that each have their merits and weaknesses. An overview.**

There is no universally accepted definition of “Artificial Intelligence” (AI) – not least so because there is also no single definition of “intelligence.” Setting aside terms like “emotional” or “social intelligence”, intelligence definitely has something to do with cognitive performance. For instance, it has to do with solving problems, combining different factors, learning, or logical reasoning. With each field of research having its own perspective.

Where AI is concerned, there are at least as many views. The German version of the online encyclopedia “Wikipedia” contains the following passage (translated here): “Artificial Intelligence generally refers to the attempt to emulate certain decision-making structures of humans, for example by building and programming a computer in such a way that it can process problems relatively independently.” Some experts emphasize the ability to solve problems or achieve complex goals. Others consider a system to be intelligent if it has capabilities that are really only available to humans. Still other experts see logical thinking or learning as key. Some emphasize aspects such as understanding or decision-making. And still others say, in short: Artificial Intelligence is intelligence that is not natural.

The U.S. philosopher John R. Searle established a helpful distinction nearly 50 years ago: On the one hand, he spoke of “strong AI” when referring to the technical replica of general human intelligence, which is a universal tool for solving problems in virtually all subfields. Today, the term “Artificial General Intelligence” (AGI) is also commonly used for this. Many hotly debated issues are being discussed in this context – such as whether computers will eventually outperform humans (“singularity”), whether machines can develop something like consciousness, or whether cyborgs could be the next stage of evolution. AGI systems do not currently exist; they merely exist in theory.

so etwas wie ein Bewusstsein entwickeln können oder ob Cyborgs eine nächste Stufe der Evolution sein könnten. AGI-Systeme gibt es derzeit nicht, sie sind eine reine Gedankenspielerei.

Dem stellte Searle die „schwache KI“ gegenüber, die sich mit der Entwicklung konkreter technischer Systeme beschäftigt, die wohldefinierte Aufgaben übernehmen können, die bisher dem Menschen vorbehalten waren (oder überhaupt nicht lösbar waren). Alle heute realisierten KI-Systeme fallen in diese Kategorie. Wenn es um Anwendungen von KI-Methoden in bestimmten Domänen (etwa bei autonomen Fahrzeugen, bei der Mustererkennung oder Sprachverarbeitung) geht, spricht man meist von „vertical AI“ – und genau darum geht es in diesem Jahrbuch.

### **Viele unterschiedliche Methoden der KI**

In den vergangenen 70 Jahren – beginnend mit der legendären Dartmouth-Konferenz, bei der 1956 der Begriff „Künstliche Intelligenz“ erstmals explizit formuliert wurde – wurden sehr viele verschiedene Methoden entwickelt, die einer Maschine ein intelligentes Verhalten verleihen können (oder zumindest den Anschein dazu).

Grundsätzlich lassen sich KI-Systeme auf Basis der Repräsentation des Wissens in sogenannte symbolische und sub-symbolische Systeme unterteilen. In einem symbolischen KI-System werden Regeln und Beziehungen für Konzepte angewendet, die für Menschen verständlich sind. Darunter fallen beispielsweise regelbasierte Systeme oder Expertensysteme. Sub-symbolische KI-Systeme sind hingegen für Menschen weitgehend eine „black box“, deren Inhalte nicht einfach zu verstehen sind. In diese Kategorie fallen insbesondere künstliche neuronale Netze.

### **Symbolische KI-Systeme**

Schon lange in Anwendung sind regelbasierte Systeme, bei denen man versucht, auf Basis von Wenn-Dann-Regeln intelligentes Verhalten nachzubilden. Ein einfaches Beispiel ist ein Entscheidungsbaum, der vorgibt, wie sich ein System verhalten soll, wenn bestimmte Bedingungen gegeben sind. Regelbasierte Systeme bestehen aus einer Datenbank von Fakten (Datenbasis), einer Menge von Regeln (Regelbasis) und einem Kontrollsystem mit Regelinterpretier (Inferenzmaschine). In diese Gruppe fallen auch herkömmliche Methoden des „Natural Language Processing“, bei denen beispielsweise vorgegeben wird, wie Inputs miteinander verknüpft werden – wo etwa eine neue Information in ein vorgegebenes Textgerüst eingefügt werden soll. Solche Systeme sind vergleichsweise einfach zu implementieren und vor allem leicht zu verstehen.

Bei wissensbasierten Systemen ist das Wissen in symbolischer Form explizit in Computersystemen repräsentiert. Darüber hinaus gibt es vielfältige Regeln, wie diese Wissensbausteine verknüpft werden. Schon seit den 1970er-Jahren werden Expertensysteme genutzt, in denen Daten in formalisierter Form (Ontologien) gespeichert sind und aus denen logische Schlüsse gezogen werden können. Für Mediziner:innen bieten solche Systeme wertvolle Hilfestellungen bei der Diagnose und Therapie von Erkrankungen.

Searle contrasted this with “weak AI,” which is concerned with the development of concrete technical systems that can take on well-defined tasks that were previously reserved for humans (or were not solvable at all). All of the AI systems being implemented today fall into this category. When it comes to applications of AI methods in specific domains (such as autonomous vehicles, pattern recognition or speech processing), one usually speaks of “vertical AI” – and this is exactly what this yearbook is about.

### **Many different methods of AI**

Over the past 70 years – starting with the legendary Dartmouth Conference, where the term “Artificial Intelligence” was explicitly formulated for the first time in 1956 – a great many different methods have been developed that can endow a machine with intelligent behavior (or at least make it display such behavior).

Generally, AI systems can be divided into so-called symbolic and sub-symbolic systems based on the representation of knowledge. A symbolic AI system applies rules and relationships to concepts that are understandable to humans. This includes, for example, rule-based systems or expert systems. Sub-symbolic AI systems, on the other hand, are mostly a “black box” for humans, a box whose contents are not easy to understand. Artificial neural networks in particular fall into this category.

### **Symbolic AI systems**

Rule-based systems have been in use for a long time. These try to emulate intelligent behavior on the basis of if-then rules. A simple example is a decision tree that specifies how a system should behave given certain conditions. Rule-based systems consist of a database of facts (database), a set of rules (rule base) and a control system with rule interpreter (inference engine). This group also includes conventional methods of “Natural Language Processing”, in which, for instance, it is specified how inputs are to be linked with each other – e.g., where new information is to be inserted into a given text framework. These types of systems are comparatively simple to implement and, above all, easy to understand.

In knowledge-based systems, knowledge is explicitly represented in symbolic form in computer systems; moreover, there are a variety of rules for how these knowledge building blocks are linked. Since as early as the 1970s, expert systems have been used in which data is stored in formalized form (ontologies) and from which logical conclusions can be drawn. Such systems offer physicians valuable assistance in the diagnosis and treatment of diseases. A knowledge-based computer system is also the now famous “Watson”, which surprised the public in 2011 when it beat human players in the US quiz show “Jeopardy!”. The system is based on a high-level semantic search engine that captures the meaning of a question asked in natural language and

Ein wissensbasiertes Computersystem ist auch das berühmt gewordene „Watson“, das im Jahr 2011 die Öffentlichkeit überraschte, als es in der US-Quizsendung „Jeopardy!“ menschliche Spieler besiegte. Das System basiert auf einer hochwertigen semantischen Suchmaschine, die den Sinn einer in natürlicher Sprache gestellten Frage erfasst und in einer großen Datenbank, die ebenfalls Texte in natürlicher Sprache enthält, innerhalb kurzer Zeit die relevanten Passagen und Fakten findet. Diese von IBM entwickelte Software wird mittlerweile in vielen Anwendungsbereichen praktisch eingesetzt. Der Nachteil dieser Systeme ist der immense Aufwand bei der Erfassung des menschlichen Wissens und der Umwandlung in eine maschinenlesbare Wissensbasis.

Regelbasiert arbeiten auch sogenannte Software-Agenten. Darunter versteht man Computerprogramme, die zu genau spezifiziertem eigenständigem Verhalten fähig sind. In Abhängigkeit von verschiedenen Zuständen laufen definierte Verhaltensprogramme ab, ohne dass eine Steuerung von außen nötig wäre. Ein Beispiel dafür sind „Webcrawler“ von Internet-Suchmaschinen, die selbsttätig Webseiten besuchen und den Links folgen.

### **Sub-symbolische KI-Systeme**

In sub-symbolischen KI-Systemen werden Daten verarbeitet, zum Beispiel in verschiedene Gruppen klassifiziert, ohne dass vom Menschen exakte Regeln dafür vorgegeben werden. Ein verbreitetes Verfahren sind „Support Vector“-Maschinen, die einen multidimensionalen Datenraum mit bestimmten mathematischen Verfahren strukturieren. Häufig genutzt werden auch statistische Verfahren wie „Markov-Ketten“ (Markov chains). Diese beruhen darauf, dass in vielen Fällen ein bestimmter Datenpunkt nur vom vorherigen Datenpunkt abhängt – typischerweise in einem zeitlichen Verlauf. In einem stochastischen Prozess lernt ein derartiges Modell, Wahrscheinlichkeiten für das Eintreten des nächsten Datenpunktes anzugeben. So können aus gesammelten Daten Vorhersagen für das künftige Systemverhalten getroffen werden. Angewandt wird das beispielsweise bei der Überwachung von Maschinen („predictive maintenance“): Wenn eine Vorhersage, etwa für die Temperatur oder die Motordrehzahl, von der realen Entwicklung (die mit Sensoren gemessen wird) abweicht, signalisiert diese Anomalie ein Problem.

### **Neuronale Netze**

Die größten Fortschritte in der KI werden derzeit durch Machine Learning mit Hilfe von künstlichen neuronalen Netzen erzielt. In neuronalen Netzen werden im Computer die Eigenschaften von natürlichen Nervenzellen nachgeahmt, die auf vielfältige Weise miteinander verknüpft sind. Ein Lernprozess manifestiert sich in der Stärke der Verknüpfungen zwischen den künstlichen Neuronen. Durch wiederholte Aktivierung im Zuge einer Trainingsphase besitzen neuronale Netze die Fähigkeit, ein Modell auf Basis von Daten automatisiert zu erlernen. Wenn viele neuronale Netze (layers) übereinandergestapelt werden, wobei die Neuronen auch über diese Ebenen hinweg miteinander verknüpft sein können, spricht man von „Deep

rapidly finds the relevant passages and facts in a large database that also contains texts in natural language. This software, developed by IBM, is now in practical use in many application areas. The drawback of these systems is the immense effort required for capturing human knowledge and converting it into a machine-readable knowledge base. So-called software agents also work on a rule-based basis. These are computer programs that are capable of precisely specified independent behavior. Defined behavioral programs run depending on various states, without the need for external control. An example of this are “web crawlers” of Internet search engines, which autonomously visit websites and follow the links.

### **Sub-symbolic AI systems**

In sub-symbolic AI systems, data is processed, e.g., classified into different groups, without exact rules being specified by humans. A widely used method are “support vector” machines which structure a multidimensional data space with certain mathematical methods. Statistical processes such as “Markov chains” are also frequently used. These are based on the fact that in many cases a certain data point depends only on the previous data point – typically in a time sequence. In a stochastic process, such a model learns to specify probabilities for the occurrence of the next data point. This way, collected data can be used to predict future system behavior. This is applied, for example, in the monitoring of machines (“predictive maintenance”): If a prediction, e.g., for temperature or engine speed, deviates from the real development (which is measured with sensors), this anomaly signals a problem.

### **Neural networks**

The greatest advances in AI are currently being made through machine learning using artificial neural networks. In neural networks, computers mimic the properties of natural neurons, which are interconnected in a variety of ways. A learning process manifests itself in the strength of the connections between the artificial neurons. Through repeated activation in the course of a training phase, neural networks have the ability to perform automated learning of a model based on data. When many neural networks (layers) are stacked on top of each other, with neurons also being able to link across these layers, one speaks of “deep learning”. There are hundreds of thousands to many billions of network nodes and edges, whose behavior is output in an output layer. Humans can no longer understand or interpret the processes within the network at all. Machine learning with neural networks comes in three main varieties:

- **Supervised learning**  
Supervised learning means that the AI system is fed with so-called “annotated” training data – the algorithm learns a relationship between given pairs of inputs and outputs. A case in point:

Learning“. Dabei gibt es hunderttausende bis viele Milliarden Netzwerkknoten und -kanten, deren Verhalten in einem Output-Layer ausgegeben wird. Für den Menschen sind die Vorgänge innerhalb des Netzwerks völlig unüberschaubar und nicht mehr zu deuten. Machine Learning mit neuronalen Netzen tritt in drei wesentlichen Spielarten auf:

- **Überwachtes Lernen (supervised learning)**

Überwachtes Lernen bedeutet, dass das KI-System mit sogenannten „annotierten“ Trainingsdaten gefüttert wird – der Algorithmus lernt dabei einen Zusammenhang zwischen gegebenen Paaren von Ein- und Ausgaben. Ein Beispiel: Ein Bild eines Tieres wird von Experten als „Hund“ oder als „Katze“ gelabelt. Anhand sehr vieler Beispiele wird das künstliche neuronale Netz immer besser, diese Klassifikation selbstständig vorzunehmen. Da häufig große Mengen an Trainingsdaten benötigt werden, um zu guten Ergebnissen zu kommen, ist der Aufwand hoch. Entscheidend ist in jedem Fall die Qualität der Trainingsdaten – etwa wie zuverlässig diese annotiert sind, ob der Datensatz vollständig ist (oder Daten für bestimmte Bereiche unterrepräsentiert sind oder ganz fehlen) oder ob es systematische Verzerrungen (bias) gibt. Gute Modelle, so eine Faustregel von KI-Experten, können nur auf Basis guter Daten erstellt werden. Überwachtes Lernen wird heute in vielen Bereichen sehr erfolgreich angewendet – etwa bei der Erkennung von Bildern, bei der Handschriftenerkennung, bei Übersetzungssoftware oder bei der Synthese von Sprache (GPT 3). Sehr erfolgreich sind auch Machine-Learning-Systeme, die mit künstlich generierten Lerndaten trainiert werden – etwa wenn es darum geht, rasch Ergebnisse für sehr aufwendigen Berechnungen zu bekommen.

- **Unüberwachtes Lernen (unsupervised learning)**

Beim unüberwachten Lernen benötigt man keine Trainingsdaten. Vielmehr analysiert das KI-System Daten hinsichtlich ihrer Ähnlichkeit beziehungsweise ihrer Distanz. Ein Beispiel für unüberwachtes Lernen ist die Suche nach Klassen in einer Menge von Datenpunkten: In der Regel ist die Suche nach Klassen in einer Menge von Datenpunkten die einzige Eingabe beim unüberwachten Lernen die Anzahl der Cluster, die gefunden werden sollen. Andere Methoden des unüberwachten Lernens bedienen sich mathematischer Verfahren wie der Dimensionsreduktion, der Hauptkomponentenanalyse oder der Dichteanalyse. Angewandt werden solche Methoden beispielsweise bei der Vorhersage des Kauf- und Nutzer:innen-Verhaltens im Onlinehandel oder bei Empfehlungssystemen auf Onlineplattformen. Möglich wird es weiters, in kontinuierlichen Datenströmen Anomalien zu finden – also etwa Abweichungen vom erwarteten Systemverhalten, Fehler in einer Produktionsstraße oder einen Hackerangriff auf ein Stromnetz.

- **Bestärkendes Lernen (reinforcement learning)**

Unter dem Begriff bestärkendes Lernen werden Lernverfahren zusammengefasst, die aus Erfahrung lernen – nicht aber durch die Vorgabe von Trainingsbeispielen. Hierzu interagiert das System mit seiner Umgebung

An image of an animal is labeled by experts as a “dog” or as a “cat”. Based on very many examples, the artificial neural network becomes better and better at making this classification on its own. Since large amounts of training data are often needed to get good results, this requires a lot of effort. The quality of the training data is crucial in any case – i.e. how reliably it is annotated, whether the data set is complete (or whether data for certain areas are underrepresented or missing altogether), or whether there are any systematic biases. According to a rule of thumb among AI experts, good models can only be built on the basis of good data. Today, supervised learning is used very successfully in many areas – e.g., in image recognition, in handwriting recognition, in translation software, or in speech synthesis (GPT 3). Another very successful set are machine learning systems that are trained with artificially generated learning data – e.g., when the aim is to quickly obtain results for very complex calculations.

- **Unsupervised learning**

Unsupervised learning does not require training data. Rather, the AI system analyzes data in terms of their similarity and distance. An example of unsupervised learning is searching for classes in a set of data points: Typically, the only input in unsupervised learning is the number of clusters to find. Other methods of unsupervised learning use mathematical techniques such as dimensionality reduction, principal component analysis, or density analysis. Such methods are applied, for example, in predicting purchase and user behavior in online retailing or in recommendation systems on online platforms. It is also possible to find anomalies in continuous data streams – e.g., deviations from expected system behavior, errors in a production line, or a hacker attack on a power grid.

- **Reinforcement learning**

The term reinforcement learning is used to describe learning methods that learn from experience – but not by providing training examples. To achieve this, the system interacts with its environment and receives a reward for correct results. Similar to an animal being trained, the AI system tries to remember the consequences of its actions and thereby to maximize the reward. Probably the most famous example of reinforcement learning is AlphaGo Zero: This system learned the game of Go without prior knowledge and without being given the rules simply by being rewarded for winning a game through trial and error; within just three days it had become so good that it played better than its programmed predecessors and also far better than human Go masters. This principle is of great interest for many areas of



und erhält für richtige Ergebnisse eine Belohnung. Ähnlich einem Tier, das dressiert wird, versucht das KI-System, sich die Konsequenzen seines Handelns zu merken und dadurch die Belohnung zu maximieren. Das wohl berühmteste Beispiel für bestärkendes Lernen ist AlphaGo Zero. Dieses System erlernte das Spiel Go ohne vorherige Kenntnisse und ohne Vorgabe der Regeln – einfach indem es dafür belohnt wurde, wenn es durch Ausprobieren ein Spiel gewann. Innerhalb von nur drei Tagen war es so gut geworden, dass es besser spielte als seine programmierten Vorgänger und auch weitaus besser als menschliche Go-Meister. Dieses Prinzip ist für viele Bereiche der Technik, etwa in der Automatisierung oder in der Robotik, sehr interessant – denn Verstärkungslernalgorithmen setzen keine Kenntnis eines exakten mathematischen Modells voraus.

### **Kombination verschiedener Netzwerktypen**

Für diese verschiedenen Arten des Machine Learning wurden mittlerweile viele Architekturen von künstlichen neuronalen Netzwerken entwickelt, die jeweils ihre Stärken und Schwächen haben. Häufig eingesetzt werden derzeit etwa „Convolutional Neural Networks“ (CNN), bei denen es die mathematische Methode der „Faltung“ ermöglicht, abstrakte Merkmale der Input-Daten herauszufiltern. Neuronale Netzwerke werden je nach den konkreten Anforderungen in unterschiedlichsten Konfigurationen mit anderen Komponenten verknüpft. Ein prominentes Beispiel sind „Generative Adversarial Networks“ (GAN). Diese bestehen aus zwei künstlichen neuronalen Netzwerken, die gemeinsam ein Nullsummenspiel durchführen. Eines davon (der Generator) erstellt zum Beispiel ein Bild, das das zweite neuronale Netzwerk (der Diskriminator) bewertet. Im Verlauf lernen beide dazu und die Ergebnisse werden immer besser.

### **Schwachstellen von neuronalen Netzen**

Die Erfahrungen der jüngsten Vergangenheit zeigen, dass künstliche neuronale Netze für viele Aufgaben eine sehr brauchbare Methode zur Lösung von Problemen sind. Doch sie haben auch einige grundsätzliche Schwachstellen (im Nachfolgenden dargestellt nach: Stefan Buijsman. Ada und die Algorithmen. Wahre Geschichten aus der Welt der künstlichen Intelligenz. Übersetzt von Bärbel Jänicke. 236 Seiten, Verlag C.H.Beck):

- **Neuronale Netze treffen Entscheidungen auf intransparente Weise:** Die formalen mathematischen Regeln hinter neuronalen Netzen beschreiben nur eine Summe sowie eine Methode, um diese Summe möglichst gering zu halten – was bedeutet, dass möglichst wenige Fehler gemacht werden. Es gibt aber keine expliziten Vorgaben, wie dies am besten zu erreichen ist. Daher ist es sehr schwierig, herauszufinden, wie das Netz intern arbeitet. In dieser Hinsicht ist es wie eine „black-box“ - man kann nicht hineinsehen, wie das System zu einer Entscheidung kommt. Daher mangelt es neuronalen Netzen grundsätzlich an „explainability“. Eine höhere Transparenz ist aber in vielen Fällen, etwa aus Haftungsgründen, erwünscht. Dafür werden derzeit Methoden entwickelt.

technology, such as automation or robotics – because reinforcement learning algorithms do not require knowledge of an exact mathematical model.

### **Combination of different network types**

Many artificial neural network architectures have now been developed for these different types of machine learning, each with its strengths and weaknesses. “Convolutional Neural Networks” (CNN) are currently frequently used, e.g., where the mathematical method of “convolution” makes it possible to filter out abstract features of the input data. Neural networks are linked to other components in a wide variety of configurations, depending on the specific requirements. A prominent example are “Generative Adversarial Networks” (GAN). These consist of two artificial neural networks that jointly execute a zero-sum game. For example, one of them (the generator) creates an image that the second neural network (the discriminator) evaluates. As they progress, both learn and the results get better and better.

### **Weak points of neural networks**

Recent experience shows that artificial neural networks are a very useful method for solving many problems. But they also have some fundamental weak points (in the following presented by: Stefan Buijsman. Ada und die Algorithmen. [“Ada and the Algorithms”]. Wahre Geschichten aus der Welt der künstlichen Intelligenz. [“True stories from the world of Artificial Intelligence”] Translated into German by Bärbel Jänicke. 236 pages, published by C.H.Beck):

- **Neural networks make decisions in a non-transparent way:** The formal mathematical rules behind neural networks describe only a sum and a method to keep this sum as small as possible – which means that as few mistakes as possible are made. But there are no explicit guidelines on how best to achieve this. It is therefore very difficult to figure out how the network works internally: It is a “black box” in this respect; one cannot look inside to see how the system arrives at a decision. Therefore, neural networks basically lack “explainability”. In many cases, however, greater transparency is desirable, for example for liability reasons. Methods are currently being developed for this purpose.

- **Neuronale Netze sind nie besser als ihre Daten:**  
KI, die aus Daten lernt, ist auf die Qualität der Daten angewiesen. Der Computer übernimmt damit auch alle Verzerrungen und Diskriminierungen, die sich in den Daten verbergen. So erwiesen sich beispielsweise Algorithmen, die aus einem Gesicht das Geschlecht bestimmen sollten, bei weißen Männern als sehr erfolgreich, bei farbigen Frauen hingegen als sehr fehlerhaft. Der Grund dafür liegt in den Lerndaten, in denen verschiedene Bevölkerungsgruppen nicht repräsentativ vertreten waren. Das Problem wird dadurch verschärft, dass neuronale Netze hinsichtlich der Kriterien für Entscheidungen – siehe oben – intransparent sind und man selbst offensichtliche Fehler nur schwer finden kann.
- **Neuronale Netze können nicht gut generalisieren:**  
KI läuft schnell aus dem Ruder, wenn man mit Situationen arbeitet, die für den Computer ungewohnt sind. Das betrifft beispielsweise fehlende Daten für bestimmte Situationen, sodass die KI-Systeme schlicht Unsinn produzieren und spektakuläre Fehlleistungen zeigen – wie etwa bei den berühmt gewordenen kapitalen Unfällen von autonomen Autos in außergewöhnlichen Situationen.
- **Neuronale Netze nehmen alles wörtlich:**  
Lernende KI-Systeme sind auch von den Vorgaben und Anweisungen abhängig, die Menschen ihnen eingeben, wenn sie eine bestimmte Leistung abrufen wollen. Wenn diese Regeln missverständlich sind – wenn ein Computer eine Regel anders interpretiert als von uns Menschen intendiert, weil der Maschine unser Hintergrundwissen fehlt –, folgt das System blind den formalen Regeln und liefert unsinnige Ergebnisse. Diese Fehler zu entdecken ist einmal mehr wegen der mangelnden Transparenz der Entscheidungsfindung schwierig.
- **Neuronale Netze sind spezialisiert:**  
Ein künstliches neuronales Netz mag eine bestimmte Aufgabe sehr gut erledigen, ist bei einer anderen Aufgabe aber völlig überfordert. Denn heutige neuronale Netze sind extrem spezialisiert – sie werden gezielt für jene Aufgabe maßgeschneidert, für die sie herangezogen werden. So arbeiten KI-Systeme, die Bilder erkennen, völlig anders als Systeme, die mit Sprache umgehen können. Derzeit besteht keinerlei Aussicht auf einen Ansatz, mit dem verschiedene Fähigkeiten miteinander verknüpft werden können. ✕

- **Neural networks are never better than their data:**  
AI that learns from data relies on the quality of the data. The computer therefore also inherits all biases and discriminations hidden in the data. For example, algorithms designed to determine gender from a face proved very successful with white men, but very error-prone with women of color. The reason for this resides in the learning data, where different groups were not represented according to their proportion in the general populace. The problem is exacerbated by the fact that neural networks are not transparent with respect to the criteria for decisions – see above – and it is difficult to find even obvious errors.
- **Neural networks do not generalize well:**  
AI quickly gets out of hand when working with situations that are unusual for the computer. This applies, for example, to missing data for certain situations, so that the AI systems simply produce nonsense and show spectacular failures – such as in the now famous fatal accidents of autonomous cars in extraordinary situations.
- **Neural networks take everything literally:**  
Learning AI systems also depend on the specifications and instructions that humans feed them when they want to invoke a certain performance. If these rules are misunderstood – if a computer interprets a rule differently than what we humans intended because the machine lacks our background knowledge – the system blindly follows the formal rules and delivers nonsensical results. The detection of these errors is once again difficult because of the lack of transparency in the decision-making process.
- **Neural networks are specialized:**  
An artificial neural network may do a particular task very well, but may be completely overwhelmed at another task. This is because today's neural networks are extremely specialized – they are specifically tailored for the one task for which they are called upon. For instance, AI systems that recognize images work completely differently from systems that can handle speech. At present, there is no prospect of an approach that could combine different capabilities. ✕

# Metaversum: Wo Cyberspace und Künstliche Intelligenz auf die reale Welt treffen

**Im (geplanten) künftigen Metaversum werden digitale Technologien unser Leben in bisher ungeahntem Ausmaß durchdringen. Dafür ist die Integration zahlreicher Schlüsseltechnologien in ein Gesamtsystem erforderlich. Ein Überblick.**

Wie so häufig bei Innovationen im Bereich der Digitalisierung hat auch der Begriff „Metaversum“ seine Wurzeln in der Science-Fiction-Literatur. Im Jahr 1992 schrieb Neal Stephenson den Roman „Snow Crash“, in dem er eine Art digitaler Parallelwelt entwirft, in der sich Menschen mithilfe von Avataren so bewegen können, wie wir in der wirklichen Welt. Seit Facebook-Gründer Mark Zuckerberg im Herbst 2021 seinen Konzern in „Meta“ umtaufte und den Aufbau eines Metaversums ankündigte, ist darum ein regelrechter Hype entstanden. Laut Zuckerbergs Ankündigung ist das Metaversum nicht mehr nur ein Internet, auf das man schaut, sondern in dem man teilnimmt: Man begibt sich in Form eines Avatars mithilfe von Virtual oder Augmented Reality in virtuelle Räume, wo man mit anderen kommuniziert, spielt oder kreativ ist, wo man Dinge ein- und verkauft, Dienstleistungen abruft, Medien konsumiert, lernt usw.

Sämtliche Elemente des Alltags sollen miteinander vernetzt werden, sodass die User:innen sie mit einem Klick, einem Wischen oder einer anderen Aktion abrufen können. Egal ob es darum geht, Anrufe zu tätigen, Geld abzuheben, Freizeit zu planen, Urlaube zu buchen, Erlebnisse zu teilen oder Zeit mit Freunden zu verbringen. Das alles soll im Metaverse möglich sein.

Eine einheitliche Definition für „Metaversum“ gibt es – bisher – nicht. Zusammengesetzt aus „meta“ (griech. inmitten, zwischen oder nach) und „Universum“, drückt der Begriff aus, dass sich etwas auf einer höheren Ebene befindet oder dem übergeordnet ist, was damit in Verbindung gebracht wird. Beschrieben wird also ein Universum aus technischen Entwicklungen, das über das hinausgeht, was wir bisher kennen.

Der Wissenschaftsautor und Investor Matthew Ball hat sich als einer der ersten mit dem Metaversum beschäftigt und 2020 den einflussreichen Artikel „The Metaverse: What It Is, Where to Find it, and Who Will Build It“ veröffentlicht. Eine erweiterte Version findet sich auf seiner Website

# Metaverse: where Cyberspace and Artificial Intelligence Meet the Real World

**In the (planned) metaverse of the future, digital technologies will permeate our lives to an unprecedented extent. This will require the integration of numerous key technologies into one overall system. Here is an overview.**

As is so often the case with innovations in the field of digitization, the term “metaverse” has its roots in science fiction literature. In 1992, Neal Stephenson wrote the novel “Snow Crash” in which he creates a kind of parallel digital world where people can use avatars to move around as we do in the real world. There has been considerable hype about this since Facebook founder Mark Zuckerberg renamed his company “Meta” in the autumn of 2021 and announced the construction of a metaverse. According to Zuckerberg’s announcement, the metaverse is no longer just an Internet that you view, but rather one that you participate in: You enter virtual spaces in the shape of an avatar using virtual or augmented reality, and there, you communicate with others, play, or create, you buy and sell, access services, consume media, study, and so on.

All elements of everyday life are to be networked with each other so that users can access them with one click, swipe or other action. Whether it’s making phone calls, withdrawing money, planning leisure time, booking vacations, sharing experiences, or spending time with friends – the intention is for all this to be possible within the metaverse.

There is – so far – no uniform definition of “metaverse”. The term is made up of “meta” (Greek for in the middle, between, or after) and “universe” and wants to express that something is at a higher level or is superior to what it is associated with. It is therefore used to describe a universe of technological developments that goes beyond what we have known so far.

Science writer and investor Matthew Ball was one of the first to delve into the metaverse, publishing the influential 2020 article “The

(<https://www.matthewball.vc>) unter dem Titel „The Metaverse Primer“. Ball beschreibt die wichtigsten Eigenschaften des Metaversums so:

- Es kann niemals beendet oder pausiert werden. Es läuft immer weiter.
- Es ist immer live. Zwar gibt es wie in der realen Welt auch zeitlich begrenzte Events, doch das Metaversum als Ganzes lebt stets in Echtzeit.
- Die Zahl von Teilnehmerinnen und Teilnehmern ist nicht begrenzt.
- Das Metaversum unterhält seine eigene Wirtschaft. Man bestellt, kauft, zahlt, handelt, schuldet, tilgt, akquiriert und veräußert Güter, Leistungen und Firmen in einem holistischen Universum.
- Das Metaverse umfasst die digitale Welt genauso wie die physische. In der Mengenlehre würde man sagen: Das Metaverse stellt die Obermenge von digitaler und physischer Welt dar.

### **Verschränkung von digitaler und realer Welt**

In gewisser Weise ist das Metaversum ein logischer nächster Schritt bei der Entwicklung digitaler Technologien, die immer weitere Bereiche unseres Lebens umfassen und durchdringen. Nach Internet, Mobilfunk, Social Media und Internet of Things geht es nun insbesondere darum, die virtuelle Welt immer stärker immersiv (virtuelle Umgebungen werden zunehmend als real empfunden) zu machen, um komplett in sie einzutauchen – und gleichzeitig die Grenzen zwischen digitaler und realer Welt immer durchlässiger zu machen und im Endeffekt beide zu verschmelzen. Jede Handlung in der virtuellen Welt ist mit einer Handlung in der realen Welt verschränkt, und umgekehrt; jede Veränderung eines Zustands löst eine Spiegelung im anderen Universum aus. Dadurch bekommen digitale Technologien, wie etwa Künstliche Intelligenz und durchgreifende Vernetzung, einen noch stärkeren Einfluss auf unser aller Leben – ein Leben, in dem nichts und niemand mehr offline ist. „Das Metaversum wird ein ständiger Begleiter des täglichen Lebens sein“, befindet etwa die Zukunftsforscherin und Gründerin des Future Today Instituts, Amy Webb (Handelsblatt, 11. 1. 2022). „Jeder Alltagsvorgang wird in einem Kosmos der vernetzten Geräte und Gegenstände stattfinden, die mit KI lernen, immer besser zu funktionieren“, heißt es in einer Analyse der Süddeutschen Zeitung („Was kommt nach dem Handy?“ 2./3. 10. 2021). KI wage nun „den Sprung aus dem Cyberspace in die wirkliche Welt“.

### **Wirtschaftlich große Erwartungen**

Von vielen als „Zukunft des Internets“ oder als „nächster Schritt nach der Mobilkommunikation“ bezeichnet, gibt es aber auch sehr viel Kritik – etwa daran, dass sich nun ein Weltkonzern diese Idee zwecks Maximierung seiner Gewinne unter den Nagel reißen will. Manche empfinden schon das Wort Metaversum als „Anmaßung und Größenwahn einer Technokratie mit Allmachtsanspruch“. Alle Schritte in Richtung Metaversum verschärfen klarerweise auch alle derzeitigen Problemfelder des Internets – von Datenschutz und Privacy-Aspekten über Sicherheitsbedenken bis hin zu (immer der Entwicklung hinterherhinkenden) rechtlichen und ethischen Herausforderungen.

Metaverse: What It Is, Where to Find it, and Who Will Build It.” An expanded version is available on his website (<https://www.matthewball.vc>) under “The Metaverse Primer“. Ball describes the most important features of the metaverse like this:

- It can never be stopped or paused. It keeps on going.
- It is always live. While there are limited-time events, just like in the real world, the metaverse as a whole always lives in real-time.
- There is no limit on the number of participants.
- The metaverse maintains its own economy. You order, buy, pay, trade, owe, repay, acquire and sell goods, services and companies within a holistic universe.
- The metaverse encompasses both the digital and physical world.
- Set theory would say: The metaverse represents the superset of the digital world and the physical world.

### **Intertwining of the Digital and Real World**

In a way, the metaverse is a logical next step in the development of digital technologies that are encompassing and permeating more and more areas of our lives. After the Internet, mobile communications, social media and the Internet of Things, the real aim now is to make the virtual world increasingly immersive (virtual environments are increasingly perceived as real) in order to completely immerse oneself in them – and at the same time to make the boundaries between the digital and the real world more and more permeable, ultimately merging both. Every action in the virtual world is intertwined with an action in the real world, and vice versa; every change of state is reflected in the other universe. As a result, digital technologies such as artificial intelligence and comprehensive networking are having an even greater impact on all of our lives – a life in which nothing and nobody will be offline anymore. “The metaverse will be a constant companion in everyday life,” says futurologist and founder of the Future Today Institute, Amy Webb (Handelsblatt, 11 January 2022). “Every day-to-day process will take place within a cosmos of networked devices and items that learn to work increasingly better with AI,” stated an analysis by the Süddeutsche Zeitung (“What comes after the cell phone?” 2/3 October 2021) It claimed that AI is now “making the leap from cyberspace to real life”.

### **High Economic Hopes**

Described by many as the “future of the Internet” or the “next step after mobile communications”, there is also a lot of criticism, for example that one global corporation now wants to hijack this idea in order to maximize its profits. Some are already considering the mere term “metaverse” as the “arrogance and megalomania of a technocracy with a claim to omnipotence”. All of the steps in the direction of the metaverse clearly also exacerbate all of the current problem areas of the Internet

Laut einer Analyse des kanadischen Forschungsinstituts Emergen Research lag das globale Marktvolumen aller für das Metaversum nötigen Technologien im Jahr 2020 bei knapp 48 Milliarden us-Dollar – erwartet wird ein Wachstum bis zum Jahr 2028 um jährlich mehr als 40 Prozent auf rund 830 Milliarden Dollar. Die us-Investmentbank Morgan Stanley schätzt das mögliche Volumen von mit dem Metaversum verbundenen Produkten und Dienstleistungen allein in den USA auf 8,3 Billionen Dollar.

### **Spiele und Digitale Zwillinge**

Es gibt einige Vorläufer des Metaversums. So etwa das 2003 ins Leben gerufene „Second Life“. Dabei handelt es sich um eine von Benutzer:innen gestaltete virtuelle Welt, in der Menschen durch Avatare interagieren, spielen, Handel treiben und anderweitig kommunizieren können. Die Second-Life-„Welt“ existiert in einer großen Serverfarm, die von Linden Lab betrieben wird und die von der Client-Software als kontinuierliche 3D-Animation dargestellt wird, in die zusätzliche Audio- und Videostreams eingebunden werden können. Personen und/oder Unternehmen können auf diese Weise miteinander in Kontakt treten oder sich gegenseitig virtuelle Waren oder Dienstleistungen anbieten. Die Kommunikation erfolgt per öffentlichem oder privatem Chat, optional auch mündlich. Gleichgesinnte können Gruppen bilden. Das Programm wurde bereits für Schulungen und virtuelle universitäre Vorlesungen genutzt, auch Livekonzerte lassen sich virtuell durchführen. Durch die Einbindung einer virtuellen Währung (L\$, Linden-Dollar), die in eine reale Währung (us-\$) transferiert werden kann, ist Second Life in den realen Wirtschaftskreislauf eingebunden.

Ein anderer Vorläufer des Metaversums ist „Fortnite“: In diesem Online-Spiel gibt es unzählige Charaktere, die berühmte Figuren oder Persönlichkeiten ins Spiel bringen. In diversen Online-Events gab es bereits Live-Konzerte von Künstlern zu bestaunen. Spieler:innen versammeln sich ganz einfach mit ihren virtuellen Avataren vor der Bühne und hören zu, tanzen oder chatten miteinander.

Stärker in der realen Welt verankert ist eine Technologie, die sich „Digitaler Zwilling“ nennt. In immer mehr Bereichen werden Prozesse und Produkte digital modelliert, bevor sie in die Realität umgesetzt werden. In diesen Digitalen Zwillingen werden beispielsweise Arbeitsprozesse simuliert, um Planungsfehler schon in einem frühen Stadium zu vermeiden und alle Abläufe in der realen Fabrik berechnen und dadurch optimieren zu können. Ähnliches gilt für die Medizin, wo komplizierte chirurgische Eingriffe in allen Details in einem digitalen Modell geplant und geübt werden können. Oder bei Crashtests, die immer häufiger (und damit materialschonend) im Computer simuliert werden.

– from data protection and privacy aspects to security concerns and legal and ethical challenges (which always lag behind).

According to an analysis by the Canadian research institute Emergen Research, the global market volume for all technologies required for the metaverse was almost 48 billion us dollars in 2020 – with annual growth of more than 40% to around \$830 billion expected by 2028. The us investment bank Morgan Stanley estimates the potential volume of products and services associated with the metaverse at \$8.3 trillion in the us alone.

### **Games and Digital Twins**

There have been some precursors to the metaverse. A case in point is “Second Life”, launched in 2003. This is a user-created virtual world in which people can interact, play, trade, and communicate in other ways using avatars. The Second Life “world” exists in a large server farm operated by Linden Lab, which the client software presents as a continuous 3D animation into which it is possible to incorporate additional audio and video streams. This means that people and/or companies can get in touch with each other or offer each other virtual goods or services. Communication takes place via public or private chat, with verbal communication also being an option. Like-minded people can form groups. The program has already been used for training courses and virtual college lectures. Even live concerts can take place virtually. By integrating a virtual currency (L\$, Linden dollar), which can be transferred into a real currency (us\$), Second Life is integrated into the real economic cycle.

Another precursor of the metaverse is “Fortnite”. There are countless characters in this online game, even including famous characters or personalities. There have already been live concerts by artists during various online events. Players simply gather in front of the stage as their virtual avatars and listen, dance, or chat with each other.

More firmly anchored in the real world is the “digital twin” technology. In an increasing number of areas, processes and products are being modeled digitally before they are implemented in reality. These digital twins are used, for example, to simulate work processes in order to avoid planning errors already at an early stage and to be able to calculate and therefore optimize all of the processes to take place in a real factory. The same applies to medicine, where it is possible to plan and practice complicated surgical procedures in a digital model. Or in crash tests, which are increasingly being simulated on the computer (and therefore in a way that does not destroy the material).



## Integration vieler Einzeltechnologien

Das Metaversum beruht auf der nahtlosen Integration vieler verschiedener Einzeltechnologien, die sich allesamt derzeit in stürmischer Entwicklung befinden. Eine Forschergruppe um Lik-Hang Lee (Korea Advanced Institute of Science and Technology/KAIST) hat kürzlich eine umfassende Studie über die einzelnen Komponenten inklusive des aktuellen Forschungsbedarfs erstellt (Journal of LATEX Class Files, Vol. 14, No. 8, Sept. 2021; DOI: 10.13140/RG.2.2.11200.05124/8). Aufgelistet werden insbesondere folgende acht Schlüsseltechnologien:

- **Virtual/Augmented/Mixed Reality:** Die wichtigste Schnittstelle zwischen User:innen und Metaversum werden auf absehbare Zeit Virtual Reality-Headsets sein, über die man in die virtuelle Welt eintauchen kann und die ein Einblenden von virtuellen Inhalten in ein Bild der realen Welt erlauben – bis sich vielleicht einmal Hologramme, riesige Bildschirme, Großprojektionen oder ähnliche Technologien durchsetzen. Entwicklungspotenzial gibt es dabei insbesondere beim Tragekomfort, bei der Auflösung und bei der Latenzzeit, denn jede Verzögerung kann zu Unwohlsein (Virtual Reality Sickness) führen.
- **Interaktivität:** Tastaturen, Mäuse und Joysticks sind wohl nicht die ultimativen Input-Geräte für das Metaversum. In Entwicklung sind zum Beispiel Kleidungsstücke mit eingebauten Sensoren oder EEG-Elektroden zur Steuerung. Auch an der Einbindung anderer Sinne wird gearbeitet.
- **Internet der Dinge / Robotik:** Der Interaktion zwischen virtueller und realer Welt kommt besondere Bedeutung zu. Etwa durch Daten über die reale Welt durch Sensoren in Geräten. Oder durch kollaborative Roboter.
- **Künstliche Intelligenz:** Immer mehr Dienstleistungen, auf die künftig über das Metaversum zugegriffen werden kann, nutzen Methoden der Künstlichen Intelligenz. Aber auch das Design der Avatare und autonome Akteure in der virtuellen Welt erfordert KI. Um dies Web-basiert zu ermöglichen, sind effiziente Modelle nötig.
- **Blockchain:** Diese Technologie kann zum einen für die Etablierung von Währungen oder zum Handel von Non-Fungible Tokens (NFT) genutzt werden. Zum anderen bieten Blockchains auch eine Möglichkeit zur sicheren Dezentralisierung von Datenbeständen.
- **Computer Vision:** Für Systeme der Augmented und Extended Reality sind realistische 3D-Modelle der Umgebung nötig – erforderlich dafür sind fortgeschrittene Methoden der Computer Vision (Lokalisierung, Mapping, Objekterkennung, Scene Understanding).

## Integration of Many Individual Technologies

The metaverse is based on the seamless integration of many different individual technologies, all of which are currently undergoing very rapid development. A research group led by Lik-Hang Lee (Korea Advanced Institute of Science and Technology/KAIST) recently conducted a comprehensive study of the individual components, including the current research needs (Journal of LATEX Class Files, Vol. 14, No. 8, Sept. 2021; DOI: 10.13140/RG.2.2.11200.05124/8). The following eight key technologies are listed in particular:

- **Virtual/Augmented/Mixed Reality:** For the foreseeable future, the most important interface between users and the metaverse will be virtual reality headsets, which can be used to immerse oneself in the virtual world and which allow virtual content to be superimposed on an image of the real world – until perhaps holograms, huge-sized screens, large-scale projections or similar technologies take over. There is potential for development in particular in terms of wearing comfort, resolution, and latency periods, because any delay can lead to discomfort (virtual reality sickness).
- **Interactivity:** Keyboard, mouse, and joystick are probably not the ultimate input devices for the metaverse. For example, items of clothing with built-in sensors or EEG electrodes for control are being developed. Work is also being done with regard to connections to the other senses.
- **The Internet of Things / Robotics:** The interaction between the virtual and the real world is of particular importance. This can take place, for example, through data about the real world from sensors in devices, or through collaborative robots.
- **Artificial Intelligence:** An increasing number of services that it will be possible to access via the metaverse in the future will use methods of artificial intelligence. The design of avatars and autonomous actors in the virtual world also requires AI, however. Efficient models are required to make this possible in a web-based manner.
- **Blockchain:** This technology can be used to establish currencies or to trade in non-fungible tokens (NFT) On the other hand, blockchains also offer the opportunity to decentralize databases in a secure way.
- **Computer Vision:** For augmented and extended reality systems, realistic 3D models of the environment are required – advanced methods of computer vision (localization, mapping, object recognition, scene understanding) are required for this purpose.

- **Edge / Cloud Computing:** Auf mobilen Endgeräten (VR-Brillen, Headsets, Smartphones etc.) können keine komplexen Prozesse wie z. B. 3D-Simulationen oder KI-Modelle berechnet werden. Um die Datenübertragung zu zentralen Servern zu entlasten und gleichzeitig zu beschleunigen, sind neue Methoden und Architekturen zur dezentralen Berechnung im Netz (Edge Computing) erforderlich, die allen Anforderungen von Sicherheit und Privacy genügen.
- **Netzwerk:** Zur Bewältigung der großen Datenmengen im Metaversum werden zuverlässige und intelligente Netzwerktechnologien, wie 5G, benötigt, die insbesondere die Quality of Experience (QoE; Erlebnis-Qualität für die User:innen) maximieren.

Ob das Konzept des Metaversums wirklich aufgeht, kann derzeit niemand sagen. Faktum ist jedenfalls, dass alle großen Technologiekonzerne und unzählige Start-ups sehr viel in die einzelnen Technologien und deren Integration investieren. Allein schon der Weg in diese Richtung wird daher viele Innovationen bringen. Oder wie Matthew Ball es umschreibt: „Selbst wenn das Metaverse nicht den fantastischen Visionen der Science-Fiction-Autoren entspricht, wird es als neue Computerplattform oder Inhaltsmedium wahrscheinlich einen Wert in Billionenhöhe schaffen. Aber in seiner vollen Vision wird das Metaverse das Tor zu den meisten digitalen Erfahrungen, eine Schlüsselkomponente aller physischen Erfahrungen und die nächste große Arbeitsplattform.“ ✖

- **Edge/Cloud Computing:** No complex processes such as 3D simulations or AI models can be performed on mobile devices (VR glasses, headsets, smartphones, etc.). In order to relieve and at the same time accelerate data transmission to central servers, new methods and architectures for decentralized processing on the network (edge computing) are required and must meet all security and privacy requirements.
- **Network:** In order to handle the large amounts of data in the metaverse, reliable and intelligent network technologies such as 5G are required, which in particular maximize the quality of experience (QoE) for users.

Whether the concept of the metaverse will really work out, nobody can tell right now. In any case, it is a fact that all large technology groups and countless start-ups are investing heavily in the individual technologies and their integration. The path in this direction alone will therefore bring many innovations. Or as Matthew Ball puts it: “Even if the metaverse doesn’t live up to the fantastical visions of science fiction writers, it is likely to create trillions of dollars in value as a new computing platform or content medium. But in its full vision, the metaverse will become the gateway to most digital experiences, a key component of all physical experiences, and the next great working platform.” ✖

„KI wird definitiv bleiben“



**HELGA NOWOTNY** ist eine international anerkannte Wissenschaftsforscherin und Professorin emerita der ETH Zürich. Sie war Gründungsmitglied und Vizepräsidentin des 2007 etablierten Europäischen Forschungsrats und von 2010 bis 2013 dessen Präsidentin. Sie ist Mitglied des Rates für Forschung und Technologieentwicklung.

**Für die Wissenschaftsforscherin Helga Nowotny ist Künstliche Intelligenz (KI) ein weiterer Schub für die kulturelle Evolution des Menschen. Sie sieht in der Technologie die Chance, die unbeabsichtigten Folgen unseres Handelns früher zu erkennen und entsprechend anders zu handeln. Das Unbehagen, das viele Menschen bei KI haben, soll durch gute Beispiele – etwa aus dem Gesundheitsbereich – sowie durch eine bessere Vermittlung der Technologie vermindert werden.**

#### Wie verändert die KI unsere Gesellschaft?

Helga Nowotny: KI verändert unsere Beziehungen zueinander, die Art und Weise, wie wir miteinander kommunizieren, aber auch, wie wir uns selbst sehen und wahrnehmen – etwa hinsichtlich unserer Zeiterfahrung. KI verändert auch unsere Beziehung zur natürlichen Umwelt, weil wir mit Hilfe von digitalen Verfahren die Möglichkeit haben, einen Teil der umweltschädigenden Handlungen, die wir kontinuierlich setzen, besser hinsichtlich ihrer Folgen zu analysieren. Wir können KI einsetzen, um Schlimmeres abzuwehren. Aber KI verändert auch unser Arbeitsleben: Hier ist im Unterschied zu früheren Automatisierungswellen vor allem die Mittelklasse betroffen. Niemand weiß freilich, wie schnell neue Arbeitsplätze geschaffen werden, um die verschwindenden zu ersetzen.

“AI is definitely here to stay”

**HELGA NOWOTNY** is an internationally recognized scientific researcher and professor emerita of ETH Zurich. She was a founding member and Vice-President of the European Research Council and its President from 2010 to 2013. She is a member of the Austrian Council for Research and Technology Development.

**For science researcher Helga Nowotny, artificial intelligence (AI) is yet another booster for the cultural evolution of humankind. She sees the technology as an opportunity to identify unintended consequences of our actions early on and correct course accordingly. The discomfort that many people have about AI should be eased by pointing to good examples—for example, in healthcare—and through better communication of the technology.**

#### How is AI changing our society?

Helga Nowotny: AI is changing our relationships with another other, the way we communicate but also how we see and perceive ourselves—for example, with respect to our experience of time. Also, AI is changing our relationship with the natural environment, because with the help of digital processes we have the possibility of analyzing some of the environmentally damaging actions that we continuously take with regard to their consequences. We can use AI to prevent worse things from happening. But AI is also changing our working lives: Here, unlike in previous waves of automation, it is primarily the middle class that is affected. No one knows, of course, how quickly new jobs will be created to replace the ones that are vanishing away.

Generell trägt KI dazu bei, die kulturelle Evolution, die ja seit langem die biologische Evolution überholt hat, weiter zu beschleunigen. Mit den digitalen Artefakten, die wir erzeugt haben und weiterhin erzeugen werden, sind wir in eine ko-evolutionäre Phase eingetreten, deren Ausgang offen und somit ungewiss ist.

#### **Gibt es in Sachen Ko-Evolution mit Technologie eine historische Parallele?**

Soweit man das überhaupt rekonstruieren kann, war die Erfindung des Feuers eine Technologie, die enorme Auswirkungen hatte. Sie hat dazu beigetragen, das Stammesgefüge und die sozialen Beziehungen untereinander anders zu gestalten. Archäolog:innen denken dabei an das gemeinsame Kochen und Essen, aber auch an die Weiterentwicklung von Sprache, weil man um das Feuer herum geredet hat. Feuer hat dazu beigetragen, die Ernährungsweise zu verändern, was wiederum unserem Gehirn und unserem Körper genützt hat. Es war also der Beginn einer Ko-Evolution, von der wir sehr profitiert haben – aber auch der Beginn einer Entwicklung, in der der Homo sapiens alles andere überrannt und dabei sehr viel zerstört hat.

#### **Ist KI ihrer Meinung nach – wieder einmal – ein Hype?**

##### **Oder wird die Technologie nun bleiben?**

KI wird definitiv bleiben; wir stehen erst am Anfang ihres Potenzials. Ein Hype versucht uns immer einzureden, dass das Allerletzte auch das Allerbeste ist. Dem ist bestimmt nicht so. Die Anfänge der KI reichen mehr als 70 Jahre zurück, in dieser längeren Phase gab es auch Sackgassen, wie etwa den „KI-Winter“. Das Abheben der digitalen Technologien, wie wir sie heute kennen, hat erst vor 15, 20 Jahren so richtig begonnen – mit den neuronalen Netzwerken, auf denen Machine Learning und Deep Learning beruhen. Mindestens so wichtig war die Zunahme an Rechenleistung, cleveren Algorithmen und die enormen Datenmengen, über die wir jetzt verfügen. Doch wir müssen erst lernen, diese Daten besser zu nützen. Wir sehen auch, wo wir viel mehr daraus machen könnten. Gerade in Österreich haben wir in der Pandemie erlebt, wie schlecht wir aufgestellt sind: Die Daten sind nicht interoperabel, sind zum Teil in Silos verwahrt, können nicht brauchbar vernetzt werden usw. Das Gesundheitssystem ist jener Anwendungsbereich, wo der Nutzen von KI am einsichtigsten sein sollte – um den Menschen zu zeigen: Was bringt mir diese Technologie? Was geht mich das an? Gesundheit geht alle an!

#### **An diesem Beispiel sieht man aber auch gut das Unbehagen, das manche mit KI haben.**

Ja, aber das hängt auch damit zusammen, dass den Menschen viel zu wenig vermittelt wird, welche Vorteile es bringt, ein lernendes Gesundheitssystem zu haben. Man kann an diesem Beispiel schön zeigen: Schaut, wenn wir die Daten besser nützen und vernetzen,

Generally speaking, AI helps to further accelerate cultural evolution that has long since overtaken biological evolution. With the digital artifacts we have created and will continue to create, we have entered a co-evolutionary phase the outcome of which is open and therefore uncertain.

#### **Does this co-evolution with technology have a historical parallel?**

As far as it can be reconstructed at all, the invention of fire was a technology that had an enormous impact. It contributed to the transformation of tribal structures and social relationships. Archaeologists would refer to the act of cooking and eating together, but also to the development of language, because people were talking around the fire. Fire helped change the way we ate, which in turn benefited our brains and our bodies. So it was the beginning of a co-evolution that was much to our benefit—but also the beginning of a development in which homo sapiens overran everything else, leaving a trail of destruction.

#### **In your opinion, is AI—once again—a hype? Or is the technology here to stay?**

AI is definitely here to stay; we are only seeing the very beginnings of its potential. A hype always tries to persuade us that the very latest also is the very best. The beginnings of AI date back more than seventy years, and this prolonged phase also had its dead ends, like the “AI winter”. The takeoff of digital technologies as we know them today did not really start until fifteen or twenty years ago—with neural networks that are at the basis of Machine Learning and Deep Learning. At least as important is the increase in computing power, clever algorithms, and the enormous mass of data that are now available to us. But we will first have to learn to better utilize this data. We are also seeing where we could make a lot more of it. Especially here in Austria, we have seen during the pandemic how ill-prepared we were, with the data not being interoperable, in part siloed, and not cross-linkable in any useful way, etc. The healthcare system is the area of application where the benefits of AI should be most obvious—to show people: What do I get out of this technology? How does it affect me? Health affects everybody!

#### **However, this example also illustrates the discomfort that some have about AI.**

Yes, but that also has to do with the fact that far too few people are being made aware of the advantages of having a learning healthcare system. This is a good example to demonstrate to them: Look, if we make better use of and cross-link the data, if we look to anonymity and ethical aspects—a matter of course for science—we are able, for example, to make you aware of

wenn wir auf Anonymität und ethische Aspekte achten – und das ist von Seiten der Wissenschaft selbstverständlich –, dann können wir euch zum Beispiel auf Komorbiditäten aufmerksam machen oder die Warnzeichen für Fettleibigkeit von Kindern rechtzeitig berücksichtigen.

Das muss man den Menschen vermitteln und gleichzeitig Vertrauen schaffen. Das herrschende Unbehagen hängt stark damit zusammen, dass wir das Vertrauen in öffentliche Institutionen verloren haben – die Sozialen Medien tragen das ihre dazu bei. Dieses Vertrauen in öffentliche Institutionen und in die Wissenschaft wiederherzustellen, ist wichtig. Im Gesundheitssystem könnte man gut zeigen, dass dieses wesentlich besser funktionieren würde, wenn wir die richtigen Daten haben, sammeln und vernetzen dürfen: Dann könnte euch das Gesundheitssystem viel mehr bieten als das, was ihr jetzt bekommt. Der Vorschlag, der vom Complexity Science Hub kam und von vielen mitunterzeichnet wurde, nämlich eine unabhängige medizinische Datenstelle einzurichten, hat Hand und Fuß.

**Ein Teil des Unbehagens mit KI hängt wohl auch damit zusammen, dass diese Technologie vor allem von einer Hand voll großer Konzerne genutzt wird, um damit noch mehr Geld zu machen.**

Das stimmt leider. Die großen Player scheffeln damit sehr viel Geld. Und für Europa besonders ärgerlich ist, dass sie nicht einmal Steuern dafür bezahlen – was eigentlich eine Selbstverständlichkeit sein sollte. Geld zu verdienen ist ja an sich nichts Schlechtes, aber man sollte seine Steuern bezahlen. Dennoch sind wir nicht nur Opfer: Wir liefern freiwillig unsere Daten an die Konzerne, weil wir Gegenleistungen erhalten, von denen wir uns ein angenehmeres Leben versprechen. Wir vergessen dabei: Wenn du etwas gratis erhältst, bist du selbst das Produkt.

**Das Gesundheitswesen ist auch ein gutes Beispiel für sogenannte „predictive algorithms“, die Sie in Ihrem jüngsten Buch „In AI we trust“ thematisieren ...**

„Predictive algorithms“ ermöglichen es uns, aufgrund der Extrapolation aus der Vergangenheit weiter in die Zukunft zu sehen. Das beruht auf der Erfahrung, dass es sehr wahrscheinlich ist, dass gewisse Strukturen und Phänomene von heute auch morgen noch da sein werden. Dabei wird übersehen, dass die Zukunft ungewiss ist – und dass kein Algorithmus die Zukunft kennt. Überraschungen kann es immer geben. Wir leben gerade jetzt in einer Periode, die wir uns vor zwei Jahren nicht vorgestellt haben: Niemand dachte, dass wir so eine Pandemie mit allen Folgeerscheinungen, bei denen wir erst am Anfang stehen, erleben. Dazu kommen der Krieg in der Ukraine, die Inflation und vieles mehr.

In dieser Situation lassen sich Risiken nicht mehr einzeln betrachten. Diese Sichtweise geht auf das 19. Jahrhundert und die

comorbidities or pay timely attention to the warning signs of child obesity.

This is what you have to get across to people while trying to build trust. The overriding discomfort is very much related to the fact that we have lost trust in public institutions—which social media really had its part in. Restoring trust in public institutions and in science is important. In the health-care system, it could be well demonstrated that it would work a lot better if we were allowed to have, collect, and link the right data: The healthcare system could then offer you a lot more than what you get now. The proposal that came from the Complexity Science Hub and was co-signed by many, namely, to establish an independent medical data center, really does make sense.

**Part of the discomfort about AI is probably due to the fact that the technology is primarily used by a handful of big corporations to make even more money.**

Unfortunately, that's true. The big players are raking in a lot of money from it. And what is particularly annoying for Europe is that they don't even pay taxes on it—which should in fact be a matter of course. Making money is not a bad thing per se, but you should pay your fair share of taxes. Nevertheless, we are not just victims here: we voluntarily supply our data to the corporations because we receive something in return, which we hope will make our lives more pleasant. What we forget is that when you get something for free the real product is you.

**Healthcare is also a good example of so-called “predictive algorithms”, which you address in your recent book “In AI We Trust” ...**

“Predictive algorithms” allow us to see further into the future based on extrapolation from the past. At its basis is the experience that if there is a high probability that certain structures and phenomena that exist today will still be here tomorrow. This overlooks the fact, though, that the future is uncertain—and that no algorithm knows the future. There can always be surprises. Right now, we are living through a period that we couldn't have imagined two years ago: No one thought we would experience such a pandemic with all its aftermath, where we are now only seeing the beginning. Add to that the war in Ukraine, inflation and much more.

In this situation, risks can no longer be considered individually. This is a view that dates back to the 19th century and the invention of insurance based on calculations of probability of occurrence multiplied by extent of damage. Today, we have risks that occur in clusters and cascades: We see that the war in Ukraine, along with China's crackdown on COVID-19, has massively disrupted supply chains. As a result,



Erfindung von Versicherungen zurück, die auf der Berechnung von Eintrittswahrscheinlichkeit multipliziert mit dem Ausmaß des Schadens beruhen. Wir haben heute Risiken, die als Cluster und als Kaskaden auftreten: Wir sehen, dass der Krieg in der Ukraine zusammen mit den scharfen Maßnahmen gegen COVID-19 in China die Lieferketten massiv gestört hat. In Nordafrika fehlt es dadurch an Nahrung, wir haben keine Mikrochips usw. Dieses Ineinander-verbundene von Risiken haben wir uns in dieser Weise nicht vorgestellt.

Man muss den Anwendungsbereich von „predictive algorithms“ kennen und definieren – und man muss wissen, dass es sich letzten Endes nur um Wahrscheinlichkeitsaussagen handelt. Wir Menschen und die Gesellschaft können aber nur schlecht mit Wahrscheinlichkeiten und mit Ungewissheit umgehen. Wenn wir blind daran glauben, dass die von einem Algorithmus bestimmte Vorhersage eintreten wird, verhalten wir uns entsprechend. So kommt es zu „self-fulfilling prophecies“.

#### Wie helfen uns Algorithmen dabei, mit diesen Krisen umzugehen?

Sie helfen uns dabei, aufzuzeigen, was passieren könnte. Es geht um „what – if“. Die Algorithmen zeigen uns, was passiert, wenn man diese oder jene Entscheidung trifft. Sie erweitern unseren Vorstellungsräum und die Möglichkeiten, die es dort gibt. Dadurch wird auch neues Bewusstsein für die Offenheit der Zukunft geschaffen. Wir dürfen uns von der Ungewissheit nicht einschüchtern lassen – es gäbe dann nur den Rückzug in Angst, die schlechteste Lösung. Wir müssen lernen, uns mit diesen What-if-Szenarien auseinanderzusetzen und zu überlegen, was wir wollen – und was nicht. Dann müssen wir freilich danach handeln. Es geht in erster Linie um Orientierungswissen, das man dann auf einzelne Ziele herunterbrechen kann. Wir brauchen mehr Orientierungswissen als eine Art von Leitschiene, die auch längerfristiges Denken ermöglicht. Derzeit wissen wir zwar viel, aber wir tun das Gegenteil, weil wir zu sehr im kurzfristigen Denken der Gegenwart verfangen sind.

#### Ein Vorbehalt gegen KI ist, dass sie eine Art „black box“ ist, in die man nicht hineinschauen kann, wie sie zu Entscheidungen kommt. Wie wichtig ist Ihrer Meinung nach „explainability“?

Es klingt schön, es ist aber technisch unheimlich schwer umzusetzen. Letztendlich werden wohl neue Algorithmen eingesetzt werden, um die „explainability“ zu liefern. Wir müssen uns davon verabschieden, zu glauben, dass es DIE EINE Lösung gibt, um KI transparenter, verstehbarer, erklärbarer zu machen. Stattdessen muss man lernen: Wo gibt es Intervention Points? Wo gibt es die Möglichkeit, im Design einzugreifen? Wo gibt es Möglichkeiten, bei den vielfältigen Kontexten der Anwendung einzugreifen? Man muss sich auf die sehr komplexen KI-Systeme als Systeme einlassen.

North Africa has a food shortage, we don't have microchips, etc. We have not envisaged risks being interwoven in this way. You have to know and define the scope applicability of “predictive algorithms”—and know that in the end they are only probability assessments. But we humans and society at large are bad at dealing with probabilities and with uncertainty. If we blindly believe that the prediction informed by an algorithm will come true we behave accordingly. That leads to “self-fulfilling prophecies”.

#### How can algorithms help us deal with such crises?

They help us point out what could happen. It's all about “what if”. Algorithms show us what happens when you make this decision or that. They expand the scope of our imagination as well as the possibilities that are there. This also creates a new awareness of the openness of the future. We must not be daunted by uncertainty—this would only lead to retreat in fear, the worst option. We must learn to face up to these what-if scenarios and think about what we want—and what we don't. And then, of course, we must act accordingly. It's basically a matter of orientational knowledge, which can then be broken down to individual objectives. We need more orientational knowledge as a kind of guideline that also enables longer-term thinking. Currently, we know a lot, but we do the opposite because we are too caught up in present-day short-term thinking.

#### One misgiving about AI is that it is a kind of “black box” that you cannot really look into to see how it arrives at decisions. How important, in your opinion, is “explainability”?

Sounds nice but is incredibly difficult to implement, technically. Eventually, it will probably take new algorithms to provide for “explainability”. We need to move away from the belief that there is ONE solution to make AI more transparent, more comprehensible, more explainable. Instead, we need to learn: Where are intervention points? Where is there a possibility to interfere in the design? Where are possibilities to interfere in the various different contexts of application? You have to engage with those very complex AI systems as systems. Technical possibilities and obstacles are ultimately inextricable from the humanistic guidelines they are supposed to serve. In the end, “black box” also is just a metaphor for one part of this complex architecture. In past times, an automobile also was a “black box” to people who'd never had anything to do with one. You could open the hood, but usually didn't know what to do next. However, there were always people who were able to repair a car themselves. Today, that's impossible. Even if

Die technischen Möglichkeiten und Hindernisse sind letztlich nicht von den humanistischen Leitlinien zu trennen, denen sie dienen sollten. „Black box“ ist letztlich auch nur eine Metapher für einen Teil dieser komplexen Architektur. Früher war ein Automobil für Menschen, die nie mit einem zu tun hatten, auch eine „black box“. Man konnte zwar die Kühlerhaube öffnen, wusste aber meist nicht weiter. Es gab allerdings immer Menschen, die ein Auto selbst reparieren konnten. Heute ist das unmöglich. Selbst wenn Sie ein geschickter Bastler sind: Bei einem jetzigen Auto können Sie nichts selbst reparieren – selbst wenn Sie wüssten, was. Das ist ein qualitativer Sprung in unserem Umgang mit Technologie. Mit der KI kommt noch eine Größenordnung dazu. Sie braucht eine enorme Infrastruktur, Energie, Mikrochips, Datenspeicher – also eine technische und soziale Ordnung, die das alles erst ermöglicht. Doch letzten Endes sind wir selbst Teile dieser Ordnung, wir ermöglichen sie, wir gestalten sie mit.

**In Ihrem neuen Buch verwenden Sie sehr stark den Begriff „mirror world“. Was meinen Sie damit genau?**

Wir bilden die reale Welt dank der unzähligen Sensoren, die wir installiert haben, immer stärker in der digitalen Welt ab. Diese Spiegelwelt ist zum Teil sehr nützlich, man kann zum Beispiel, wie es die NASA seit Jahrzehnten macht, Maschinen im Weltraum zunächst am digitalen Zwilling reparieren. Operationen lassen sich leichter planen, Simulationen werden konkretisiert. Ein Spiegel ist aber auch eine Metapher: Was sehen wir, wenn wir uns im Spiegel anschauen? Er gibt uns die Möglichkeit, über uns selbst nachzudenken – und darüber, wie die Technik unser Verhältnis zu uns selbst verändert.

**Geht es nach den Vorstellungen der Industrie, soll das „Metaverse“ eine umfassende Spiegelwelt werden. Viele Menschen, mich eingeschlossen, sehen das mit großer Sorge.**

Das ist derzeit sehr viel Hype. Ich habe vor mehr als 15 Jahren an der ETH Zürich ein Experiment gesehen, bei dem sich zwei Forschungsgruppen, die in verschiedenen Laboren sitzen, miteinander so vernetzen, dass sie nicht nur miteinander sprechen, sondern sich so verhalten, als wären sie alle im selben Labor. Das ist sehr nützlich. In Singapur habe ich vor Jahren Demonstrationen gesehen, wie man mit einem Avatar Basketball trainieren kann. Das sind großartige Dinge!

Das Unbehagen kommt daher, dass das jetzt kommerzialisiert wird. Die Konzerne verkaufen uns Dinge, die wir angehalten sind, als Paket zu kaufen. Sie setzen die Bedingungen, unter denen wir zustimmen – ansonsten bekommen wir deren Leistungen nicht –, und nutzen das aus, um uns noch mehr abhängig davon zu machen. Alternativen, aber auch Wettbewerb, werden dadurch

you are a gifted do-it-yourselfer: in a modern-day car, you can't fix anything yourself—even if you knew what. That's a qualitative leap in our approach to technology. AI adds a whole different dimension. It needs an enormous infrastructure, energy, microchips, data storage—in other words, a technological and social system that enables it all in the first place. But in the end, we ourselves are part of these systems, we enable them, we help shape them.

**In your new book, you keep emphasizing the term “mirror world”. What exactly do you mean by that?**

Thanks to the countless sensors we have installed, we can increasingly image the real world in the digital world. This mirror world is, in part, very useful to have; for example, you can, as NASA has been doing for decades, make repairs on machines in space on the digital twin first. Surgery is plannable more easily, simulations can be made more concrete. But a mirror is also a metaphor: Looking at ourselves in the mirror, what do we see? It gives us an opportunity to think about ourselves—and about how technology is changing our relationship with ourselves.

**If the industry has its way, the “Metaverse” is set to become an encompassing mirror world. Many people, myself included, see this with great concern.**

There is a lot of hype about that right now. I saw an experiment more than 15 years ago at ETH Zurich where two research groups sitting in different labs were networked with one another so that they not only talked to each other, but acted as if they were all in the same lab. That's very useful. Years ago in Singapore, I saw demonstrations of basketball training with an avatar. These are great things!

The discomfort comes from the fact that this is now being commercialized. Corporations are selling us things that we are being urged to buy as a package. They define the conditions that we have to agree to—we won't be able to get their services otherwise—and use that to make us even more dependent on them. It makes alternatives, as well as competition, an impossibility. I think it's great to be able to use this technology for science, for music, for sports. But I don't want the metaverse to dictate what I have to buy and what I should do with it. It is reducing us to being digitally incapacitated.

**How can we escape this situation?**

The only alternative I see is Open Source and its further development. In science, for example, I can imagine using Open Source and Open Metaverse for own purposes. We have to protect certain areas against commercialization. Otherwise,

unmöglich gemacht. Dass man diese Technologie für die Wissenschaft, für Musik, für Sport nutzen kann, finde ich großartig. Aber ich möchte nicht, dass mir das Metaverse diktiert, was ich zu kaufen habe und was ich damit machen soll. Wir werden dadurch in eine digitale Unmündigkeit zurückversetzt.

#### **Wie können wir dieser Situation entgegengehen?**

Die einzige Alternative, die ich sehe, ist Open Source und deren Weiterentwicklung. Für die Wissenschaft kann ich mir zum Beispiel vorstellen, dass man Open Source und Open Metaverse für die eigenen Zwecke nützt. Wir müssen gewisse Bereiche vor einer Kommerzialisierung schützen. Andernfalls haben wir Konsum hoch drei, die Menschen sitzen dann nur mehr vor dem Bildschirm, die passive Konsumentenhaltung, die wir ja häufig beklagen, wird potenziert. Dem muss man entgegenwirken: durch Erziehung, durch Aufklärung im besten Sinn des Wortes und durch Empowerment der Jugendlichen und das Aufzeigen und Umsetzen von Alternativen.

#### **Wie könnte man den Menschen besser erklären, was bei den Anwendungen von KI geschieht?**

Wir müssen Wissenschaft besser kommunizieren. Lange Zeit gab es ein großes Versäumnis. Man hat viel zu sehr versucht zu erklären, was Wissenschaft als Endprodukt für die Gesellschaft leistet. Man hat beeindruckende Ergebnisse, neue Technologien, innovative Medikamente usw. kommuniziert – aber viel zu wenig, wie Wissenschaft tatsächlich funktioniert. Wissenschaft ist ein Prozess, und die Grundlagenforschung ist immer eine Reise ins Ungewisse. Wenn man weiß, was herauskommt, ist es keine Grundlagenforschung mehr. Auf dieser Reise kann viel schiefgehen, die Wissenschaftler:innen machen trotzdem weiter, weil sie glauben, dass etwas Sinnvolles dabei herauskommt. Das muss man vermitteln.

Dafür muss man sich neue Formate einfallen lassen. In Göttingen zum Beispiel gab es eine Vereinbarung zwischen dem Max-Planck-Institut und einem Gymnasium. Nachdem man die Lehrer:innen überzeugt hatte, sind die Schüler:innen in die Labore des MPI gegangen und haben dort ihren Unterricht mit neuestem Wissen bekommen. Solche Partnerschaften müssen aufgebaut werden.

Man braucht auch neue Arten von Feedbackschleifen. Ein Beispiel: Bei der Erforschung seltener Krankheiten haben sich die betroffenen Familien zusammengetan und das Wissen, das sie über die Krankheit haben, an die Wissenschaft weitergegeben. Vielleicht sollte es neue Jugendclubs geben, die sich dem aktiven Lernen von dem widmen, was die Jugendlichen antreibt. Oder beim Sport: Auch dort kann man sehr viele wissenschaftliche Themen und den Umgang damit anhängen.

we will have consumerism running riot, with people only sitting in front of the screen and that passive consumer attitude that we often complain about multiplying. This calls for counteracting: through education, through enlightenment in the best sense of the word, and through empowerment of young people as well as through indication and implementation of alternatives.

#### **How could we better explain to people what AI applications do?**

We need to better communicate science. For a long time, there was one major shortcoming. Far too much effort was spent on explaining what science as a final product does for society, on communicating impressive results, new technologies, innovative medications, etc.—but far too little on how science actually works. Science is a process, and basic research is always a journey into the unknown. Once you know the outcome it's no longer basic research. A lot can go wrong on this journey, but the scientists keep on going because they believe that something useful will come out of it. That is what you need to communicate.

To do so, you have to come up with new formats. In Göttingen, for example, there was an agreement between the Max Planck Institute and a high school. Once their teachers were convinced, the students went to the MPI lab for classes teaching cutting-edge knowledge. That's the kind of partnerships that need to be built.

Also needed are new types of feedback loops. One example: in rare disease research, families affected have come together and shared what they know about the disease with the scientific community. Perhaps there should be new youth clubs dedicated to actively learning about what drives young people. Or in sports: there, too, you can bring in all sorts of science issues and how to deal with them.

**Wenn man über KI diskutiert, gibt es rasch zwei Lager: Die einen sehen in der Technologie eine positive Utopie, die viele unserer Probleme löst; die anderen sehen eine Dystopie, ein Katastrophenszenario. Wie ist Ihre Meinung dazu?**

Man darf weder in die eine noch in die andere Falle tappen. Die Dystopie führt dazu, dass man letzten Endes depressiv wird. Und die Nerd-Seite, die glaubt, mit Technik allein könnten wir alle Probleme lösen, führt auch in eine Sackgasse. Es ist nie die Technik allein: Mit Technik können wir einige Probleme lösen, aber sie ist eingebettet in soziales Verhalten, in soziale Zusammenhänge, in Organisationen, in rechtliche und ethische Rahmen usw. Das alles muss man mitdenken und mitgestalten, wenn man etwas verändern will.

**In welchem Bereich würden Sie persönlich sich rasche Fortschritte mit KI wünschen?**

Ich würde mir wünschen, dass wir kausale Zusammenhänge besser verstehen. Und dass wir dadurch lernen, die unbeabsichtigten Folgen unseres Handelns früher zu erkennen und entsprechend anders zu handeln. ✕

**Whenever AI is discussed, there are two camps that quickly emerge: those who see in technology a positive utopia that solves many of our problems, and the others who see a dystopia, a scenario for disaster. What is your opinion on this?**

You can't fall into either trap. Dystopianism ends up in depression. And the nerd side, which believes we can solve all problems with technology alone, also leads to a dead end. It's never technology alone: we can solve some problems with technology, but it's always embedded in social behaviors, in social contexts, in organizations, in legal and ethical frameworks, etc. All of this needs to be considered and involved if you want to bring about change.

**In what area would you personally most hope for swift progress in AI?**

I would like us to better understand causal relationships. And that we learn from them to identify unintended consequences of our actions earlier and choose a different path accordingly. ✕

Praktische  
Anwendungen  
Practical  
applications



„Ohne spezifisches Domänen-Wissen hat man kaum Chancen auf Erfolg“

**Für Andreas Kugi sind Methoden der Künstlichen Intelligenz (KI) ein weiteres Werkzeug in der Toolbox der Automatisierungsmethoden: Sie ermöglichen es beispielsweise, dass Roboter ihre Umgebung erkennen und Szenarien verstehen – als Voraussetzung, dass sie intelligent reagieren können.**

**Sie beschäftigen sich mit dem Forschungsthema „AI-enabled automation and robotics“. Wo kommt bei der Automatisierung KI ins Spiel?**

**Andreas Kugi:** Die klassische Automatisierungstechnik hatte folgende Logik: Man erfasst etwas messtechnisch mit Sensoren, die Daten werden von Algorithmen verarbeitet und damit werden dann Aktoren angesteuert, die das System gezielt beeinflussen. In modernen Automatisierungssystemen erweitert man dieses Konzept durch Kombination unterschiedlicher Sensormodalitäten mit daten- und physikalisch basierten Modellen und setzt zunehmend bildgebende Verfahren ein, die wesentlich mehr Informationen beinhalten – und das in Echtzeit. Das dient im Wesentlichen dazu, ganze Situationen erfassen zu können und in weiterer Folge darauf aufbauend Entscheidungen zu treffen, wie das Automatisierungssystem gezielt in das System eingreifen soll. Dabei spielt KI eine entscheidende Rolle. Die große Herausforderung besteht darin, wie man diese Daten interpretieren kann. Dazu kombiniert man Methoden der Sensorfusion mit maschinellem Lernen und a-priori-Wissen in Form von physikalischen Modellen oder semantischer Information. Ein Beispiel: Wenn man beim autonomen Fahren die Szene eines Kamerabildes interpretieren möchte und über die Objektdetek-



**ANDREAS KUGI** ist Vorstand des Institutes für Automatisierungs- und Regelungstechnik (ACIN), Professor für komplexe dynamische Systeme an der Technischen Universität Wien sowie Co-Leiter des „Center for Vision, Automation & Control“ am AIT Austrian Institute of Technology. Seine Hauptinteressen in Forschung und Lehre liegen im Bereich der physikalisch basierten mathematischen Modellierung, der Systemtheorie, des nichtlinearen Regler- und Beobachterentwurfes, des mechatronischen Systementwurfes, der Echtzeitoptimierung sowie in der Robotik und Prozessautomatisierung. Er ist wirkliches Mitglied der Österreichischen Akademie der Wissenschaften (ÖAW) und Mitglied im Expertengremium der Deutschen Exzellenzinitiative.

Practical applications

“Without specific domain knowledge, there’s little chance of success”

**ANDREAS KUGI** is the Head of the Automation and Control Institute (ACIN) and professor for complex dynamic systems at the TU Wien as well as Co-Head of the Center for Vision, Automation & Control at the AIT Austrian Institute of Technology. His main interests in research and teaching include the areas of physically based mathematical modelling, systems theory, non-linear controller and observer design, mechatronic system design, real-time optimization, robotics and process automation. He is a full member of the Austrian Academy of Sciences (ÖAW) and a member of the panel of experts for the German Excellence Initiative.

**For Andreas Kugi, the methods of AI are just another tool in the toolbox of automation, enabling, for example, robots to identify their environment and understand scenarios – a prerequisite for them to respond intelligently.**

**You work on the subject of “AI-enabled automation and robotics” in your research. Where does AI come into play in automation?**

Andreas Kugi: Classical automation technology had the following logic: you record something using sensors and measurement technology, the data is processed by algorithms, and then actuators are activated to bring a targeted influence to bear on the system. In modern automation systems, this concept is expanded by combining different sensor modalities with data and physics-based models and by increased use of imaging processes that contain substantially more information – and in real time. It essentially serves to enable comprehending the whole of a situation and subsequently deciding, on that basis, on how the automation system will be set up to interfere in the system. AI plays a crucial role in this. The big challenge is how to interpret this data. To do so, you combine methods of sensor fusion with machine learning and a-priori knowledge in the form of physical models or semantic information. To give an example: If in autonomous driving the scene of a camera image is to be interpreted and the object detection function identifies a tree, you then know that, unlike other road users, that tree

Praktische Anwendungen

tion wird ein Baum erkannt, dann weiß man, dass sich der Baum im Gegensatz zu anderen Verkehrsteilnehmer:innen nicht bewegen kann. Dieses Wissen wird dann systematisch für die Vorhersage von möglichen Kollisionen und der Fahrsituation sowie die Steuerung des Fahrzeugs herangezogen.

Dieses Erfassen von ganzen Situationen ist wesentlich mehr als das Erfassen eines einzelnen sensorischen Signals. Durch die Rechenpower, die man heute zur Verfügung hat, kann man dann diese Informationen nutzen, um das Systemverhalten zu optimieren, Ressourcen zu sparen und Entscheidungen zu treffen, was im nächsten Schritt zu tun ist. Auch das ist eine Rückkopplungsschleife, aber auf einer höheren Ebene. Wenn wir also von „AI-enabled automation“ sprechen, dann nutzt man nach wie vor die klassischen Elemente der Automatisierung, fügt aber eine zusätzliche Ebene der Steuerung, die ein weit höheres Maß an kognitiven Fähigkeiten besitzt, hinzu. Dieses Konzept kennt man aus dem Bereich des autonomen Fahrens, doch setzen wir diese Prinzipien auch zunehmend für die Automatisierung von Produktionsanlagen ein, beispielsweise zur Steigerung der Ressourcen-Effizienz und zur Verringerung von Qualitätsdefekten.

#### **Können Sie uns ein Beispiel dafür geben?**

In unserem Leuchtturm-Projekt am AIT Center for Vision, Automation & Control beschäftigen wir uns unter anderem mit autonomen Arbeitsmaschinen – wobei wir uns in einem ersten Schritt auf intelligente Assistenz-Funktionen und autonome Teilaufgaben fokussieren. Als repräsentatives Szenario betrachten wir autonome Lade- und Entladevorgänge von Paletten, Baumstämmen, Schüttgut etc. Spannend dabei ist, dass man auf der einen Seite Kollisionen der beteiligten Maschinen mit anderen Objekten oder dem Menschen vermeiden muss, aber auf der anderen Seite muss die Maschine in Kontakt mit Gegenständen treten, wenn man etwas greifen oder bewegen möchte. Dazu muss das System wissen: Das ist das Objekt, das ich greifen möchte und das ist ein Objekt, mit dem ich nicht in Kontakt kommen darf. Und dazu braucht das System ein tiefgehendes Verständnis der Szenen und dabei setzen wir massiv auf Methoden des maschinellen Lernens. Zum Beispiel für das Manipulieren von Baumstämmen. Wir arbeiten hier mit Convolutional Neural Networks (CNN) und nutzen die Methode des sogenannten „geometric cue learning“. Dabei wird das a-priori-Wissen der Geometrie eines Objekts beim Lernen systematisch berücksichtigt. Beispielsweise bei einem Baumstamm weiß man, dass es in erster Näherung ein zylindrisches Objekt ist. Mit Hilfe von 3D-Grafik-Programmen werden nun geeignete synthetische Daten generiert und damit die neuronalen Netze trainiert. Die Erfolgsquote bei realen Daten ist dabei extrem

is not a moving object. This knowledge is then systematically used to predict possible collisions, the driving situation, and the steering of the vehicle.

Comprehension of the whole of a situation is a lot more than capturing a single sensor signal. With the computing power available today, the information can then be used to optimize system behavior, save resources, and make decisions on what to do in the next step. This, too, is a feedback loop, but on a higher level. So, if we speak of “AI-enabled automation”, it is still about using the classical elements of automation but inserting an additional level of control with a far greater extent of cognitive capability. This is a concept known from autonomous driving, but we are also increasingly applying these principles to the automation of production facilities, for example to improve resource efficiency and reduce quality defects.

#### **Can you give us an example?**

In our lighthouse project at the AIT Center for Vision, Automation & Control, we are dealing, among other things, with autonomous labor machines – focusing in a first step on intelligent assistance functions and autonomous subtasks. One typical scenario we are looking at is autonomous loading and unloading processes of pallets, logs, bulk goods, etc. The exciting thing here is that, on the one hand, you have to avoid collisions between the machinery involved and other objects or humans, but on the other hand, the machine has to get in physical contact with objects if you want to grab or move something. To do so, the system must know: this is the object I want to grab – and that is an object I must not come in contact with. The system therefore needs a deep understanding of the scenes, and for this we massively rely on machine learning. For example, the handling of tree trunks. We work here with Convolutional Neural Networks (CNN) here and use the method of so-called “geometric cue learning”. It means that a-priori knowledge of an object’s geometry is systematically taken into account during learning. A tree trunk, for example, is known, in first approximation, to be a cylindrical object. Now suitable synthetic data is generated with the help of 3D graphics programs to train the neural networks. The success rate with real data is extremely high, even though the classifier has never seen a real image before. Moreover, this method is distinguished by the fact that its results are robust with respect to the texture and surface characteristics of the objects under consideration.

hoch, obwohl der Klassifikator zuvor niemals ein reales Bild gesehen hat. Darüber hinaus zeichnet sich diese Methodik dadurch aus, dass die Ergebnisse robust gegenüber der Textur und der Oberflächenbeschaffenheit der betrachteten Objekte sind.

#### **Das heißt: Sie nutzen in diesen Fällen a-priori-Wissen, um gute Trainingsdaten zu produzieren?**

Ja. Für maschinelles Lernen braucht man qualitativ hochwertige Daten, die gelabelt werden müssen. Das ist insbesondere für 3D-Daten ein Riesenaufwand, weshalb wir dies rein synthetisch machen. Dies funktioniert bei menschlich gemachten Objekten (Paletten, Container, etc.) besonders gut, ist aber auch für Objekte mit typischer Geometrie, wie beispielsweise Baumstämme, sehr gut anwendbar.

#### **Welche weiteren Schritte sind nötig, wenn eine Maschine einmal die Umgebung erkannt hat?**

Nehmen wir als Beispiel einen klassischen Industrieroboter, der die Aufgabe hat, eine Schraube an einer ganz bestimmten Position zu entnehmen, an eine andere festgelegte Position zu bringen und dort die Schraube hineinzuschrauben. Würde man das Objekt um einen Zentimeter verschieben, würde dieser Roboter die Schraube um einen Zentimeter falsch hineinschrauben. Dies ist bei einem Großteil der eingesetzten Systeme im Stand der Technik der Fall. Wenn der Roboter mit Kamerasystemen ausgestattet ist und man mit Algorithmen in der Lage ist, eine Schraube und das passende Schraubloch zu erkennen, so kann man die Schraube aus einer Box entnehmen, das zugehörige Loch suchen und den Schraubvorgang ausführen. Dies erfordert bereits ein gewisses – noch immer sehr eingeschränktes – Verständnis des Prozesses, ist aber mit heutigen Methoden umsetzbar, wenn auch die Ausführungsgeschwindigkeit der Roboter in der Regel noch zu langsam ist. Aber so richtig intelligent ist das auch noch nicht. Insbesondere wenn man in Richtung der automatisierten Herstellung kleiner Stückzahlen mit einer großen Variantenvielfalt denkt, dann braucht man ein tiefergehendes Wissen über den Prozess und die zur Verfügung stehenden Werkzeuge sowie ein umfangreicheres Szenenverständnis. Wenn man beispielsweise einem Menschen einen Schraubenzieher hinlegt, dann weiß er oder sie, dass das ein Schraubenzieher ist und was man mit diesem Werkzeug alles machen kann: Man kann eine Schraube reindrehen, aber man kann es unter anderem auch dazu verwenden, den Deckel einer Lackdose zu öffnen. Dieses Wissen hat der Roboter in der Regel nicht. Er weiß a priori nicht, was er mit dem Werkzeug alles anfangen kann. Wir Menschen haben dieses Verständnis: Wenn wir eine Lackdose und einen Schraubenzieher sehen und die Aufgabe „Lackdose öffnen“ bekommen, dann wissen wir, was zu tun ist.

#### **That means: You use a-priori knowledge in those cases in order to produce good training data?**

Yes. For machine learning, you need high-quality data that has to be labeled as well. It's a huge effort, especially with 3D data, which is why we do it in a purely synthetic way. This works particularly well for human-made objects (pallets, containers, etc.), but is also highly applicable to objects with a typical geometry, such as tree trunks.

#### **What other steps are needed once the machine has identified the environment?**

Let's take for example a classic industrial robot, which has the task of picking up a screw from one specific place, moving it to another predefined place and screwing it in there. If the object were to be moved one centimeter, the robot would screw in the screw one centimeter off from the correct position. This is the case with most of the state-of-the-art systems currently in use. If such a robot is equipped with camera systems and enabled by algorithms to identify a screw and the matching screw hole, it becomes possible to pick up a screw from a box, find the matching screw hole and carry out the process of screwing in the screw. This already requires a certain – still very limited – understanding of the process, but is implementable with today's methods, even if robotic execution speed generally is still too slow. But that is not yet actually intelligent either. Particularly if you start thinking in the direction of automated production of small series with a large number of variants, you need deeper knowledge of the process and the tools available as well as a more comprehensive scene understanding. For example, if you put a screwdriver in front of a human, they will know that this is a screwdriver and what you can do with it as a tool: You can screw in a screw, but you can also use it to pry the lid off a paint can, and other things. The robot, as a rule, does not have that knowledge. It does not know a-priori what it can do with the tool. We humans have that understanding: If we see a can of paint and a screwdriver and are given the task 'Open paint can', we know what to do.

## Wie kann man dem Roboter dieses Wissen geben und wo wird dies gebraucht?

Das ist die große Frage, an der weltweit geforscht wird. Im Rahmen unseres Leuchtturm-Projekts am AIT bzw. bei Profactor in Kooperation mit der TU Wien konzentrieren wir uns in einem ersten Schritt auf bestimmte Reparaturvorgänge – konkret auf die Reparatur von Freiform-Oberflächen, wie etwa Polieren, Lackieren oder das Auftragen einer Paste. Wir wollen anhand dieses Beispiels einmal den gesamten Prozess durchspielen. Dabei beobachten wir auch, wie Menschen diese Reparaturvorgänge vornehmen, und wollen davon lernen. Wir verwenden dazu neben Kamerasystemen auch instrumentierte Werkzeuge, mit denen wir Bewegungen und Kräfte messen können. Daraus extrahieren wir Informationen, wie man das Werkzeug an der Oberfläche führen muss, um eine bestimmte Oberflächenstruktur zu erzielen. Dies wird anhand einer gewissen Anzahl von Objekten demonstriert, und dann soll der Roboter mit Hilfe von Transfer-Learning-Methoden in die Lage versetzt werden, die Reparaturaufgaben auch an Objekten vorzunehmen, die er zuvor so noch nie gesehen hat.

Dahinter steht eine klare Vision: Bisher wurde Robotik und Automatisierung primär dafür eingesetzt, dass man Produkte sehr schnell und effizient zusammenbaut. In Hinblick auf die notwendige Steigerung der Ressourceneffizienz und Nachhaltigkeit wird sich in Zukunft die Frage stellen, ob man Produkte automatisiert reparieren und effizient und werthaltig zerlegen kann. Bei solchen Prozessen ist die Variantenvielfalt vom Prinzip her extrem hoch und dies erfordert neue Automatisierungsarchitekturen und robotische Lösungen. Natürlich wird dies nur dann umsetzbar sein, wenn man die Möglichkeit des automatisierten Zerlegens und Reparierens bereits im Produktdesign vorsieht. Dazu kommt, dass viele KMUs Roboter als intelligente Werkzeuge einsetzen könnten – etwa ein Tischlereibetrieb beim Schleifen von Oberflächen.

## Wie wichtig ist „explainability“ – also die Nachvollziehbarkeit von Entscheidungen der KI – im industriellen Kontext?

Die Europäische Union hat hier einen risikobasierten Ansatz gewählt, was ich als sehr vernünftig erachte. Man muss jedoch zwischen verschiedenen Anwendungsfällen unterscheiden: Wenn man zum Beispiel eine KI für eine Produktionsanlage entwickelt, die im Mittel die Hälfte des Wasserverbrauchs einspart, dann denke ich nicht, dass es notwendig ist, genau zu wissen, warum die Maschineneinstellung zu jedem Zeitpunkt so gewählt wurde, wenn das Ergebnis im Schnitt passt. Es ist auch keine Katastrophe, wenn dies in Ausnahmefällen dann einmal nicht der Fall ist. Hier braucht man meiner Einschätzung nach keine Erklärbarkeit und Beweisbarkeit der Methodik – hier reicht es aus, dass es unterm Strich eine Verbesserung gibt. Auch bei herkömmlichen Automatisierungssys-

## How can this knowledge be imparted to the robot and where is this needed?

That is the big question and the subject of research worldwide. In our lighthouse project at AIT, or with Profactor in cooperation with the Vienna University of Technology, we are focusing in a first step on certain repair processes – specifically, on repairs of freeform surfaces, such as polishing, painting, or the application of a paste. We want to use this as an example to go through the entire process once. And we also observe how people go about performing such repairs and learn from them. Aside from camera systems, we also use instrumented tools with which we can measure movements and forces applied. From this, we extract information on how to move a tool over a surface to achieve a certain surface structure. This is demonstrated on a certain number of objects, and then the robot is to be enabled, through transfer learning methods, to perform repair tasks on objects it has never seen before.

There is a clear vision behind this: until now, robotics and automation have primarily been used for high-speed and efficient product assembly. With respect to the necessary increases in resource efficiency and sustainability, the question in the future will be, if automation be used to repair products and disassemble them efficiently and with value retention? In such processes, there is an extremely wide variety of variants, at least in principle, and this requires new automation architectures and robotic solutions. That, of course, will only be feasible if the possibility of automated disassembly and repair is already provided for in the product design. Also, many small and medium-sized businesses might be able to use robots as intelligent tools – a joiner's workshop, for example, to smooth and polish surfaces.

## How important is “explainability” – that is, the traceability of AI decisions – in an industrial context?

The European Union has chosen a risk-based approach here, which I consider very reasonable. However, you have to distinguish between different use scenarios: If, for example, you develop an AI for a production plant that saves half the water consumption on average, I don't think it's necessary to know exactly at every given point in time why the machine setting was chosen the way it was if the overall result is, on average, okay. And it's not a catastrophe, either, if, as an exception, this is not the case. In my opinion, there is no need for explainability and provability of the method here – it is enough to see that, on balance, there is an improvement. Even with conventional automation systems, you usually do not have systematic proof of stability, particularly when nonlinear and adaptive algorithms are used. Of course, however, you



temen hat man in der Regel keinen systematischen Stabilitätsbeweis, insbesondere wenn nichtlineare und adaptive Algorithmen zum Einsatz kommen. Natürlich müssen aber immer Maßnahmen geplant und umgesetzt werden, sodass eine Fehlfunktion zu keinem wie immer gearteten Schaden, insbesondere beim Menschen, führen kann. Das heißt, dass in diesem Zusammenhang die Konzeption des gesamten Sicherheitskonzepts bei der Automatisierung eine wesentliche Rolle spielt.

Die Situation sieht natürlich vollkommen anders aus, wenn Entscheidungen der KI den Menschen unmittelbar betreffen. In diesem Fall möchte man natürlich schon nachvollziehen können, wie es zu diesen Entscheidungen gekommen ist. Es gibt eine Reihe von Forschungsarbeiten, bei denen man versucht, die Lernarchitektur so zu konzipieren, dass per Konstruktion eine Nachvollziehbarkeit der Ergebnisse gegeben ist. Es gibt Fortschritte, aber ein fertiges System ist zur Zeit nicht verfügbar. Im industriellen Kontext verfolgen wir sogenannte hybride Ansätze. Dabei kombinieren wir physikalisch basierte mathematische Modelle mit maschinellem Lernen und können dabei garantieren, dass das gelernte hybride Modell immer innerhalb der physikalisch plausiblen und zulässigen Grenzen bleibt. Dies erlaubt auch eine gewisse Beweisbarkeit der Ergebnisse.

#### Wo liegen aus Ihrer Sicht die Grenzen des Einsatzes von KI?

Man muss den neuen Methoden gegenüber offen sein und deren Möglichkeiten gezielt nutzen. Aber man muss auch vorsichtig sein und nicht jede neue Methode als Allheilmittel betrachten: Nicht jeder neue Besen kehrt immer und überall besser. Vielmehr sollte man genau überlegen, wann, wo und wie man diese neuen Methoden einsetzt. Bei Automatisierungslösungen bin ich kein Vertreter des starken End-to-End Learnings, bei dem man vorne Daten hineinfüttert und erwartet, dass man am Ausgang alles herausbekommt, was man möchte. Unser Ansatz besteht eher darin, dass wir möglichst viel a-priori-Wissen und spezifisches Domänen-Wissen in den Lernprozess mit einfließen lassen. Unsere Erfahrungen in unzähligen industriellen Anwendungen zeigen, dass die geeignete Kombination aus klassischen Automatisierungsmethoden, physikalisch basierten mathematischen Modellen und modernen Methoden des maschinellen Lernens zu den besten Ergebnissen führt. Ein Beispiel: Die Bewegung eines Roboters kann ich über die Grundgesetze der Mechanik hervorragend beschreiben. Über hochdynamische Regelungen kann ich auf klassische Art und Weise die gewünschten Gelenksgeschwindigkeiten einprägen, für die Modellierung der Umgebung sind aber moderne Verfahren des maschinellen Lernens unabdingbar.

must always plan and implement measures to make sure that a malfunction cannot cause any harm whatsoever, especially for humans. This means that, in this context, the design of the entire safety concept plays an essential role in automation. The situation is, of course, completely different when AI decisions directly affect humans. In that case, one would naturally like to be able to understand how those decisions were made.

There is a number of research projects in which people try to design the learning architecture in such a way that results are traceable by design. There is progress being made, but a complete system is currently not yet available. In an industrial context, we are pursuing so-called hybrid approaches. We combine physically based mathematical models with machine learning and thus can guarantee that the learned hybrid model always stays within physically plausible and permissible boundaries. This also allows for a certain provability of the results.

#### What are, in your views, the limits of AI use?

You have to be open to the new methods and make purposeful use of their possibilities. But you also have to be careful and not view every new method as a cure-all: Not every new broom sweeps cleaner everywhere all the time. Rather, you should consider carefully when, where and how to use these new methods. When it comes to automation solutions, I'm not an advocate of strong end-to-end learning, where you feed in data at the front and expect to get everything you want from it out of the back. Our approach is more about incorporating as much a-priori and specific domain knowledge into the learning process as possible. Our experience in countless industrial applications shows that the suitable combination of classical automation methods, physically based mathematical models, and modern machine-learning methods leads to the best results. To give an example: I can excellently describe robot motion through the basic laws of mechanics, and with highly dynamic controls I can train the desired joint speeds in a classical manner, but for the modeling of the environment, modern machine-learning methods are indispensable.



**Das heißt: Sie setzen in Ihren Forschungs- und Entwicklungsprojekten KI dann ein, wenn man mit klassischen Methoden nicht weiterkommt?**

Ich würde es eher so formulieren: Welches Werkzeug – klassisch oder KI-basiert – ist am besten geeignet, bestimmte Aufgaben zu erfüllen? Wir setzen maschinelles Lernen auch dann ein, wenn die klassischen Methoden zu rechen- und speicherintensiv sind, beispielsweise im Bereich von Computational Fluid Dynamics. Dabei werden mit Hilfe der klassischen Modelle neuronale Netze trainiert, die man nach der Lernphase dann in Echtzeitanwendungen nutzen kann. Auch im Bereich der Bildverarbeitung ersetzen wir einzelne Funktionsblöcke der klassischen Bildverarbeitungs-Algorithmen durch recheneffizientere und robustere Verfahren des maschinellen Lernens.

Generell ist KI für uns ein zusätzliches Werkzeug in der Toolbox. Dabei gilt dasselbe wie bei klassischen Methoden: Ohne spezifisches Domänen-Wissen, ohne a-priori-Wissen über das System, den Prozess, das Werkzeug, den Gegenstand oder das Phänomen, das man analysieren möchte, hat man kaum Aussichten auf Erfolg. Unserer Erfahrung im industriellen Umfeld nach sind die Chancen, ans Ziel zu kommen, umso größer, je weiter man den Anwendungsbereich einschränkt. Je breiter der Anspruch ist, umso schwieriger wird es.

**Maschinen und Roboter werden in den nächsten Jahren unter anderem dank KI immer leistungsfähiger werden. Haben Sie persönlich Sorge, dass sie uns eines Tages über den Kopf wachsen – dass es so etwas wie eine „Singularität“ geben könnte?**

Nein, da habe ich überhaupt keine Sorge. Das würde ich eher in die Kategorie Science-Fiction einordnen. Mich schreckt es auch überhaupt nicht, wenn eine Maschine etwas kann, was ich nicht kann – das ist ja überhaupt kein neues Phänomen. Vor der Technologie selbst habe ich keine Angst – Angst ist ohnehin kein guter Ratgeber. Wovon ich viel mehr Respekt habe, sind die Menschen, die diese Technologien einsetzen. Man weiß ja, dass jede Technologie nicht nur zum Positiven genützt werden kann. Darauf muss man stets ein Auge haben. Das gilt aber nicht nur für KI, sondern für jede Technologie. Im Weiteren müssen wir uns überlegen, wie wir damit umgehen, dass die Komplexität dieser Systeme stets im Steigen begriffen ist und es immer weniger Menschen gibt, die diese Technologie tatsächlich verstehen und beherrschen.

**That means, you use AI in your research and development projects when classical methods don't cut it?**

I would rather put it this way: which tool – classical or AI-based – is best suited to perform certain tasks? We also use machine learning when classical methods are too computation or memory intensive, for example, in the field of Computational Fluid Dynamics. There, neural networks are trained with the help of the classical models, which can then, after the learning phase, be used in real-time applications. In image processing, too, we substitute individual function blocks of classical image processing algorithms with more computationally efficient and robust machine learning methods.

Generally speaking, AI, to us, is one additional tool in the toolbox. And the same things apply as for classical methods: Without specific domain knowledge, without a-priori knowledge of the system, the process, the tool, the object or phenomenon you want to analyze, there's little chance of success. In our experience in an industrial environment, the chances of reaching your goal are better the more you narrow down the area of application. The broader the scope, the more difficult it becomes.

**Machines and robots will become ever more powerful in the coming years. Are you personally worried that they might one day surpass us – that there might indeed be something like a “singularity”?**

No, that does not worry me at all. I'd rather put that in the category of science fiction. Nor does it scare me when a machine can do something I can't – indeed, this is not a new phenomenon at all. I'm not afraid of the technology itself – and fear is not a good advisor anyway. What I am having a lot more qualms about are the people using these technologies. You know that any technology can be used not only for positive things. That is something you always have to keep an eye on. But this is not just true of AI but of any technology. Furthermore, we have to think about how to deal with the fact that the complexity of these systems is ever increasing and that there are fewer and fewer people who are able to actually understand and master that technology.

Dazu kommt noch etwas anderes: Was den Menschen ausmacht, ist neben der Kognition auch die Empathie, die Innovation, die Neugier, die Fingerfertigkeit und Feinfühligkeit und vieles mehr. Diese Kombination ist für technische Systeme unerreichbar. Ich würde mir zum Beispiel einen Roboter wünschen, der bei mir zu Hause den Siphon reinigt. Aber den wird es noch sehr lange nicht geben: Man braucht dafür gezielte feinfühligkeitsbewusste Bewegungen und Koordination auf engstem Raum. Und wenn man die Teile nicht ordentlich übereinander montiert hat, dann tropft es. Es wird immer gesagt, dass das eine einfache Arbeit ist, für die man keine spezielle Ausbildung braucht. Aber für eine Maschine ist das in absehbarer Zeit nicht lösbar.

**Bedeutet das, dass Sie auch keine allzu großen Sorgen haben, dass uns KI und intelligente Roboter die Arbeit wegnehmen?**

Manche leben noch in dieser Denkwelt aus den 1970er-Jahren, dass Automatisierung mit Arbeitsplatzverlust gleichzusetzen ist. Über das sind wir längst hinweg, und auch die Geschichte hat uns gezeigt, dass zwar einzelne Tätigkeiten für den Menschen weggefallen sind, sich aber dadurch gleichzeitig viele neue Möglichkeiten eröffnet haben. Automatisierung hilft, wettbewerbsfähige Arbeitsplätze in Europa zu halten. Heute geht es in der Automatisierung vor allem um intelligente Assistenzsysteme, Erhöhung der Flexibilität und Adaptivität, Steigerung der Ressourceneffizienz und Reduktion von Ausschuss, Abfall und Emissionen. Automatisierung ist eine domänenübergreifende Enabling-Basistechnologie, die in Zukunft der Deal-Breaker sein kann, wenn es um Nachhaltigkeit und Kreislaufwirtschaft geht. Das bedeutet aber, dass man die gesamte Kette neu denken muss: Man muss die Produkte so bauen, dass man sie später automatisiert reparieren sowie zerlegen und werthaltig in den Kreislauf zurückführen kann.

Ein weiterer kritischer Punkt für unsere Wirtschaft ist das Fehlen von geeigneten Fachkräften. Dies erzeugt einen massiven Druck auf die Automatisierung, neue intelligente Assistenzsysteme, Werkzeuge und Maschinen zu entwickeln, die die Menschen in der Arbeit bestmöglich unterstützen, deren Belastung senken und die Arbeitsplätze attraktiver machen. ✖

And there is something else: What defines being human, aside from cognition, also is empathy, innovation, curiosity, dexterity and sensitivity and many other things. This combination is unattainable for technological systems. I, for one, would really wish to have a robot to clean the sink drain in my home. But it's not going to happen for a very long time to come: you need targeted delicate movements and coordination in a very confined space. And if you don't put the parts properly together again, it will be dripping. It's always said that this is a simple piece of work that does not need any special training. But for a machine it will remain unachievable in the foreseeable future.

**Does that mean you are not too worried about AI and robots taken our jobs away from us?**

Some still live in this world of 1970s thinking in which automation is tantamount to job loss. We're long past that, and history has also shown us that while individual human jobs have been lost, many new opportunities have opened up as a result. Automation helps to keep competitive jobs in Europe. Today, automation is primarily about intelligent assistance systems, increasing flexibility and adaptivity, improving resource efficiency and reducing scrap, waste, and emissions. Automation is a cross-domain basic enabling technology that might be a deal-breaker in the future when it comes to sustainability and the circular economy. But that means you have to rethink the entire chain: you have to build products in a way that you can later automate their repair as well as disassembly and return them to the cycle with retained value.

Another crucial point for our economy is the shortage of skilled labor force. This creates massive pressure on automation efforts to develop new intelligent assistance systems, tools and machinery that provide the best possible support for people at work, reduce their stress, and make workplaces more attractive. ✖

# Intelligente Helfer im Alltagsleben

**Vor allem das Smartphone öffnet die Tür für viele Anwendungen mit Künstlicher Intelligenz: KI-Apps unterstützen uns in vielen Alltagssituationen, helfen uns bei der Suche nach einer Information oder nach dem richtigen Weg und erlauben es selbst Laien, Vogelstimmen richtig zuzuordnen.**

Auch wenn es viele nicht bemerken: Künstliche Intelligenz ist längst in unserem Alltag angekommen. Aus vielen IT-Dienstleistungen, die wir tagtäglich wie selbstverständlich nutzen, sind KI-Methoden nicht mehr wegzudenken. Der wohl wichtigste Zugang zu KI in unserem Alltag ist das Smartphone – etwa wenn wir das Handy per Gesichtserkennung entsperren, den Weg zu einem Restaurant suchen, auf einer Internet-Plattform einkaufen, auf einer Musik-Plattform einen neuen Song empfohlen bekommen oder auch nur in einer Suchmaschine etwas finden wollen. KI nutzen wir weiters, wenn eine Messaging-App automatisch ein Wort, das wir eintippen, ergänzt, wenn wir mit den Siris und Alexas dieser Welt reden oder wenn uns ein Assistenzsystem im Auto dabei hilft, Abstand und Spur zu halten. Die meisten dieser Anwendungen nutzen die neuen Möglichkeiten des Maschinenlernens. Im Vorteil sind dabei die großen Technologiekonzerne, allen voran die „Big Five“ (GAFAM: Google [Alphabet], Amazon, Facebook [Meta], Apple und Microsoft). Diese Unternehmen verfügen zum einen über eine Unzahl an Daten, die sie in ihrer Tätigkeit (mehr oder weniger legal) sammeln, und zum anderen über die erforderliche Rechenleistungen, um diese Daten in große Lernsysteme einzuspeisen und die Ergebnisse entsprechend – nämlich für ihre Zwecke – zu nutzen.

## Mit den Geräten sprechen

Ein Beispiel ist die Suchmaschine Google: Groß geworden durch die Erfindung und perfekte Nutzung des sogenannten „PageRank“-Algorithmus – eines Rechenverfahrens, das die Verlinkungsstruktur und insbesondere die Popularität bestimmter Links bewertet und daraus die Ergebnisreihenfolge bei einer Suchabfrage ableitet –, nutzt der Konzern zunehmend auch KI-Methoden, die in den potenten konzerneigenen Forschungsstätten entwickelt werden. Ein Beispiel ist das System BERT (Bidirectional Encoder Representations from Transformers), das seit 2019 unter anderem natürliche Sprache bei der Sucheingabe ermöglicht. Dabei werden nicht nur die eingetippten (oder gesprochenen) Wörter, sondern auch deren Beziehung zueinander in die Auswertung einbezogen. Viele weitere KI-Anwendungen, etwa in Google Maps oder Gmail, werden durch die von Google entwickelte plattformunabhängige Open-Source-Softwarebibliothek TensorFlow ermöglicht – alles auf Basis der unzähligen Daten, die Google wie ein Schwamm aufsaugt.

# Intelligent tools in everyday life

**Smartphones in particular open the door to many Artificial Intelligence applications: AI apps support us in a large number of everyday situations, assist us in finding information or the right route to our destination, and even allow non-experts to correctly classify bird calls.**

Even if many people don't notice it: artificial intelligence has already arrived in our everyday lives. Many IT services we use as a matter of course every day would not be imaginable without AI methods. Probably the most important access to AI in our everyday life is the smartphone – e.g., when we unlock our mobile phone by facial recognition, when we search for the way to a restaurant, when we shop on an internet platform, when we get a recommendation for a new song on a music platform or when we simply want to find something in a search engine. We also use AI when a messaging app automatically completes a word we type, when we talk to the Siris and Alexas of this world or when an assistance system in our car helps us to keep our distance and stay in our lane. Most of these applications use the new possibilities of machine learning. The big technology groups, above all the “Big Five” (GAFAM: Google [Alphabet], Amazon, Facebook [Meta], Apple and Microsoft), enjoy an advantage here. These companies have, on the one hand, a vast amount of data that they collect (more or less legally) in the course of their activities and, on the other hand, the necessary computing power to feed this data into large learning systems and use the results correspondingly – namely for their own purposes.

## Talking to the devices

One example is the search engine Google: Having become big through the invention and perfect use of the so-called “PageRank” algorithm – a calculation method that evaluates the linking structure and, in particular, the popularity of certain links and derives the result order for a search query from this – the Group is also increasingly using AI methods that are being developed in the Group's own powerful research facilities. One example is the BERT (Bidirectional Encoder Representations from Transformers) system, which, among other things, has enabled using natural language as search input since 2019. Not only the typed (or spoken) words are analyzed, but also their relationship to each other. Many other AI applications, such as in Google Maps or Gmail, are made possible by the platform-independent open-source software library TensorFlow developed by Google – all

Auch Multimedia-Plattformen oder Streaming-Dienste kennen ihre Kund:innen und ihre (bisherigen) Vorlieben recht genau – und dieses Wissen wird in Recommender-Systemen mithilfe von KI-Algorithmen dazu genutzt, um weitere Medieninhalte vorzuschlagen. Ein erklecklicher Anteil – Schätzungen reichen von 10 bis 30 Prozent – des Volumens geht auf solche Empfehlungen zurück; Ähnliches gilt für Online-Händler. Netflix & Co. setzen außerdem auf maschinelles Lernen, wenn es um die Finanzierung von Film- und Serienproduktionen geht: Dank KI können die Unternehmen vorausberechnen, welche Titel sich in bestimmten Märkten voraussichtlich lohnen.

### **Kontrolle, Sicherheit und Privacy**

KI ist auch beim leidigen Thema der Kontrolle von Inhalten in Sozialen Medien ein großes Thema. Angesichts der Überfülle an Postings ist eine händische Überprüfung fragwürdiger Inhalte völlig illusorisch. Bei der Identifizierung von gewaltverherrlichenden, pornografischen oder politisch extremen Inhalten kommen KI-Methoden zum Einsatz, die Auffälligkeiten automatisch aufspüren. Dass dabei viele Fehler passieren, ist offensichtlich – etwa dass die 30.000 Jahre alte „Venus von Willendorf“ von Facebook als pornografisches Objekt eingestuft wurde. Ein heiß diskutiertes Thema ist auch die Gesichtserkennung – eine Aufgabe, die KI-Lernsysteme mittlerweile beinahe perfekt beherrschen. Was zum Beispiel beim Entsperren eines Smartphones sehr angenehm ist, wird beim Einsatz in größeren Datenbeständen schnell zu einer Gefahr für Datenschutz und Privacy. Aufgrund weltweiter Proteste musste Google den ersten Versuch, sein „Glass“ mit eingebauter Kamera auf den Markt zu bringen, abbrechen; auch Facebook zog die Möglichkeit zur automatischen Identifizierung von Personen auf Fotos wieder zurück. Auch im Büroalltag ist KI längst angekommen – man denke nur an ziemlich zuverlässige Spam-Filter, die lästige oder gefährliche E-Mails von uns abhalten, oder an frei über das Internet verfügbare Übersetzungsdienste, die einfache Alltagstexte in verblüffender Qualität in unzählige Sprachen übertragen.

Dass KI auch die Welt der Computerspiele revolutioniert, darf nicht überraschen – immerhin waren es Spiele wie Schach oder Go, die der Allgemeinheit erstmals die Leistungsfähigkeit des Maschinenlernens demonstrierten. KI-Systeme können sich die Regeln eines Spiels durch das sogenannte „reinforcement learning“ binnen kürzester Zeit aneignen und dann Strategien entwickeln, die der menschlichen Vorausschau bei weitem überlegen sind. Das betrifft mittlerweile auch Computerspiele, bei denen man bisher dachte, dass Intuition und ein Gefühl für die Sache entscheidend seien, um die vielen Einflussfaktoren optimal unter einen Hut zu bringen. So sorgten Forscher:innen des Elektronikkonzerns Sony zu Jahresbeginn mit einer Publikation in „Nature“ (Vol. 602, 223) für Aufsehen, in der sie ein KI-System namens „GT Sophy“ beschrieben, das alle menschlichen Akteure beim Autorennspiel „Gran Turismo“ besiegte – einfach weil es Bremskraft und -zeitpunkte vor Kurven sowie die Ideallinie durch Kurven noch besser hinbekam.

based on the innumerable data that Google soaks up like a sponge. Multimedia platforms or streaming services are also quite familiar with their customers and their (previous) preferences – and this knowledge is used in recommender systems with the help of AI algorithms to suggest further media content. A considerable proportion – estimates range from 10 to 30 per cent – of the volume goes back to such recommendations; the same applies to online retailers. Netflix & Co. also rely on machine learning when it comes to financing film and series productions: Thanks to AI, the companies can predict which titles are likely to be profitable in certain markets.

### **Control, security, and privacy**

AI is also a big issue in the troublesome topic of content control in social media. Given the overabundance of postings, a manual check of questionable content is totally illusory. When identifying content that glorifies violence, is pornographic or politically extreme, AI methods are used that automatically detect conspicuous features. It is obvious that many mistakes are made in the process – for instance, the 30,000-year-old “Venus of Willendorf” was classified as a pornographic object by Facebook. Another hotly debated topic is facial recognition – a task that AI learning systems have now mastered almost perfectly. But what may be very convenient when unlocking your smartphone can quickly turn into a threat to data protection and privacy when used in larger data sets. As a result of worldwide protests, Google had to abandon the first attempt to launch its “Glass” with a built-in camera; Facebook also withdrew the possibility of automatically identifying people in photos. AI has long since arrived in everyday office life as well – just think of fairly reliable spam filters that keep annoying or dangerous emails from us, or translation services freely available on the internet that translate simple everyday texts into innumerable languages with astonishing quality.

It should come as no surprise that AI is also revolutionizing the world of computer games – after all, it was games like Chess or Go that first demonstrated the power of machine learning to the general public. AI systems are able to learn the rules of a game within a very short time through so-called “reinforcement learning” and then develop strategies that are far superior to human foresight. This now applies to computer games as well, where it was previously thought that intuition and a feeling for the matter were decisive in order to optimally reconcile the many influencing factors. At the beginning of the year, for example, researchers from the electronics company Sony caused a stir with a publication in Nature (Vol. 602, 223) in which they described an AI system called “GT Sophy” that beat all human players in the car racing game “Gran Turismo” – simply because it managed braking power and timing before bends and the ideal line through bends even better.

### Wie heißt denn diese Blume?

Auch viele Naturliebhaber:innen wollen und können nicht mehr auf KI-Systeme verzichten, die in Form von Smartphone-Apps gratis oder um sehr wenig Geld angeboten werden. So hat beispielsweise das Cornell Lab of Ornithology die App BirdNET programmiert, die gelernt hat, tausende Vogelarten anhand ihrer Stimmen zu erkennen. Eine deutschsprachige Version hat die TU Chemnitz herausgegeben, die selbst aus einem gleichzeitigen Gezwitscher unterschiedlichster Arten mit erstaunlicher Genauigkeit die Vogelart herausfindet (und in Zweifelsfällen entsprechend der Wahrscheinlichkeit in einer Liste reiht). Sehr beliebt ist auch die App „Flora Incognita“ der TU Ilmenau, die tausende Blumen, Gräser, Farne und Bäume anhand eines Fotos identifizieren kann.

Diese Tools dienen zugleich der Forschung: Wenn die User:innen der Weiterverarbeitung der Daten und der Übermittlung der Standorte zustimmen, können aus den vielen einzelnen Bestimmungen automatisch Verbreitungskarten verschiedenster Lebewesen erstellt werden. KI-Methoden und die Einbindung von Bürger:innen in die Forschung ermöglichen auf diese Weise die Erstellung von globalen Inventurlisten zur Biodiversität – so wie es derzeit zum Beispiel in dem internationalen Projekt LIFEPLAN geschieht, in dem (mit starker österreichischer Beteiligung) einheitliche Standards zur Messung der Artenvielfalt unter anderem anhand von Tiergeräuschen entwickelt und etabliert werden.

Es wird nicht mehr lange dauern, bis auch Werkzeug oder Haushaltsgeräte mit KI-Funktionen ausgestattet werden. Dies hat etwa der Technologiekonzern Bosch zu Jahresbeginn angekündigt. Als Beispiel wurde das Sensorsystem SoundSee genannt, das mit nachgeschalteter KI Geräusche analysiert, um herauszufinden, wann eine Wartung erforderlich ist. Dieselbe Technologie wird seit einigen Jahren in der Internationalen Raumstation ISS eingesetzt, um ungewöhnliche und auf potenzielle Gefahren hindeutende Geräusche herauszufiltern. Oder hier auf der Erde für die Diagnose von Lungenerkrankungen, wie etwa Asthma, schon in einer frühen Phase.

### KI für den Gaumen

Und selbst die Kulinarik wird in absehbarer Zeit nicht von der Anwendung von KI-Methoden ausgespart bleiben. So entwickelten Forscher:innen des Karlsruher Instituts für Technologie (KIT) einen mobilen Sensor, der flüchtige Substanzen registriert und mit einem Smartphone (zur Anbindung an eine KI) gekoppelt wird. Dieser war imstande, verschiedene Minzesorten zu unterscheiden. Oder ein anderes Beispiel: Der Schweizer Aromen- und Duftstoff-Hersteller Firmenich hat kürzlich bekanntgegeben, dass in Kooperation mit Spitzenköch:innen an einem KI-System gearbeitet wird, das pflanzenbasierte Fleischalternativen optimal würzen soll. Als Trainingsdaten für die KI diente eine Datenbank mit tausenden Zutaten und Geschmäckern aus jahrzehntelanger Erfahrung. Das System soll Ähnlichkeiten zwischen Aromen und deren Wechselwirkungen aufspüren und schließlich gezielt gewisse Geschmacks- und Geruchsvorstellungen erfüllen. ✖

### What is the name of this flower?

Even many nature lovers no longer want to do or even can do without AI systems which are offered free of charge or for very little money as smartphone apps. For instance, the Cornell Lab of Ornithology programmed the app BirdNET which can recognize thousands of bird species by their voices. A German-language version published by the Technical University of Chemnitz identifies bird species with amazing accuracy even from a simultaneous chirping of a wide variety of species (and in cases of doubt ranks them in a list according to probability). The app “Flora Incognita” from the TU Ilmenau, which can identify thousands of flowers, grasses, ferns, and trees from a photo, is also very popular.

These tools also serve research purposes: provided the users agree to the further processing of the data and the transmission of the locations, distribution maps of a wide variety of living organisms can be automatically created from the many individual identifications. In this way, AI methods and the involvement of citizens in research make it possible to create global biodiversity inventories – as is currently happening, e.g., in the international LIFEPLAN project, in which (with strong Austrian participation) uniform standards for measuring biodiversity are being developed and established on the basis of animal sounds and other parameters.

It won't take long before tools or household appliances will also be fitted with AI functions. The technology group Bosch, e.g., announced this at the beginning of the year. The sensor system SoundSee, which analyzes sounds with downstream AI to find out when maintenance is required, was mentioned as an example. The same technology has been used for several years on the International Space Station (ISS) to filter out unusual noises and those indicating potential danger. Or here on Earth for diagnosing lung diseases such as asthma at an early stage.

### AI for the palate

And even the culinary realm won't be able to escape the application of AI methods in the foreseeable future. Researchers at the Karlsruhe Institute of Technology (KIT), for example, have developed a mobile sensor that registers volatile substances and is coupled with a smartphone (for connection to an AI). It was able to distinguish between different types of mint. Or another example: The Swiss flavor and fragrance manufacturer Firmenich recently announced that it is working on an AI system in collaboration with top chefs to optimally season plant-based meat alternatives. A database with thousands of ingredients and tastes from decades of experience served as training data for the AI. The system is supposed to detect similarities between flavors and their reciprocal effects and eventually provide products precisely geared to certain taste and smell preferences. ✖



# Sprachalgorithmen: Wenn Computer Geschichten schreiben

**Sprachalgorithmen zählen derzeit zu jenen Bereichen der KI, die sich am stürmischsten entwickeln. Diese Systeme können in der Tat schon viel – aber sie haben auch grundsätzliche Mängel.**

Kaum eine andere Technologie aus dem Bereich Künstliche Intelligenz hat zuletzt weltweit solche medialen Wellen geschlagen wie die Sprachverarbeitung. So publizierte etwa der britische „Guardian“ einen Text, der von der Software GPT-3 produziert wurde. Im deutschen Sprachraum sorgte ein Text von Daniel Kehlmann für Aufsehen: In „Mein Algorithmus und ich“ dokumentierte er seinen Selbstversuch, gemeinsam mit einer Sprachsoftware im Ping-Pong-Verfahren eine Kurzgeschichte zu schreiben (Daniel Kehlmann: Mein Algorithmus und Ich. Stuttgarter Zukunftsrede. ISBN: 978-3-608-98480-4, 2021, 64 Seiten, Klett-Cotta).

## **Durchwachsender Selbstversuch eines Schriftstellers**

Kehlmanns Ergebnis wurde allgemein, und auch von ihm selbst, zwar als interessant, aber unterm Strich als doch etwas dürrig beurteilt – auch wenn der Schriftsteller nach eigenem Bekunden immer wieder von manchen Inputs des Sprachalgorithmus überrascht war. Aber: Das Hauptziel des Experiments wurde nicht erreicht – es kam keine Kurzgeschichte zustande. Das lag zu einem guten Teil an der verwendeten Software: Der kalifornische Sprachalgorithmus CTRL ist ein sogenanntes Corpus-basiertes KI-System – ein System mit einer großen Datensammlung mit Hunderttausenden von Büchern, Zeitungen und Online-Foren (dem „Corpus“) und einem statistischen Auswertungsalgorithmus. Dieser greift auf ein Gedächtnis aus Abermillionen von Sätzen und Wortfolgen, die in Form von Wortfragmenten (Tokens) erfasst sind, zurück und reiht diese Bausteine gemäß ihrer Wahrscheinlichkeit. Wenn es statistisch am wahrscheinlichsten ist, dass auf das Wort A das Wort B folgt, dann bringt das System nach dem Wort A auch das Wort B.

Dieses Verfahren produziert Sätze, die nicht falsch sind und durchaus Sinn ergeben – für kurze Texte und für Antworten auf einfache Fragen funktioniert das recht gut. Allerdings hat die Methode im Fall eines längeren Textes mehrere eklatante Schwächen: Abgesehen davon, dass das System nach einer gewissen Textlänge abstürzt (weil die erforderliche Rechenleistung deutlich schneller steigt als der Umfang des Textes bzw. des Corpus), ist

# Language Algorithms: When computers write stories

**Language algorithms are currently among those AI areas developing at a most dynamic pace. These systems are in fact quite capable already – however, they also have fundamental shortcomings.**

Hardly any other technology in the field of artificial intelligence has recently made such media waves all over the world as speech processing. The British “Guardian” published a text that was produced by the software GPT-3. In the German-speaking world, a text by Daniel Kehlmann created quite a stir: In “Mein Algorithmus und Ich” [My Algorithm and I], he documented his self-attempt to write a short story together with a language software using the ping-pong method (Daniel Kehlmann: Mein Algorithmus und Ich. Stuttgarter Zukunftsrede. ISBN: 978-3-608-98480-4, 2021, 64 pages, Klett-Cotta).

## **A writer’s self-experiment with mixed results**

Kehlmann’s result was widely considered interesting, including by himself, but on the whole somewhat lacking – even if the writer was, by his own admission, repeatedly surprised by some of the speech algorithm’s inputs. But: the main goal of the experiment was not achieved – no short story came to life. This was largely due to the software used: the Californian language algorithm CTRL is a so-called corpus-based AI system – a system with a large data collection with hundreds of thousands of books, newspapers and online forums (the “corpus”) and a statistical evaluation algorithm. This algorithm draws on a memory of millions and millions of sentences and word sequences, which are recorded in the form of word fragments (tokens), and ranks these building blocks according to their probability. If it is statistically most likely that word A will be followed by word B, then the system will also put word B after word A.

This method produces sentences that are not wrong and make quite a bit of sense – which works quite well for short texts and for answers to simple questions. However, when it comes to a longer text, the method has several evident weaknesses: besides the fact that the system crashes after a certain text length (because the required computing power increases much faster than the size of the text or corpus), it is only as good as the quality of the texts fed in, with all its

es nur so gut wie die Qualität der eingespeisten Texte mit allen Schwächen und Fehlern. Enthält die Datenbasis beispielsweise rassistische Texte, so finden sich diese auch in den Texten der Maschine. „Künstliche Intelligenz ist ein Zweitverwerter. Alles, was sie tun kann, speist sich aus der vom Internet verfügbar gemachten Tätigkeit unzähliger Menschen“, resümiert Kehlmann. Zudem fehlt es dem Sprachalgorithmus an einer inneren Logik. „Zufall ist nicht Kreativität“, bemerkte der Publizist Hans Rudolf Straub treffend in seiner Analyse des Kehlmann-Selbstversuchs. Eine funktionierende Geschichte entstehe erst, wenn es einen logischen Zusammenhang zwischen den Elementen gebe. „Und diese innere Logik fehlt einem Corpus-basierten Programm prinzipiell.“ (<https://hrstraub.ch/kuenstliche-intelligenz-daniel-kehlmann-und-ctrl/>)

### **Robo-Autor im „Guardian“**

Von ungleich größerer Qualität war da der Text, den die britische Tageszeitung „The Guardian“ am 8. September 2020 veröffentlichte, in dem der Sprachalgorithmus GPT-3 argumentierte, dass der Mensch keine Angst vor KI zu haben brauche – nachdem ihm Guardian-Redakteur:innen zwei kurze Absätze als Anleitung vorgegeben hatten (<https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>). GPT-3 ist die Abkürzung für „Generative Pre-trained Transformer 3“. Dabei handelt es sich um einen von der Organisation OpenAI entwickelten Sprachalgorithmus, der die KI-Methode des Deep Learning nutzt und mit einem Sprachmodell verknüpft ist, das anhand der Daten immer weiter verfeinert wird. Als Trainingsdaten dienten fünf Datensätze (inklusive Wikipedia) mit insgesamt knapp 500 Milliarden Tokens. Aus diesem Lerndatensatz leitete die Software in einem mehrschichtigen neuronalen Netzwerk nicht weniger als 175 Milliarden Parameter ab, die implizit beschreiben, wie die (englische) Sprache funktioniert.

Die Ergebnisse begeistern auf der einen Seite Expert:innen – die Texte sind das mit Abstand Beste, was KI-Sprachalgorithmen derzeit zu bieten haben. Im Fall des Guardian-Artikels, so berichten die Redakteur:innen, hat GPT-3 acht verschiedene Essays produziert, die alle grammatikalisch perfekt und differenziert und in sich schlüssig argumentiert waren; für den in der Zeitung publizierten Essay hat die Redaktion Textteile aus acht vom Algorithmus produzierten Texten ausgewählt – auch um zu zeigen, dass GPT-3 unterschiedliche Stile beherrscht. Unterm Strich sei die Produktion des Essays deutlich rascher gegangen als die meisten menschlichen Redakteur:innen für einen entsprechenden Text gebraucht hätten, heißt es von Seiten des Guardian.

### **KI kaum geeignet für Qualitätsjournalismus**

Das bedeutet allerdings – und das ist die andere Seite – nicht automatisch, dass GPT-3 für den Routineeinsatz in Zeitungen und anderen Medien prädestiniert wäre. Denn: Wie bei allen Maschinenlernsystemen kann der Algorithmus nicht mehr als die Trainingsdaten. Es gibt viele Fälle, in denen der Sprachalgorithmus Texte produzierte, die schwer rassistisch oder

flaws and errors. If the database contains racist texts, for example, these will also be found in the machine’s texts. “Artificial intelligence is a recycler. Everything it can do is fed from the activity of countless people made available by the internet,” Kehlmann sums up. Furthermore, the language algorithm lacks an inner logic. “Coincidence is not creativity,” the publicist Hans Rudolf Straub aptly noted in his analysis of Kehlmann’s self-experiment. A functioning story only comes about when there is a logical connection between the elements. “And this inner logic is missing in a corpus-based programme as a matter of principle.” (<https://hrstraub.ch/kuenstliche-intelligenz-daniel-kehlmann-und-ctrl/>)

### **Robo-Autor in the “Guardian”**

A text published by the British daily newspaper “The Guardian” on 8<sup>th</sup> September 2020, in which the language algorithm GPT-3 argued that humans need not be afraid of AI, was of incomparably greater quality – after Guardian editors had given it two short paragraphs as a guide (<https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>). GPT-3 is the acronym for “Generative Pre-trained Transformer 3”. It is a language algorithm developed by the organization OpenAI which uses the AI method of Deep Learning and is connected to a language model that is refined further and further based on the data. Five data sets (including Wikipedia) with a total of almost 500 billion tokens served as training data. From this learning data set, the software derived no less than 175 billion parameters in a multi-layer neural network that implicitly describe how the (English) language works.

On the one hand, the results enthuse experts – the texts are by far the best that AI language algorithms currently have to offer. According to the editors, GPT-3 produced eight different essays in the case of the Guardian article, all of which were grammatically perfect, differentiated and coherently argued; the editors selected parts from eight texts produced by the algorithm for the essay published in the newspaper – also to demonstrate that GPT-3 can master different styles. All in all, the production of the essay was much faster than most human editors would have needed for a corresponding text, according to the Guardian.

### **AI hardly suitable for quality journalism**

However – and that is the flip side – this does not automatically imply that GPT-3 is predestined for routine use in newspapers and other media. For, as with all machine learning systems, the algorithm is not capable of doing more than its training data allow. There are many cases in which the language algorithm produced texts that are severely racist or misogynistic, sometimes full of swear words. Another serious weakness is that the system is prone to repetitions. And that it also tells fibs from time to time, i.e. it produces incorrect

frauenfeindlich sind und teils vor Schimpfwörtern strotzen. Ein weiterer gravierender Schwachpunkt ist, dass das System zu Wiederholungen neigt. Und dass es hin und wieder auch flunkert, also unrichtige Dinge und Fantasiegebilde produziert – wenn es zum Beispiel noch gilt, einen Text zu verlängern. „GPT-3 denkt sich offenbar Informationen aus, so dass sie zum erfragten Muster passen“, merkt etwa die Bloggerin Tina Nord nach einem Selbstversuch an (<https://www.lernen-wie-maschinen.ai/ki-pedia/>). Daher ist GTP-3 für ernsthaften Journalismus nur wenig geeignet – maximal nach aufwendigem Lektorat und Faktencheck.

Das ist auch einer der zentralen Schlüsse aus dem Whitepaper „Journalistische Textautomatisierung“, das die Austria Presse Agentur (APA) kürzlich veröffentlicht hat (Download: <https://apa.at/service/whitepaper>). Basierend auf Erkenntnissen der Studie „AI.AT.Media“ (durchgeführt von der APA gemeinsam mit Joanneum Research) und drei Jahren Erfahrung der Nachrichtenagentur mit Systemen zur automatischen Generierung von Texten werden der Status Quo beschrieben sowie die Potenziale und Limitationen aufgezeigt. In manchen Bereichen ist Textautomatisierung mithilfe von Methoden der „Natural Language Generation (NLG)“ bereits möglich – etwa bei Bilanzberichterstattung, Sportberichten, Wahlergebnissen oder der täglichen Meldung der aktuellen COVID-19-Zahlen. Diese Softwaresysteme zählen zur Gruppe der regelbasierten Automatisierungsverfahren – sie enthalten im Kern also keine Elemente des Maschinenlernens (die freilich für Hilfszwecke wie etwa zur Recherche integriert werden können). „Aus Gründen der Nachvollziehbarkeit, der Qualitätssicherung und der in den Texten verbürgten Fakten ist regelbasierte Automatisierung derzeit das einzige anwendbare Verfahren im Journalismus“, schreiben die Studienautor:innen.

### **Intelligente Lückentexte, gefüllt mit aktuellen Daten**

In den genannten Fällen liegen strukturierte Daten vor, die automatisch in zuvor definierte „Lückentexte“ eingefügt werden. Dieser Rahmen für die aktuellen Daten folgt in diesen Bereichen typischerweise einem bekannten Muster. Um so weit zu kommen, ist im Vorfeld viel journalistische Arbeit erforderlich – beginnend bei der Analyse des Quelldatensatzes und der Beurteilung des Informationsgehalts bis hin zum Storytelling, also dem Aufbau und der Gestaltung des „Lückentextes“, und der Definition der Variablen, die eingefügt werden sollen. „Die Redakteur:innen müssen antizipativ arbeiten und jede mögliche Geschichte planen und schreiben“, heißt es in dem Papier.

Die Systeme ersetzen die Arbeit von Journalist:innen nicht, wird betont, aber sie ergänzen diese: Die Algorithmen schaffen die Basis für zusätzliche Textformate, etwa maßgeschneidert für unterschiedliche Plattformen. Überdies werden Dinge möglich, die bisher nicht machbar waren: Die APA beispielsweise produziert mit Hilfe von NLG nach Wahlen blitzschnell Ergebnismeldungen für jede Gemeinde – etwas, das händisch an einem Wahlabend undenkbar wäre. ✖

things and fantasies – for example, when it is necessary to lengthen a text. “GPT-3 apparently makes up information so that it fits the pattern being asked,” notes blogger Tina Nord after a self-experiment (<https://www.lernen-wie-maschinen.ai/ki-pedia/>). Therefore, GTP-3 is hardly suitable for serious journalism – at best after elaborate editing and fact-checking.

This is also one of the central conclusions of the white paper “Journalistic Text Automation” recently published by the Austria Press Agency (APA) (Download: <https://apa.at/service/whitepaper>). It describes the status quo along with the potentials and limitations based on the findings of the study “AI.AT.Media” (conducted by APA together with Joanneum Research) and three years of experience of the news agency with systems for the automatic generation of texts. In some areas, text automation using methods of “Natural Language Generation (NLG)” is already possible – for example in balance sheet reporting, sports reports, election results or the daily reporting of the current COVID-19 figures. Such software systems belong to the group of rule-based automation methods – they do not contain any elements of machine learning at their core (which can, of course, be integrated for auxiliary purposes such as research). “For reasons of traceability, quality assurance and the facts vouched for in the texts, rule-based automation is therefore currently the only applicable procedure in journalism,” the authors of the study write.

### **Intelligent gap texts, filled with current data**

In the cases described above, structured data is at hand that is automatically inserted into previously defined “gap texts”. This framework for the current data typically follows a known pattern in these areas. To get that far, a lot of journalistic work is required beforehand – starting with the analysis of the source data set via assessing the information content all the way to the storytelling, i.e. the structure and design of the “gap text” and the definition of the variables to be inserted. “Editors have to anticipate future events and plan and write every possible story,” the paper says.

It emphasizes that the systems do not replace the work of journalists, but they do supplement it: the algorithms create the basis for additional text formats, such as ones tailored for different platforms. They also open the door to things that were not feasible before: APA, for example, uses NLG to produce results reports for each municipality in a flash after elections – something that would be unthinkable to do manually on election night. ✖

## Sicherheit & Fake News

**KI-Methoden sind zugleich eine Gefahr für IKT-Systeme als auch eine große Chance, um die Sicherheit in vielen Bereichen stark zu verbessern. Einige aktuelle Beispiel aus den Laboren des AIT Austrian Institute of Technology.**

Die Verbreitung von Desinformation ist eine ernste Bedrohung: Fake News beeinflussen Entscheidungen, führen zu Verunsicherung, zum Aufbau von Unzufriedenheit in der Öffentlichkeit und zur Destabilisierung von Gesellschaft und Demokratie. In Krisenzeiten wie der aktuellen COVID-19-Pandemie oder dem Krieg in der Ukraine hat Desinformation einen neuen negativen Gipfelpunkt erreicht – allerorts ringt man um adäquate Maßnahmen zur Eindämmung. Hoffnungen ruhen dabei auf neuen Technologien: Mit neuen, auf KI basierenden Methoden sowie mit der Zusammenführung von Informationen aus unterschiedlichen Quellen soll es möglich werden, den Wahrheitsgehalt von Nachrichten künftig automatisch zu überprüfen. Angesichts der Vielzahl an ständig im Netz kursierenden Meldungen ist eine händische Überprüfung chancenlos. KI wird freilich auch dazu eingesetzt, Fake News zu produzieren. So zum Beispiel „Deep Fakes“, also manipulierte Videos, in denen die handelnden Personen ausgetauscht wurden bzw. ihnen Aussagen in den Mund gelegt werden, die sie niemals gesagt haben. Das ist heute in immer besserer Qualität mit „Generative Adversarial Networks“ (GAN), also zwei miteinander gekoppelten Deep-Learning-Systemen, möglich: Ein „Generator“ erzeugt dabei künstliche Bilder, ein „Diskriminator“ versucht, diese als echt oder falsch einzustufen. In vielen Runden lernt der Generator, Bilder besser zu fälschen, und der Diskriminator, diese korrekt zu klassifizieren – diese beiden neuronalen Netze trainieren einander also gegenseitig. Zur Erkennung solcher „Deep Fakes“ gibt es eine Reihe forensischer Techniken, zu denen nun auch KI-gestützte Methoden kommen, die nach Unregelmäßigkeiten oder bestimmten Mustern suchen.

### Computer findet Fake News

Ebenso anspruchsvoll ist es, automatisch Fake News in Texten zu finden. Für Computer ist ein Text kein Wissen, sondern nur Worte, die aneinandergereiht werden. Ein Abgleich von Fakten mit Datenbanken wie z. B. Wikipedia funktioniert daher maximal bei einer sehr formellen Sprache. Mit Hilfe von KI-Systemen zur Sprachverarbeitung können Wörter aber zunehmend in einen relevanten Kontext gesetzt werden. Die Methoden dazu werden immer ausgefeilter. Am AIT Austrian Institute of Technology zum Beispiel wird an einem System gearbeitet, das zur Analyse von Fake News Hinweise aus unterschiedlichen Bereichen – Bild, Video, Audio, Text – miteinander kombiniert und überdies die Verbreitung von Meldungen in

## Security & Fake News

**AI methods are both a threat to ICT systems and a great opportunity to improve security significantly in many areas. Some current examples from the AIT Austrian Institute of Technology labs.**

Spreading of disinformation is a serious threat: fake news may influence decisions, lead to insecurity, public dissatisfaction and the destabilization of society and democracy. In times of crisis such as the current COVID-19 pandemic or the war in Ukraine, disinformation has reached a new negative peak – people everywhere are struggling to find adequate countermeasures. Hopes rest on new technologies: With new, AI-based methods and the combination of information from different sources, it should in the future be possible to check automatically to see whether a piece of news is actually true. In view of the large number of reports constantly circulating on the Internet, it would be almost impossible to perform a manual check. On the other hand, it must be conceded that AI is also being used to produce fake news. An example is “deep fakes”, i.e. manipulated videos in which the people involved have been exchanged or they are made to seem as if they are saying things they in fact never said. Nowadays, it is increasingly possible to improve the quality of these using “Generative Adversarial Networks” (GAN), i.e. two coupled deep learning systems: a “generator” that creates artificial images while a “discriminator” tries to classify them as real or false. Over many iterations, the generator learns to falsify images better and the discriminator learns to classify them correctly – the two neural networks actually train each other. There are a number of forensic techniques for detecting such “deep fakes”. AI-supported methods that look for irregularities or specific patterns have now been added to these.

### Computer Tracks Down Fake News

It is just as challenging to find fake news in texts automatically. For computers, a text is not knowledge, it is just words that are strung together. A comparison of facts against databases such as Wikipedia therefore works only if the language is very formal. With the help of AI systems for language processing, however, words can increasingly be placed in a relevant context. The methods used to do this are becoming increasingly sophisticated. The AIT Austrian Institute of Technology, for example, is working on a system that combines information from different areas – images, video, audio, text – to analyze fake news. It also takes into account the distribution of reports over social media. In



den Sozialen Medien berücksichtigt. In einem Projekt im Rahmen des österreichischen Sicherheitsforschungsprogramms KIRAS wird aktuell ein auf KI basierendes Fake-News-Detektionssystem erarbeitet, das auf die jeweiligen Endanwender:innen abgestimmt ist und im täglichen Arbeitsumfeld einsetzbar ist. Allerdings reicht es in der Praxis nicht aus, eine Nachricht als Ganzes als richtig oder falsch einzustufen. Man benötigt detailliertere Informationen, welche Teile eines Textes falsch sind. Um in diesem Feld weiterzukommen, arbeiten die AIT-Forscher:innen an Methoden der „explainable AI“. Dabei wird analysiert, wie eine KI zu der Entscheidung „wahr“ oder „fake“ kommt – also welche Faktoren zu einer solchen Klassifikation geführt haben. In Texten sollen dabei als kritisch angesehene Wörter angezeigt werden und idealerweise Internet-Links vorgeschlagen werden, unter denen man verdächtige Passagen gegenprüfen kann.

### **Fake Shops im Netz**

Ein anderes brennendes Sicherheitsthema im Internet sind derzeit Fake Shops: Man bestellt und bezahlt eine Ware, diese kommt aber nie an. Solche betrügerischen Seiten werden immer professioneller und zugleich auch schwieriger erkennbar. Auch in diesem Bereich kann KI helfen. Mit einer Datenbank von gut beschriebenen Fake-Seiten im Netz, die das Österreichische Institut für angewandte Telekommunikation (ÖIAT) in den vergangenen Jahren in seiner Funktion als Internet-Ombudsman gesammelt hat, trainierten Entwickler:innen am AIT und bei X-Net einen Algorithmus und stellten diesen als „Fake Shop Detector“ der Öffentlichkeit zur Verfügung. Dieser kennt aktuell rund 10.000 Fake Shops und mehr als 25.000 vertrauenswürdige Online-Händler in der DACH-Region. Im Falle eines noch unbekanntem Online-Shops kommt dann die KI zum Einsatz und diese überprüft in Echtzeit, ob Ähnlichkeiten mit bereits bekannten Fake Shops vorliegen. Insgesamt 21.000 Merkmale werden dabei berücksichtigt – darunter beispielsweise die Struktur der Website oder Kommentare im Quellcode –, aus deren Kombination der Fake Shop Detector seine Empfehlungen ableitet. Das in einen Browser integrierte Tool zeigt das Ergebnis der Analyse nach dem Ampelsystem an.

### **Eindringen in Netzwerke**

Immer mehr greifen auch Versuche von Hacker:innen um sich, in fremde Computersysteme einzudringen. Das gilt auch für computergestützte Netzwerke wie beispielsweise Smart Grids. Dort sind Energietechnik und IKT miteinander verknüpft, um die Netze flexibler zu machen und dadurch mehr erneuerbare Energie einspeisen zu können, ohne die Stabilität der Netze zu gefährden. Aufgrund der vielen Sensoren und neuen Schnittstellen zwischen physischer (Kraftwerke, Stromnetz) und virtueller Welt (digitale Systeme) bringt dies aber auch neue Herausforderungen mit sich. So können beispielsweise Smart Grids zum Ziel von Cyberangriffen werden – im schlimmsten Fall droht ein Black-out.

Daher versucht man – bei kritischer Infrastruktur noch stärker als bei anderen IKT-Systemen – bereits in der Konzeptionsphase ein möglichst

a project as part of the Austrian security research program KIRAS, an AI-based fake news detection system is currently being developed that is tailored to the needs of each end user and can be implemented in the daily working environment. In practice, however, it is not sufficient to classify a whole message as true or false. Detailed information is needed about which parts of the text are false. In order to make progress in this area, the AIT researchers are working on a methods known as “explainable AI”. This analyzes how AI has come to the decision “true” or “fake” – i.e. which factors led to this classification. In this process, words in texts that are viewed as critical are to be highlighted, ideally with suggested web links where you can double-check suspicious passages.

### **Fake Online Shops**

Another very important issue online now is fake shops: you order products and pay for them, but they never arrive. Fraudulent sites of this kind are becoming more and more professional and at the same time more difficult to detect. AI can help here, too. With a database full of well-defined fake online sites that the Österreichisches Institut für angewandte Telekommunikation (ÖIAT) [Austrian Institute for Applied Telecommunications (OIAT)] has collected in its function as Internet Ombudsman in recent years, developers at AIT and X-Net trained an algorithm and have made it available to the public as a “Fake Shop Detector”. This app now recognizes around 10,000 fake shops and more than 25,000 trustworthy online retailers in the DACH region. When an online shop is not yet known to it, AI is used: it checks in real time whether there are any similarities to previously known fake shops. The process takes into account a total of 21,000 features, including, e.g., the structure of the website and comments in the source code. The Fake Shop Detector then makes recommendations based on a combination of these factors. The tool is integrated into your browser and displays the result of the analysis using a traffic light system.

### **Invading Networks**

Attempts by hackers to penetrate third-party computer systems are becoming more common. This is the case even for computer-supported networks such as Smart Grids. These link together energy technology and ICT in order to make the grids more flexible and therefore be able to feed in more renewable energy without endangering the stability of the grids. The large number of sensors and new interfaces between the physical world (power plants, power grid) and the virtual world (digital systems) also creates new challenges, however. Smart Grids may become a target for cyber attacks – the worst case scenario would be a blackout.

For this reason, the aim is – in the case of critical infrastructure to an even greater extent than in other ICT systems – to design a system that is as secure and reliable as possible already in the concept



sicheres und zuverlässiges System zu konzipieren („Security by Design“): Durch genaue Kenntnis möglicher Bedrohungen und Abwehrmaßnahmen kann man gezielten Cyberangriffen eine möglichst geringere Angriffsfläche bieten.

Doch das alleine reicht nicht aus. Denn während des Betriebs verändern sich die Systeme ständig, und dadurch ergeben sich laufend neue Angriffsflächen. Daher braucht man auch Systeme, die selbsttätig Angriffe erkennen und entsprechend reagieren können. Und hier kommt KI ins Spiel: Das gesamte System (Datenverkehr, Systemverhalten, physikalische Prozesse usw.) wird ständig durch Sensoren überwacht, und sogenannte „Intrusion-Detection-Systeme“ sollen in diesen immensen Datenfluten erkennen, wenn etwas Ungewöhnliches passiert, das vom normalen Systemverhalten abweicht. Dazu werden Methoden des „unsupervised learning“, wie zum Beispiel Clustering, eingesetzt, die in den Daten nach Anomalien suchen. Weicht ein Prozess von der üblichen Dynamik ab, erfolgt ein Alarm an den Systemadministrator. Die große Herausforderung dabei ist, dass man einerseits keine relevanten Ereignisse übersehen darf, dass aber gleichzeitig auch zu viele falsche Alarme vermieden werden müssen.

### **KI im Dienste der Verkehrssicherheit**

Verkehrsunfälle passieren gehäuft an bestimmten neuralgischen Punkten. Je früher man solche Unfallherde erkennt – am besten bevor wirklich ein Unfall passiert –, umso stärker kann man die Zahl der Schäden, Verletzungen und Todesopfer im Straßenverkehr senken. Bisher musste man sich bei der Einschätzung weitgehend auf historische Daten sowie auf Erfahrungswerte von Expert:innen verlassen. Moderne Technik verändert die Situation: Die in den vergangenen Jahren entwickelte „AIT Mobility Observation Box“ erlaubt es, Verkehrsinfrastruktur und Konfliktsituationen objektiv zu erfassen und zu bewerten. Dabei handelt es sich um eine kompakte Box, die mit einer Kamera das Verkehrsgeschehen filmt, ohne es zu beeinflussen. Durch die automatisierte Anonymisierung aller Verkehrsteilnehmer:innen wird dem Datenschutz Rechnung getragen, und es können keine Rückschlüsse auf einzelne Personen gezogen werden. Aus den Kameradaten werden Trajektorien (Bewegungsprofile) abgeleitet. Jede:r Straßennutzer:in wird detektiert, klassifiziert und alle Bewegungslinien werden aufgezeichnet. Dies ermöglicht das hochpräzise Monitoring aller Bewegungen der verschiedenen Gruppen von Verkehrsteilnehmer:innen (Fußgänger:innen, Radfahrer:innen, Pkw, Lkw, E-Scooter, usw.) in einem bestimmten Straßenabschnitt.

Diese Daten sind die Basis für die Entwicklung eines risikobasierten Bewertungsverfahrens, das auf maschinellem Lernen beruht. Daraus werden Kennzahlen für Verkehrskonflikte wie beispielsweise Konfliktschwere oder Relativgeschwindigkeiten abgeleitet, aber auch allgemeine verkehrstechnische Parameter wie Verkehrsstärken oder Geschwindigkeiten. So wird es erstmals möglich, die Sicherheit von Verkehrsinfrastrukturen nach objektiven Kriterien zu messen und dadurch vergleichbar zu machen. Das bessere Verständnis von möglichen Unfallursachen und von neuralgischen Punkten, an denen es wiederholt zu Konflikten zwischen verschiedenen Verkehrsteilnehmer:innen kommt, hilft schließlich bei der Verbesserung von Sicherheitsmaßnahmen.

phase (“security by design”): By gaining and applying precise knowledge of possible threats and defensive measures, it is possible to ensure that there are fewer targets for cyberattacks.

But this alone is not enough. The systems are constantly changing during operation, and this constantly creates new areas that could be attacked. This means that systems also need to be able to automatically recognize attacks and react accordingly. That’s where AI comes in: The entire system (data traffic, system behavior, physical processes, etc.) is constantly monitored by sensors, and “intrusion detection systems” are designed to detect when something unusual happens that deviates from normal system behavior in this immense flood of data. “Unsupervised learning” methods such as clustering that search for anomalies in the data are used for this purpose. If a process deviates from the usual dynamics, an alarm will be sent to the system administrator. The big challenge here is not to overlook any relevant event while avoiding too many false alarms.

### **AI Aiding Traffic Safety**

Traffic accidents often occur at certain critical locations. Recognizing such sources of accidents as early as possible – preferably even before any accident actually happens – increases the likelihood of reducing the number of accidents, the damage, injuries, and fatalities resulting from road traffic. Up to now, assessments had to be largely based on historical data and the experience of experts. Modern technology is changing things: The “AIT Mobility Observation Box”, which was developed over the past few years, makes it possible to record and objectively evaluate traffic infrastructure and conflict situations. It consists of a compact box that uses a camera to film what is happening to traffic, while having no influence on it. The automated anonymization of all road users means that data protection is taken into account and no conclusions can be drawn about individual persons. The camera data is used to derive trajectories (movement profiles). Each road user is detected and classified, and all movement lines are recorded. This enables the high-precision monitoring of all movements by the different groups of road users (pedestrians, cyclists, cars, trucks, e-scooters, etc.) in a specific section of road.

This data is the basis for the development of a risk-based assessment method based on machine learning. This is used as a basis for indicators for traffic incidents such as the severity of the incident or relative speeds, but also for more general traffic-related parameters such as traffic volumes or speeds. This makes it possible for the first time to measure the safety of transport infrastructures based on objective criteria and therefore make them comparable. A better understanding of possible causes of accidents and of neuralgic points at which incidents repeatedly arise between different road users ultimately helps to improve safety measures.

## Wie KI behandelnden Ärzt:innen hilft

**Es gibt kaum eine medizinische Sparte, in der nicht bereits Deep-Learning-Verfahren Einzug gehalten haben.**

Die Medizin zählt zu den frühesten Anwenderinnen der Methoden der Künstlichen Intelligenz. Diagnostische Verfahren – insbesondere bildgebende – liefern immens viele Daten, die zu überblicken der Mensch kaum in der Lage ist. Auch die Verknüpfung von Daten aus verschiedensten Bereichen und die Suche nach Mustern in diesen Datensätzen machen KI-Methoden sehr interessant für den Einsatz im Gesundheitswesen.

Ein höchst aktuelles Beispiel ist ein Algorithmus, der erkennt, welche Personen eher einen schweren Verlauf einer COVID-19-Infektion mit Bedarf an Intensivbehandlung oder Todesfolge haben werden. Die Erkrankungsverläufe sind auch für Mediziner:innen mit viel Erfahrung nur sehr schwer zu prognostizieren. Eine technische Hilfestellung könnte die Prognose für einzelne Patient:innen verbessern. Wissenschaftler:innen um David Gómez-Varela von der Division für Pharmakologie und Toxikologie der Universität Wien sowie Kolleg:innen vom Max-Planck-Institut für Multidisziplinäre Naturwissenschaften in Göttingen und aus Finnland haben einen Maschinenlern-Algorithmus namens „COVID-19 Disease Outcome Predictors“ (CODOP) konzipiert, der das Zusammenspiel vieler einzelner Parameter analysiert. Die nötigen Trainingsdaten kamen aus Spanien, den USA, Honduras, Bolivien und Argentinien. Darin enthalten waren Daten aus routinemäßigen Blutabnahmen von nahezu 30.000 Patient:innen, die zwischen März 2020 und Februar 2022 in über 150 Krankenhäusern behandelt wurden. Verglichen wurden die Werte dann jeweils damit, ob eine Patient:in Intensivbehandlung benötigte, später entlassen werden konnte oder verstarb. Berücksichtigt wurden dabei alle dominanten SARS-CoV-2-Varianten, der Impf- und der Immunitätsstatus.

Im Zentrum der Analysen des CODOP stehen zwölf Blutwerte, die bei der Spitalsaufnahme erhoben werden. Der Algorithmus schaffte es, bis zu neun Tage im Voraus, das Überleben oder den Tod von Patient:innen vorherzuberechnen. Dieses Frühwarnsystem vor schweren Verläufen ist frei verfügbar und nach Angaben der Forscher:innen einfach zu implementieren.

## How AI is Helping Doctors

**There is hardly a branch of medicine where deep learning has not yet asserted its presence.**

Medicine is one of the earliest users of artificial intelligence methods. Diagnostic procedures – especially imaging ones – deliver an immense amount of data which humans are hardly able to review. The linking of data from a wide variety of areas and the search for patterns in these data sets also make AI methods of great interest with regard to use in healthcare.

A highly topical example is an algorithm that recognizes which people are more likely to develop a severe COVID-19 infection that will require intensive care or may even result in their death. How the disease will progress is very difficult to predict, even for doctors with a lot of experience. Technical assistance could improve the prognosis for individual patients. Scientists around David Gómez-Varela from the Division of Pharmacology and Toxicology at the University of Vienna and colleagues from the Max Planck Institute for Multidisciplinary Natural Sciences in Göttingen and from Finland designed a machine learning algorithm called “COVID-19 Disease Outcome Predictors” (CODOP) which analyzes the interaction of many individual parameters. The training data required came from Spain, USA, Honduras, Bolivia, and Argentina. It included routine blood draw data from nearly 30,000 patients treated at over 150 hospitals between March 2020 and February 2022. In each case the values were compared with whether the patient required treatment in intensive care, was later discharged, or died. All of the dominant SARS-CoV-2 variants, vaccinations and the immunity status were taken into account in this process.

The CODOP analyses focus on twelve blood values that are collected on admission to the hospital. The algorithm managed to predict the survival or death of patients up to nine days in advance. This early warning system for severe cases is freely available and, according to the researchers, is easy to implement. Next, CODOP aims to predict the need for hospitalization within 24 hours for patients in primary care

In weiterer Folge soll mit CODOP die Notwendigkeit einer Krankenhauseinweisung für Patient:innen in der Primärversorgung innerhalb von 24 Stunden und die Verlegung auf die Intensivstation für bereits hospitalisierte Patient:innen innerhalb von 48 Stunden vorhergesagt werden, berichten die Entwickler:innen.

### **Zählung von sich teilenden Zellen**

Eine Unterstützung der behandelnden Ärzt:innen kann KI auch bei der Diagnose mancher Krebserkrankungen geben. Ein wichtiges Tool für die mikroskopische Einschätzung, ob ein Tumor sich im/in Patient:in ausbreiten wird, ist die sogenannte Mitosezählung. Dabei wird die Anzahl der sich gerade teilenden Zellen in Gewebeproben gezählt. Dies gibt Hinweise auf die Zellteilungsaktivität und somit auf die Bösartigkeit von Tumoren. Trotz des unbestrittenen Nutzens hat diese Untersuchungsmethode bisher einen deutlichen Nachteil: Abhängig von der zu untersuchenden Person unterscheiden sich die Ergebnisse, was zu falschen Diagnosen führen kann. Die Patholog:innen müssen nicht nur entscheiden, welcher Teil der Gewebeprobe untersucht wird, sondern sie müssen auch Mitosen von anderen, zum Teil sehr ähnlichen Strukturen, unterscheiden.

Ein internationales Forschungsteam unter der Leitung der Veterinärmedizinischen Universität Wien hat die bisherige Methode nun mithilfe von Deep Learning verbessert. Der Algorithmus, der an Gewebeproben von Hunden entwickelt wurde, ermöglichte eine Verbesserung der Standardisierung der Mitosezählung und zwar in zweifacher Hinsicht: Zum einen unterstützen die algorithmischen Vorhersagen Patholog:innen dabei, mitotische „Hotspots“ zu erkennen, und zum anderen bei der Unterscheidung von Mitosen und anderen Zellen. Damit wird die Zuverlässigkeit der Mitosezählung deutlich erhöht.

### **Innovativer Augenscanner**

Als äußerst interessantes Werkzeug hat sich KI auch bei der Interpretation von bildgebenden Diagnoseverfahren herausgestellt. Ein großer Wurf ist beispielsweise ein innovativer Augenscanner, der bei der Früherkennung von Retinaschäden und anderen neurodegenerativen Krankheiten hilft. Schon in den vergangenen Jahren wurde an der MedUni Wien eine Untersuchungsmethode etabliert, die mittels optischer Kohärenztomografie (OCT) den Augenhintergrund abtastet. Dabei werden dreidimensionale, hochauflösende Schnittbilder hergestellt, in denen Störungen, wie beispielsweise Netzhautablösungen, sehr genau diagnostiziert werden können. Bei der Analyse der ungeheuren Datenmengen sind Lernmethoden der Künstlichen Intelligenz extrem hilfreich. Früherkennung ist bei der Retinadegeneration sehr wichtig, denn die Krankheit lässt sich nicht heilen, wohl aber kann der Verlauf bei rechtzeitiger Erkennung stark gebremst werden.

Dieses Verfahren wird nun in mehrfacher Hinsicht erweitert. So arbeitet etwa das Christian-Doppler-Labor für Künstliche Intelligenz in der Netzhaut an einer Zusammenführung von OCT-Diagnosen mit weiteren

and transfer to ICU within 48 hours for already hospitalized patients, the developers report.

### **Counting Dividing Cells**

AI can also provide support for doctors in diagnosing some types of cancer. An important tool in the microscopic assessment of whether a tumor will spread in a patient is the so-called “mitosis count”. It counts the number of cells in tissue samples that are dividing at a given moment. This provides information on cell division activity and therefore on the malignancy of tumors. Despite the undisputed benefit, this examination method has had a clear disadvantage up to now: depending on the person examining the samples, the results differ, which can lead to incorrect diagnoses. The pathologists not only have to decide which part of the tissue sample is to be examined, they also have to distinguish mitoses from other structures, some of which are very similar.

An international research team led by the University of Veterinary Medicine in Vienna has now improved on the previous method with the help of deep learning. The algorithm, developed on canine tissue samples, allowed an improvement in the standardization of mitotic counting. They did this in two ways: on the one hand, the algorithmic predictions support pathologists in recognizing mitotic “hotspots”, and on the other hand in distinguishing mitoses from other cells. This significantly increases the reliability of mitotic counting.

### **Innovative Eye Scanners**

AI has also turned out to be an exceedingly interesting tool in interpreting diagnostic procedures involving imaging. One great success, for example, is an innovative eye scanner that helps in the early detection of retinal damage and other neurodegenerative diseases. In the past few years already, an examination method has been established at the Medical University of Vienna that scans the ocular fundus using optical coherence tomography (OCT). Three-dimensional, high-resolution cross-sectional images are produced in which disorders such as retinal detachments can be diagnosed very precisely. Artificial intelligence learning methods are extremely helpful in analyzing the enormous amounts of data. Early detection is very important in the case of retinal degeneration, because although the disease cannot be cured, its progress can be significantly slowed if detected in time.

This process is now being extended in many different directions. The Christian Doppler Laboratory for Artificial Intelligence in the Retina is working on merging OCT diagnoses with other data from hospital files – here, too, AI methods are key. This should enable the extraction of quantifiable biomarkers from medical image data and therefore make the diagnosis and monitoring of diseases of the retina

Daten aus Krankenhausakten – auch hier sind KI-Methoden zentral. Dies soll die Extraktion quantifizierbarer Biomarker aus medizinischen Bilddaten ermöglichen und somit die Diagnosestellung und das Monitoring von Erkrankungen der Retina präziser machen. Darüber hinaus soll die Entdeckung neuer Subtypen an Erkrankungen durch großflächige, datenbasierende Ansätze möglich werden.

Eine andere Entwicklungsrichtung wird im EU-Projekt MOON (multimodale optische Diagnostik für altersbedingte Erkrankungen des Auges und des Zentralnervensystems) eingeschlagen: Hier wird der OCT-Augenscan mit der sogenannten Raman-Spektroskopie verknüpft. Letztere Technologie spürt durch Licht Molekülschwingungen, aus dem gemessenen Spektrum erschließt sich die Gewebszusammensetzung. Biochemische Veränderungen aufgrund von krankheitsbedingten Änderungen des Stoffwechsels setzen lange vor tatsächlichen Gewebeschäden ein, welche vor allem in der Netzhaut zu unwiederbringlichem Verlust der Sehfähigkeit führen können. Durch die KI-gestützte Kombination der zwei Methoden können also zusätzlich zur Darstellung der Struktur im Auge auch Informationen auf molekularer Ebene in die Diagnose einbezogen werden – was neben der Diagnose von Augenerkrankungen und Diabetes auch der Erkennung von subtilen neurodegenerativen Erkrankungen im Frühstadium dient.

### **Suche nach ähnlichen Bildern und Diagnosen**

Ein anderes Beispiel für den Einsatz von KI bei bildgebenden Diagnoseverfahren ist das österreichische Start-up Contextflow (mit Wurzeln an der TU Wien und der Medizinischen Universität Wien), das ein Verfahren entwickelt, das ärztliche Befunde und Computertomographie-Bilder mit anderen, ähnlichen Fällen vergleicht und dadurch die Diagnose deutlich einfacher, zuverlässiger und sicherer macht. Die Basis dafür ist der Zugriff auf große medizinische Bilddatenbanken. Mediziner:innen können in dem System bestimmte Bildregionen markieren und bekommen ähnliche Fälle angezeigt, ohne mühsam händisch Archive durchstöbern zu müssen. Überdies werden die Diagnostexte, die dazu erstellt wurden, automatisch durchsucht. Einfache Fälle können auf diese Weise viel schneller als bisher bearbeitet werden. Bei schwierigen Fällen, für die dann mehr Zeit bleibt, erhöht das KI-System durch Anzeigen passender Vergleichsfälle die Sicherheit und Genauigkeit der Diagnose. Das ist wohlgerne keine Zukunftsmusik mehr: Contextflow wurde (nach zehn Jahren der Entwicklung) im Jahr 2020 offiziell vom TÜV zertifiziert und kann damit als Medizinprodukt verwendet werden.

### **Durchblick im Lungenmikrobiom**

Die Zusammenführung unterschiedlichster Daten steht auch bei einem neuen Verfahren in der Lungenheilkunde im Zentrum. Lungentransplantationen können in vielen Fällen lebensrettend sein. Doch dabei sind viele Komplikationen möglich. Eine davon ist eine chronische Verschlechterung der Lungenfunktion (CLAD/Chronic Lung Allograft Dysfunction), oft erst Jahre nach der Transplantation. Eine frühe Diagnose ist dabei entscheidend,

more precise. In addition, it should be possible to discover new subtypes of diseases as a result of large-scale, data-based approaches.

A different direction of development is being pursued in the EU project MOON (Multi-modal Optical Diagnostics for Ocular and Neurodegenerative Disease): In this project, the OCT eye scan is combined with the so-called “Raman spectroscopy”. This technology uses light to sense molecular vibrations, and the tissue composition is derived from the spectrum measured. Biochemical changes resulting from disease-related changes in metabolism occur long before actual tissue damage occurs, which can lead to irreversible loss of vision, especially in the retina. The AI-supported combination of the two methods means that, in addition to displaying the structure in the eye, it is also possible to include information at the molecular level in the diagnosis – which, in addition to diagnosing eye diseases and diabetes, also helps to detect subtle neurodegenerative diseases in the early stages.

### **Searching for Similar Images and Diagnoses**

Another example of the use of AI in diagnostic procedures that use imaging is the Austrian startup Contextflow (with roots at the Vienna University of Technology and Vienna Medical University) which is developing a procedure that compares medical findings and computer tomography images with other, similar cases, thereby making diagnoses much easier, more reliable, and safer. The basis for this is access to large medical image databases. Physicians can identify certain image regions in the system and similar cases are displayed without it being necessary to manually rummage laboriously through archives. In addition, the diagnostic texts created for the images are searched automatically. This enables simple cases to be processed much more quickly than before. In difficult cases, for which doctors thus have more time, the AI system increases the reliability and accuracy of the diagnosis by displaying suitable comparative cases. This may sound like something from the future, but in fact, it is not: Contextflow was officially certified by the TÜV in 2020 (after ten years of development) and can therefore be used as a medical product.

### **Looking into the Lung Microbiome**

The merging of a wide variety of data is also the focus of a new procedure in pulmonary medicine: in many cases, lung transplants can be life-saving. But there are many risks from potential complications. One of these is a chronic deterioration in lung function (CLAD/Chronic Lung Allograft Dysfunction) which often occurs even years after the transplant. Early diagnosis is crucial in order to be able to implement therapeutic countermeasures. Researchers at Vienna Medical University have noticed that the “lung biome” – the sum of all microorganisms living in the lungs – plays a significant role here. In concrete terms: a change in the lung microbiome provides an indication of future changes in lung function. Some of these microorganisms



um therapeutisch gegensteuern zu können. Forscher:innen der MedUni Wien haben bemerkt, dass das sogenannte Lungenmikrobioms – die Gesamtheit aller Mikroorganismen, die in der Lunge leben – dabei eine große Rolle spielt. Konkret: Eine Veränderung des Lungen-Mikrobioms gibt einen Hinweis auf künftige Veränderungen der Lungenfunktion. Zum einen werden diese Kleinstlebewesen vom/von Spender:in mitübernommen, zum Teil gelangen sie erst im/in der Empfänger:in in die Lunge. Nach der Transplantation verändert sich das Mikrobiom und passt sich an den neuen Wirt an. Dabei gibt es viele Einflussfaktoren, etwa Alter, Geschlecht, Grund- oder Vorerkrankungen. Es zeigte sich, dass aus der Analyse des Lungenmikrobioms, der Auswertung von Multiomics-Daten – etwa des Lipidoms (Fettmuster) oder des Metaboloms (Stoffwechselbausteine der Zellen) – und von klinischen Parametern mit Hilfe von Machine Learning Prognosen über die Veränderungen der Lungenfunktion abgeleitet werden können. Das ist ein Beitrag zu einer früheren Erkennung von Risikopatient:innen.

### **Automatisierte Biosignalanalyse**

KI-Technologien bieten überdies große Chancen für die Entwicklung von Verfahren, mit denen aus biomedizinischen Signalen (Vitaldaten wie EEG oder EKG, aber auch Telemedizin- oder Verhaltensdaten) medizinisch relevante Informationen extrahiert werden können. Am AIT Austrian Institute of Technology werden beispielsweise KI-gestützte Methoden zur automatischen Auswertung von Elektroenzephalogrammen (EEG) entwickelt und in die Praxis umgesetzt. Das EEG misst die elektrische Aktivität des Gehirns und ist eine kostengünstige, nicht belastende und einfach anzuwendende Untersuchungsmethode, die in der klinischen Praxis regelmäßig, etwa zur Diagnose von Epilepsie, eingesetzt wird. Die Auswertung dieser Daten ist derzeit noch ziemlich zeitintensiv. Das betrifft etwa Langzeit-EEGs, bei denen alle 24 Stunden bis zu 3000 EEG-Seiten anfallen, die durchgesehen werden müssen, oder die Erkennung sogenannter „Spikes“, die in Zusammenhang mit epileptischen Anfällen stehen. Die behandelnden Ärzt:innen können dank KI binnen Sekunden einen schnellen Überblick über das Geschehen im Gehirn der Patient:innen bekommen.

Automatisierte EEG-Analyse bietet auch Hoffnungen auf Fortschritte bei wirklich schweren medizinischen Problemen. Eines davon ist die Früherkennung von Demenz, einer weitverbreiteten neurodegenerativen Erkrankung, die für Patient:innen und Angehörige großes Leid mit sich bringt und im fortgeschrittenen Stadium einen Verlust aller geistigen Fähigkeiten bedeutet – wodurch die Patient:innen vollständig auf Betreuung angewiesen sind. Auch wenn die Krankheit bisher oft unheilbar ist, gibt es Behandlungen, die den geistigen Verfall der Patient:innen um einige Jahre hinauszögern können. Diese Methoden sind aber nur dann von Nutzen, wenn sie frühzeitig angewandt werden, sobald die allerersten Symptome auftreten. Derzeitige diagnostische Methoden wie Hirnscans, nuklearmedizinische Methoden oder die Untersuchung der Cerebrospinalflüssigkeit sind hoch invasiv, teuer oder mit einer Strahlenbelastung der Patient:innen verbunden und daher nicht für ein Screening geeignet. ✖

are taken over from the donor, while some of them only enter the lungs in the recipient. After transplantation, the microbiome changes and adapts to the new host. There are many influencing factors, such as age, gender, underlying or previous illnesses. It has been shown that with the aid of machine learning, predictions about the changes in lung function can be made based on the analysis of the lung microbiome, the evaluation of multiomics data – such as the lipidome (fat pattern) or the metabolome (metabolic building blocks of the cells) – as well as clinical parameters. This makes a contribution to the earlier recognition of patients at risk.

### **Automated Biosignal Analysis**

AI technologies also offer great opportunities for the development of processes with which medically relevant information can be extracted from biomedical signals (vital data such as EEG or ECG, but also telemedicine or behavioral data). At the AIT Austrian Institute of Technology, e.g., AI-supported methods for the automatic evaluation of electroencephalograms (EEG) are being developed and put into practice. An EEG measures the electrical activity of the brain and is an inexpensive, non-stressful and easy-to-use examination method that is regularly implemented in clinical practice, e.g., to diagnose epilepsy. The evaluation of this data is currently still quite time-consuming. This relates to long-term EEGs, in which up to 3000 EEG pages are generated every 24 hours and have to be reviewed, or the detection of the “spikes” associated with epileptic seizures. The treating physicians can use AI to gain a quick overview of the patient’s brain activity.

Automated EEG analysis also offers hope of advances in the area of really difficult medical problems. One of these is the early detection of dementia, a widespread neurodegenerative disease that causes great suffering for patients and their families and, in advanced stages, results in a loss of all mental abilities – meaning that patients require full-time care. Although the disease is often incurable at present, there are treatments that can delay the mental decline of patients for a few years. These methods, however, are only useful if applied early, as soon as the very first symptoms appear. Current diagnostic methods such as brain scans, nuclear medicine methods or the examination of the cerebrospinal fluid are highly invasive, expensive, or involve radiation exposure of the patient and are therefore an unsuitable form of screening. ✖



## Automatisierung: Optimale Prozesse und (teil)autonome Maschinen

**In Produktionsprozessen und bei der Entwicklung autonomer Maschinen und Fahrzeuge werden KI-Methoden auf vielfältige Weise eingesetzt. Ein zentraler Punkt ist dabei die Erkennung der Umgebung sowie die Interpretation einer Szene, der Maschinen ausgesetzt sind.**

Maschinen, die selbsttätig eine Tätigkeit verrichten, faszinieren die Menschheit seit eh und je. Noch dazu, wenn ihre Aktivitäten intelligent wirken – wie zum Beispiel der sogenannte „Schachtürke“, der im Frühling des Jahres 1770 den Wiener Hof begeisterte. Der österreichisch-ungarische Hofbeamte und Mechaniker Wolfgang von Kempelen zeigte Kaiserin Maria Theresia sein neuestes Werk: einen Schachisch, an dem eine menschenähnliche, türkisch gewandete Figur saß, die mit einem mechanischen Arm Schachfiguren bewegen konnte und die gegen die meisten Gegenspieler auch gewann. Man rätselte lange Zeit, wie die Maschine, in deren Innerem deutlich hörbar Zahnräder am Werk waren, dies schaffte. Der Erfinder bewahrte Zeit seines Lebens Schweigen darüber, dass es sich schlicht um eine Täuschung, einen Betrug handelte. Im Inneren des Apparats war ein menschlicher Schachspieler versteckt, der den Mechanismus bediente.

Allerdings war es für manche Zeitgenossen nicht unplausibel, dass eine Maschine über solche Fähigkeiten verfügen könnte – schließlich hatten bereits frühe Aufklärer:innen die Idee in die Welt gesetzt, dass Tiere (inklusive dem Menschen) als bloße Maschinen angesehen werden könnten, die wie ein zwar kompliziertes, aber dennoch genau definiertes Uhrwerk funktionierten. Für Menschen, die an einen vom Leib unabhängigen Geist glaubten – etwa an einen „göttlichen Funken“, der den Menschen beseelt und mit Intelligenz ausstattet – erschien dies freilich damals als unmöglich.

Dennoch wunderte sich kaum mehr jemand, als im Jahr 1996 der IBM-Computer „Deep Blue“ erstmals einen amtierenden Schachweltmeister schlagen konnte. Die Menschheit hatte in den dazwischen liegenden zwei Jahrhunderten den Aufstieg der Technik erlebt und sich an den Gedanken gewöhnt, dass Maschinen so manche Aufgabe übernehmen können, die zuvor dem Menschen vorbehalten war, und manche Dinge sogar besser

## Automation: Optimal Processes and (Semi-)Autonomous Machines

**AI methods are used in a variety of ways within production processes and in the development of autonomous machines and vehicles. A focal point is the recognition of the environment and the interpretation of a scene to which machines are exposed.**

Machines that are able to perform an activity of their own accord have always fascinated humankind. Even more so when their activities appear to be intelligent – such as the “chess-playing Turk” who inspired the Viennese court in the spring of 1770. The Austro-Hungarian court official and mechanic Wolfgang von Kempelen showed Empress Maria Theresa his latest work: a chess table at which sat a human-like figure dressed in Turkish clothing, which was able to move chess pieces with a mechanical arm and which also beat most opponents. For a long time, people puzzled over how the machine, inside which gears could clearly be heard at work, managed to do this. The inventor kept secret for his entire life that the whole thing was just a deception, a fraud. There was a human chess player concealed inside the apparatus, operating the mechanism.

For some contemporaries, however, it was not implausible for a machine to be able to actually possess such abilities – after all, early Enlightenment thinkers had already put forward the idea that animals (including humans) could be regarded as simple machines that functioned like a complicated, yet precisely defined clockwork. For people who believed in a spirit independent of the body – for example in a “divine spark” that gives people souls and equips them with intelligence – this of course seemed impossible at the time.

Nevertheless, hardly anyone was surprised when, in 1996, the IBM computer “Deep Blue” was able to beat a reigning world chess champion for the first time. Over the course of the intervening two centuries, humans had witnessed the rise of technology and had become accustomed to the idea that machines could do quite a number of tasks previously reserved for humans and could even do some things

bewältigen können. Rechenmaschinen zum Beispiel, die unermüdlich und ohne Fehler Zahlen addieren. Oder später Industrieroboter, die in vom Menschen unerreichbarer Schnelligkeit und/oder Präzision Teile montieren oder Oberflächen bearbeiten.

### **Roboter sind bereits auf dem Acker unterwegs**

Ob man solche Geräte als „intelligent“ bezeichnen kann, darüber kann man trefflich streiten. Heute gibt es jedenfalls Maschinen, die tatsächlich autonom und weitgehend unabhängig vom Menschen Aufgaben erledigen können. Ein Beispiel aus einem Bereich, den man auf den ersten Blick als nicht allzu innovativ einschätzen würde: In der Landwirtschaft sind bereits autonome Arbeitsmaschinen im Einsatz, die – mit GPS und Sensoren gesteuert – selbsttätig ihre Furchen auf Feldern ziehen. Serienreif sind auch schon kleine, vierrädrige Roboter, die als Flotte von autonomen Spezialmaschinen zusammenarbeiten. Diese Vehikel säen Samen aus, merken sich, wo eine Nutzpflanze steht, erfassen mit Sensoren (und einer Auswertung der Daten mittels KI-Methoden) Unkräuter und töten unerwünschte Pflanzen mit Hochspannungsimpulsen ab. Mit solchen Systemen lässt sich „precision farming“ betreiben: Jedes Feldstück, ja jede Pflanze, erhält dabei genau die Behandlung, die nötig ist – von Bewässerung und Düngung bis hin zur Bekämpfung von Schädlingen. Noch haben solche Farm-Roboter ihren Preis und noch sind viele Fragen ungelöst, zum Beispiel über die Nutzung der erhobenen Daten. Doch ihnen wird eine große Zukunft vorhergesagt, nicht zuletzt deshalb, weil es immer schwieriger wird, Erntehelfer:innen und qualifiziertes Personal für die Feldarbeit zu finden.

### **Qualitätsinspektion in der Industrie**

Ein anderer Einsatzbereich von KI in der Sachgüterproduktion ist „predictive maintenance“. Dabei lernen KI-Systeme einen Zusammenhang zwischen gewissen Messdaten und der Leistung von Maschinen. Wenn es zu einer größeren Abweichung vom erwarteten Systemverhalten kommt, wird dies vom System automatisch signalisiert – dann ist eine Wartung oder Reparatur fällig. Durch die Echtzeitinformationen über den Zustand der Ausrüstung können mögliche Produktionsunterbrechungen erkannt werden, bevor sie auftreten – und somit vermieden werden.

KI-Systeme werden zunehmend auch eingesetzt, um die Qualität von Industriegütern sicherzustellen bzw. zu steigern. Hier spielen moderne Inspektionssysteme eine große Rolle, wie sie beispielsweise am AIT Austrian Institute of Technology entwickelt werden. Beim sogenannten „Inline Computational Imaging“ (ICI) werden Objekte auf einem Fließband unter Hochgeschwindigkeitskameras vorbeibewegt. Für die Auswertung der Kameradaten wurden spezielle Algorithmen entwickelt, die unter anderem blitzschnell 3D-Rekonstruktionen der zu prüfenden Objekte erstellen, in denen man kleinste Defekte an der Oberfläche detektieren kann. Dabei kommen klassische Methoden der Bildverarbeitung zum Einsatz, die mehr und mehr mit KI-Verfahren ergänzt werden. Das ermöglicht beispielsweise eine Segmentierung von Rissen in den Prüfobjekten oder eine Klassifikation

better. Calculators, for example, that tirelessly add numbers without making mistakes. Or later, industrial robots that assemble parts or process surfaces with a speed and/or precision unattainable for humans.

### **Robots are already in the fields**

Whether such devices can in fact be described as “intelligent” is up for argument. In any case, there are now machines that can actually perform tasks autonomously and largely independently of humans. One example from an area that at first glance would not appear to be particularly open to innovation: in agriculture, autonomous machines are already in use, which – controlled by GPS and sensors – independently dig furrows in fields. Small, four-wheeled robots that work together as a fleet of autonomous specialist machines are also ready for series production. These vehicles sow seeds, remember where a crop is, detect weeds with sensors (and evaluate the data using AI methods) and kill unwanted plants using high-voltage pulses. Systems of this kind are used for “precision farming”: each plot of land, indeed each plant, receives exactly the treatment that is necessary – from irrigation and fertilization to pest control. Farm robots of this kind still have their price, and there are still many unsolved issues, for example about the use of the data recorded. But a great future is predicted for them, not least because it is becoming increasingly difficult to find harvest workers and qualified personnel for field work.

### **Quality inspection in industry**

Another area of application of AI in the production of goods is “predictive maintenance”. In this process, AI systems learn about a connection between certain measurement data and the performance of machines. If there is a major deviation from the expected system behavior, this is automatically signaled by the system – then maintenance or repair is due. With real-time information about the condition of the equipment, possible production interruptions can be identified even before they occur – and therefore be avoided.

AI systems are also increasingly being used to ensure or increase the quality of industrial goods. Modern inspection systems, such as those developed at the AIT Austrian Institute of Technology, play a major role here. In “Inline Computational Imaging” (ICI), objects are moved past high-speed cameras on a conveyor belt. Special algorithms were developed to evaluate the camera data, which, among other things, create lightning-fast 3D reconstructions of the objects to be inspected, in which the smallest defects on the surface can be detected. Traditional methods of image processing are used, which are increasingly being supplemented with AI processes. This enables, for example, a segmentation of cracks in the test objects or a classification of defects. In this way, the evaluation methods are successively refined and the quality of the inspection is constantly increasing.

von Fehlern. Auf diese Weise werden die Auswertungsmethoden sukzessive verfeinert und die Qualität der Inspektion immer mehr gesteigert.

### **Autonome Maschinen und Fahrzeuge**

Ein strategisches Forschungsziel des AIT ist die Entwicklung von autonomen Arbeitsmaschinen, wie Bagger, Kräne, Gabelstapler etc. Diese sollen den Menschen in seiner Tätigkeit unterstützen und schwere, gefährliche oder monotone Aufgaben übernehmen. So wurde kürzlich am Standort Seibersdorf ein Testgelände aufgebaut, in dem beispielsweise ein autonomer Verladekran für Baumstämme entwickelt und getestet wird. Die Aufgabe „Fahr zum Baumstamm, greif den Baumstamm und bring ihn zum Laster!“ ist für den Menschen eine klar definierte und (mit dem entsprechenden Gerät) leicht lösbare Aufgabe. Für Maschinen war dies bisher kaum möglich. Denn hinter dem scheinbar einfachen Befehl stecken viele komplexe Aufgaben und Forschungsfragen. Das umfasst zum Beispiel die Regelung von Hydraulikkomponenten und des mechanischen Systems, die zuverlässige Aufgaben- und Bewegungsplanung inklusive Lokalisierung der Eigenposition – auch bei sich verändernden Umgebungen –, das richtige Greifen von Objekten, die robuste Wahrnehmung der Umgebung sowie die Objektklassifizierung zur korrekten Interpretation des Umfelds, um nur einige zu nennen. Bei diesen komplexen Aufgaben helfen KI-Systeme erheblich. Noch wesentlich größer werden die Anforderungen für Roboter, wenn sie „geschützte“ und wohldefinierte Umgebungen wie etwa Fabrikhallen, Felder oder Holzverladeplätze verlassen und sich in der realen Welt zurechtfinden müssen – mit allen möglichen störenden und oft unvorhersehbaren Einflussfaktoren, die sich unmöglich alle schon im Vorfeld berücksichtigen lassen. Problematisch sind dabei insbesondere Sicherheitsaspekte: Von autonomen Maschinen dürfen niemals Gefahren für andere Gegenstände und erst recht nicht für den Menschen ausgehen. Um das zu gewährleisten, ist eine genaue Kenntnis ihrer Umgebung nötig. Dieser Punkt ist in vielen Anwendungen die wirklich „harte Nuss“, die es zu knacken gilt.

### **Knackpunkt Umgebungserkennung**

In vielen Fällen ist es sinnvoll, einen „digitalen Zwilling“ der Umgebung zu erstellen. Das kann beispielsweise mithilfe von Kamerasystemen, Radarsensoren oder Lasermessungen geschehen. Daraus kann mit mathematischen Verfahren ein hochgenaues 3D-Modell der Umgebung erstellt werden, das in einem nächsten Schritt segmentiert wird. Einzelne Objekte – etwa ein Verkehrszeichen oder ein:e Passant:in – werden klassifiziert, und ihnen werden gewisse Eigenschaften zugeordnet (etwa dass ein Verkehrszeichen einen fixen Standort hat, sich Fußgänger:innen aber bewegen). Dabei leisten Machine-Learning-Methoden wertvolle Dienste, die zum Beispiel gelernt haben, ein Verkehrszeichen zu identifizieren, diese Information mit anderen zu verknüpfen, um in der Folge zu einem Szenenverständnis zu gelangen, auf dessen Basis Entscheidungen getroffen werden können.

### **Autonomous machines and vehicles**

A strategic AIT research goal is the development of autonomous working machines, such as excavators, cranes, forklifts, etc. These are intended to support people in their work and take on difficult, dangerous, or monotonous tasks. A test site was recently set up at the Seibersdorf location, where e.g., an autonomous loading crane for tree trunks is being developed and tested. The task “Drive to the tree trunk, grab the tree trunk and bring it to the truck!” is a clearly defined task which (using the appropriate equipment) can be easily solved by humans. So far, this had been nearly impossible for machines. The reason: behind the seemingly simple command there are many complex tasks and research questions. This includes, e.g., controlling hydraulic components and the mechanical system, reliable task and movement planning including localization of one’s own position – even in changing environments – properly gripping objects, sound perception of the environment, and object classification for the correct interpretation of the environment, to name just a few. AI systems can help considerably with these complex tasks. The requirements for robots become even greater when they leave the “protected” and well-defined environments such as factory buildings, fields or timber loading bays and have to find their way in the real world – with all sorts of disruptive and often unpredictable influencing factors that cannot possibly all be taken into account in advance. Safety aspects in particular are problematic here: autonomous machines must never pose a threat to other objects, and certainly not to humans. In order to ensure this, precise knowledge of their surroundings is required. In many applications, this point is the really “tough nut” to be cracked.

### **Environmental detection is the crucial point**

In many cases it makes sense to create a “digital twin” of the environment. This can be done, e.g., with the help of camera systems, radar sensors, or laser measurements. From this, a high-precision 3D model of the environment can be created using mathematical processes which is segmented in the next step. Individual objects – such as a traffic sign or a pedestrian – are classified and assigned certain properties (such as a traffic sign having a fixed location, but pedestrians are moving). Machine learning methods provide valuable services here; they have learned to identify a traffic sign, to link this information with other data in order to subsequently gain an understanding of the scene, on which basis decisions can then be made.

Bei autonomen Fahrzeugen ist diese Umgebungserkennung die Basis für die Bewegungsplanung und die Steuerung des Autos. Laut Expert:innen ist die richtige Interpretation des Umgebungsmodells das schwächste Glied beim autonomen Fahren. Eine fehlerhafte Klassifizierung, etwa eines entgegenkommenden Fahrzeugs oder eines querender Fußgänger:innen kann desaströse Konsequenzen haben – wie einige berühmt gewordene Unfälle von Testautos namhafter Konzerne zeigen.

### Testen und Validieren

Das Hauptproblem dabei ist, dass Machine-Learning-Systeme immer nur Wahrscheinlichkeitsaussagen liefern – und das ist mit sicherheitskritischen Anwendungen wie dem autonomen Fahren, bei denen man eine möglichst hohe Sicherheit verlangen muss, nur schwer vereinbar. Daher werden an das Testen und Validieren solcher Systeme besonders hohe Anforderungen gestellt. Solche Methoden werden unter anderen in der Gruppe „Dependable Systems Engineering“ am AIT entwickelt. Dabei geht es im Kern darum, zu verifizieren, ob eine KI richtig arbeitet. Um das herauszufinden, setzt man ein autonomes Fahrsystem beispielsweise in einer Simulationsumgebung einer kritischen Situation mit anderen Autos oder Fußgänger:innen aus und kann dadurch bei genau definierten Rahmenbedingungen testen, ob das System im Fahrzeug richtig reagiert.

Je komplexer die zu prüfenden Systeme werden, umso mehr steigt der Aufwand für das Testen; die klassischen Prüfverfahren funktionieren dabei immer schlechter. Daher ist ein zweites große Forschungsthema die Nutzung von KI, um Verifizierungs- und Testtechniken zu verbessern. Wichtige Stoßrichtungen bei der Weiterentwicklung sind es, die Verfahren schneller zu machen oder auftretende Fehler besser eingrenzen und erklären zu können. Die Optimierung solcher Tests ist auch hinsichtlich der Abwägung von Kosten und Nutzen wichtig: In manchen Fällen machen Testen und Verifizieren 50 bis 70 Prozent des gesamten Entwicklungsaufwands aus. ✖

In autonomous vehicles, this recognizing the environment forms the basis for planning the movements and controlling the car. According to experts, the correct interpretation of the environment model is the weakest link in autonomous driving. An incorrect classification of an oncoming vehicle or a pedestrian crossing can have disastrous consequences – as some famous accidents involving test cars from well-known companies have shown.

### Testing and validation

The main problem with this is that machine learning systems only provide probability statements – and this is difficult to reconcile with safety-critical applications such as autonomous driving which require the highest possible level of safety. For this reason, particularly high requirements are placed on the testing and validation of these systems. Such methods, along with others, are developed at AIT in the “Dependable Systems Engineering” group. The main focus is on verifying whether the AI is working correctly. To determine this, an autonomous driving system is exposed to a critical situation with other cars or pedestrians in a simulation environment and can therefore perform a test under precisely defined framework conditions to determine whether the system in the vehicle reacts correctly. The more complex the systems to be tested become, the greater the effort required for testing; traditional test methods are working less and less well in such situations. Therefore, a second major research topic is the use of AI to improve verification and testing techniques. Important directions in further development aim at making the processes faster and being better able to isolate and explain occurring errors. Optimizing tests of this kind is also important for balancing costs and benefits: in some cases, tests and verification may account for up to 50% or even 70% of the whole development expense. ✖

## High-Tech Stadtplanung

**Am City Intelligence Lab des AIT Austrian Institute of Technology wurde das System InFraReD (Intelligent Framework for Resilient Design) entwickelt – eine auf KI basierende Plattform für intelligente und resiliente Stadtplanung. Diese liefert in Echtzeit ein Feedback über die Auswirkungen von Planungen.**

In der Stadtplanung gilt es, unzählige Parameter zugleich im Blick zu haben, um die Konsequenzen von baulichen Maßnahmen bestmöglich auszutarieren und ein Optimum für Planer:innen, Bauherr:innen, Nutzer:innen und das Gemeinwohl herauszuholen. Solche Parameter sind beispielsweise die Erreichbarkeit von Gebäuden und Stadtvierteln, die Fußläufigkeit, die Entfernung zu öffentlichen Verkehrsmitteln oder zu Naherholungsgebieten, aber auch die Behaglichkeit von Stadtvierteln, der thermische Komfort (Stichwort: sommerliche Überhitzung), die Anfälligkeit bei Sturm, die Besonnung der Gebäude (das Solarpotenzial ist entscheidend für eine künftige Nutzung als Photovoltaik-Flächen), Lärmbelastung oder Überflutungsgefahr.

Am City Intelligence Lab (CIL) am AIT Austrian Institute of Technology wurde in den vergangenen Jahren ein System aufgebaut, das die Planung mit Hilfe modernster Technologien wie etwa Virtual und Augmented Reality oder Künstlicher Intelligenz wesentlich vereinfacht. Per Mausclick oder Fingerzeig können Gebäude vergrößert oder verkleinert werden, an einen anderen Platz verschoben oder gedreht werden – und die Konsequenzen solcher Umplanungen werden praktisch in Echtzeit berechnet und dargestellt. Auf diese Weise ist ein partizipativer Planungsprozess realisierbar, bei dem Stakeholder und Interessensgruppen gemeinsam Varianten ausprobieren können und sofort über die Folgen informiert werden. Herkömmlicherweise sind solche Rückkopplungsschleifen sehr zeitintensiv, sodass kaum eine wirkliche „Co-Creation“ zustande kommen kann. Im CIL dagegen können vom frühesten Planungsstadium an vielfältige Szenarien entworfen und sofort evaluiert werden.

## High-Tech Urban Planning

**The InFraReD system (Intelligent Framework for Resilient Design) was developed at the City Intelligence Lab at the AIT Austrian Institute of Technology – a platform for intelligent and resilient urban planning based on AI. This system delivers real time feedback on the effects of different planning options.**

Urban planning needs to keep an eye on countless parameters at the same time in order to balance the consequences of construction measures in the best possible way and to achieve the best results for planners, builders and users, while always working in the public interest. These parameters include, for example, the accessibility of buildings and districts, which amenities are located within walking distance, the distance to public transport or to local recreational areas, but also how comfortable urban areas are, comfort from the point of view of temperature (with the key aspect here being overheating in summer), susceptibility to storms, the amount of sunlight on the buildings (the solar potential is crucial for future use as photovoltaic areas), noise pollution, or the risk of flooding.

At the City Intelligence Lab (CIL) at the AIT Austrian Institute of Technology, a system has been created in recent years that significantly simplifies planning using state-of-the-art technologies such as virtual and augmented reality and artificial intelligence. Buildings can be made larger or smaller, moved to a different location or rotated just by mouse click or touch – and the consequences of such replanning are calculated and displayed practically in real time. This enables the implementation of a participatory planning process in which stakeholders and interest groups can try out different versions together and are immediately informed of the consequences. Traditionally, feedback loops of this kind are very time-consuming, meaning that real “co-creation” is difficult to achieve. At the CIL, on the other hand, a variety of scenarios can be designed and evaluated immediately, beginning at the earliest planning stages.



Damit diese interaktive Arbeitsweise gelingen kann, ist es entscheidend, dass die Konsequenzen von Planungsänderungen rasch – also binnen Sekunden – berechnet und dargestellt werden. Das ist bei vielen Parametern möglich. Doch manche Kenngrößen erfordern extrem aufwendige Berechnungen. Ein Beispiel dafür ist der Wind: Mit sogenannten „Computational Fluid Dynamics“-Simulationen (CFD) kann der Einfluss von Gebäuden, Plätzen oder Straßenzügen auf das Windgeschehen sehr genau vorherberechnet werden. Allerdings dauert das sehr lange, nämlich Stunden bis Tage. Das Windgeschehen um Gebäude herum ist aber für viele Standortfaktoren eine entscheidende Größe – etwa für den Komfort und die Sicherheit von Fußgänger:innen bei Sturm, für die thermische Behaglichkeit, für das sommerliche und winterliche Mikroklima oder für die lokale Luftverschmutzung.

Hier kommt Künstliche Intelligenz ins Spiel: Im Projekt InFraReD (Intelligent Framework for Resilient Design) wurde ein Deep-Learning-Modell darauf trainiert, das Ergebnis einer CFD-Simulation für einen bisher nicht berechneten Fall vorherzusagen. Dazu wurde ein Generative Adversarial Neural Network (GAN) aufgebaut, das mit Tausenden von CFD-Simulationen aus einem digitalen 3D-Modell der Stadt Wien trainiert wurde. In der Folge war das System in der Lage, die Windgeschwindigkeit und -richtung in Kopfhöhe (1,75 Meter über dem Boden) für jedes neue Design, das entworfen wird, vorherzusagen. Das dauert in der Praxis nur wenige Millisekunden – das ist 10.000-mal schneller als eine komplette CFD-Simulation. Die Genauigkeit ist mit 80 bis 95 Prozent zwar etwas geringer, aber für eine erste Abschätzung, ob eine Planungsvariante weiterverfolgt werden kann und soll, reicht das auf jeden Fall aus. Um eine exakte Berechnung der am Ende ausgetüftelten Variante führt dann zur Evaluierung der Annahmen ohnehin kein Weg vorbei. Ähnliches wie bei Wind gilt auch für die Sonneneinstrahlung – und weitere Faktoren sind in Arbeit. ✖

In order for this interactive manner of working to succeed, it is crucial for the consequences of planning changes to be calculated and displayed quickly – i.e. within seconds. This is now possible with many parameters. Some parameters, however, require extremely complex calculations. An example of this is wind: “Computational Fluid Dynamics” (CFD) simulations enable the influence of buildings, squares, or streets on wind to be calculated very precisely. This does take a very long time, however – hours, or even days. The wind around buildings is a decisive variable in many factors regarding location – for example regarding the comfort and safety of pedestrians in storms, comfort from a temperature perspective, the summer and winter microclimate, and local air pollution.

This is where artificial intelligence comes into play: As part of the InFraReD (Intelligent Framework for Resilient Design) project, a deep learning model was trained to predict the result of a CFD simulation for a case for which no calculations had yet been done. For this purpose, a Generative Adversarial Neural Network (GAN) was created which was trained with thousands of CFD simulations from a digital 3D model of the city of Vienna. The system is now able to predict the wind speed and direction at head height (1.75 meters above the ground) for each new design that is created. In practice, this only takes a few milliseconds – that’s 10,000 times faster than a complete CFD simulation. Although the level of accuracy is 80% to 95% and thus somewhat lower than with a comprehensive simulation, it is definitely sufficient for an initial assessment of whether a planning variant can and should be pursued. For a precise evaluation of the final variant, the assumptions must be computed comprehensively anyway, but until then, much time is gained. The same thing can be done for sun exposure as for wind – and work is currently underway to add even more features. ✖

# Wissenschaft bekommt neue Einsichten

**KI-Methoden erlauben in vielen Fachdisziplinen das Erkennen neuer Zusammenhänge. Die Kooperation zwischen Mensch und Maschine kann große Fortschritte bewirken.**

Innerhalb weniger Jahre haben die Methoden der Künstlichen Intelligenz die Welt der Wissenschaft stark verändert. Im Zuge der Erforschung von biochemischen Vorgängen, quantenphysikalischen Phänomenen oder soziologischen Zusammenhängen fallen immens viele Daten an. Diese sind die Basis, um Fragestellungen, die als Hypothesen formuliert werden, zu beantworten – etwa indem Muster in den empirischen Daten gefunden werden oder mathematische Modelle mit den Daten gestützt und verfeinert werden.

Vor allem die Entwicklung des Deep Learning hat hier einen großen Fortschritt gebracht, erlaubt diese Technik doch die Analyse immens großer Datensätze, die kein Mensch jemals überblicken könnte. Zum einen können die künstlichen neuronalen Netze und die zugehörigen Algorithmen Regelmäßigkeiten in den Daten finden – diese können einen Hinweis auf einen bisher noch nicht entdeckten Zusammenhang geben, auf dessen Basis dann ein neues Modell konzipiert wird. Zum anderen können an Daten geschulte KI-Systeme Vorhersagen für bisher ungelöste Probleme der Wissenschaft machen (die dann ihrerseits wieder einer Bewertung durch die Wissenschaftler:innen bedürfen).

## Faltung von Proteinen

Ein prominentes Beispiel dafür ist der Algorithmus „AlphaFold“, der von Google Deepmind entwickelt wurde und von der Wissenschaftszeitung „Science“ zum „breakthrough of the year 2021“ gekürt wurde. Dabei geht es um die räumliche Struktur von Proteinen. Diese bestehen aus einer Kette von Aminosäuren, die sich in Abhängigkeit von den Umgebungsbedingungen selbstorganisiert zu einer dreidimensionalen Struktur faltet. Die räumliche Anordnung der einzelnen Atome ist dabei entscheidend für die biologische Funktion – ob ein Protein beispielsweise als Katalysator für eine bestimmte chemische Reaktion dienen kann oder an ein anderes Molekül bindet und dadurch eine Krankheit heilt (oder zumindest Symptome

# New Scientific Insights

**AI methods are making it possible to recognize new connections in many specialist disciplines. Cooperation between humans and machine has the potential to bring about great progress.**

Within the space of just a few years, methods of artificial intelligence have made significant changes in the world of science. When researching biochemical processes, quantum physical phenomena or sociological relationships, immense amounts of data are generated. These are the basis for answering questions phrased as hypotheses – for example by finding patterns in the empirical data or by supporting and refining mathematical models with the data.

The development of deep learning in particular has brought great progress in this area, because this technology allows immensely large data sets to be analyzed where no human could ever gain a practical overview. On the one hand, the artificial neural networks and the associated algorithms can find regularities in the data – these can provide an indication of a previously undiscovered connection, on the basis of which a new model can then be designed. On the other hand, AI systems trained on data can make predictions for previously unsolved scientific problems (which then in turn require evaluation by scientists).

## Folding Proteins

A well-known example of this is the “AlphaFold” algorithm which was developed by Google Deepmind and was named “Breakthrough of the Year 2021” by the scientific journal “Science”. It relates to the spatial structure of proteins. These consist of a chain of amino acids, which folds into a three-dimensional structure in a self-organized manner depending on the environmental conditions. The spatial arrangement of the individual atoms is decisive with regard to the biological function – whether a protein can serve, for example, as a catalyst for a certain chemical reaction or binds to another molecule and thereby cures a disease (or at least combats symptoms). So if you are able to predict the spatial structure, you would be able to design proteins in a targeted manner for a specific purpose.

bekämpft). Wenn man also die räumliche Struktur vorhersagen kann, könnte man gezielt Proteine für einen bestimmten Zweck designen. Das Problem dabei: Für eine konkrete Aminosäurekette gibt es viele denkbare Konfigurationen, dass man Jahrtausende brauchen würde, um alle Möglichkeiten durchzuprobieren. AlphaFold hat dieses Problem nun mithilfe von Deep Learning in bisher ungeahnter Präzision gelöst. In einer Trainingsphase erlernte der Algorithmus anhand Tausender evolutionär verwandter Proteine, von denen sowohl die Sequenzen als auch die 3D-Form bekannt ist, bestimmte Muster. Diese werden dann dazu genutzt, um die Abstände zwischen Aminosäuren vorherzusagen; berücksichtigt werden dabei physikalische und geometrische Einschränkungen für die Faltung. Darüber hinaus wird – ähnlich wie bei der maschinellen Sprachverarbeitung – ein Aufmerksamkeitsmechanismus eingesetzt, der die verarbeiteten Daten in einen Kontext setzt. Dadurch wird es möglich, neue Fragmente von Proteinen zu erfinden, die dazu verwendet werden, die Genauigkeit der vorgeschlagenen Proteinstruktur zu verbessern.

AlphaFold ist zur Gänze als Open-Source-Software veröffentlicht und kann von der globalen Wissenschafts-Community frei genutzt werden. So wurde AlphaFold, um nur ein Beispiel zu nennen, kürzlich in das von AIT-Forscher Ivan Barisic im EU-Projekt MARILIA entwickelte Softwarepaket namens CATANA integriert, mit dem komplexe Biomoleküle (Proteine und DNA) dreidimensional modelliert, manipuliert und visualisiert werden können.

### Suche im Teilchen-Zoo

Auch Teilchenphysiker:innen nutzen dankbar die neuen Möglichkeiten, die ihnen KI bietet. So verwenden etwa Physiker:innen der Österreichischen Akademie der Wissenschaften (ÖAW) künstliche neuronale Netze, um in den Daten von Teilchenbeschleunigern nach bisher unbekanntem Teilchen zu suchen. Die Analyse dieser immens großen Datensätze wird dadurch erschwert, dass es viele Störeinflüsse gibt und schon minimalste Abweichungen des Winkels bei der Kollision zweier Teilchen zu komplett anderen Ergebnissen führen können; daher müssen sehr viele Teilchenzusammenstöße beobachtet werden, um durch statistische Untersuchungen festzustellen, ob möglicherweise ein unbekanntes Teilchen entstanden ist. Eine Forschergruppe um Gianluca Inguglia, der das ERC-Projekt „InterLeptons“ leitet, hat nun ein neues Werkzeug namens „Punzi-Nets“ entwickelt, das diesen Prozess durch den Einsatz von maschinellem Lernen deutlich beschleunigt. Dabei handelt es sich um ein System, das in einem ersten Schritt lernt, Signal und Rauschen zu unterscheiden, und berechnet, wie ein mögliches Signal für ein bestimmtes Teilchen aussehen könnte. In einem zweiten Schritt wird das Punzi-Net optimiert, indem es lernt, ein vom italienischen Physiker Giovanni Punzi entwickeltes Qualitätskriterium für Hochenergiephysik möglichst gut zu erfüllen. Das System lernt auf diese Weise, verschiedene mögliche Massen eines Teilchens auf einmal zu prüfen, sogar solche, die nicht in den Trainingsdaten enthalten waren. Unterm

The problem: there are so many conceivable configurations for a specific amino acid chain that it would take thousands of years to attempt all the possibilities. AlphaFold has now solved this problem with unprecedented precision using deep learning. In a training phase, the algorithm learned certain patterns from thousands of evolutionarily related proteins, for which both the sequences and the 3D shape are known. These are then used to predict the distances between amino acids; physical and geometric limitations for the folding are taken into account. In addition – similar to machine language processing – an attention mechanism is used that puts the processed data into context. This makes it possible to invent new fragments of proteins that are used to improve the accuracy of the proposed structure of the protein.

AlphaFold was released as 100% open source software that can be used freely by the global scientific community. In this way, AlphaFold, to give just one example, was recently integrated into the CATANA software package developed by AIT researcher Ivan Barisic in the EU project MARILIA which models, manipulates, and 3D-visualizes complex biomolecules (proteins and DNA).

### Searching Through a Sea of Particles

Particle physicists are also grateful for the new opportunities offered by AI. For example, physicists at the Austrian Academy of Sciences (ÖAW) use artificial neural networks to search for previously unknown particles in the data from particle accelerators. The analysis of these immensely large data sets is made more even difficult by the fact that there are many disturbing influences and even the smallest deviations in the angle when two particles collide can lead to completely different results; for this reason, a large number of particle collisions must be observed to determine through statistical investigations whether an unknown particle may have been created. A research group led by Gianluca Inguglia, who heads the ERC project “InterLeptons”, has now developed a new tool called “Punzi-Nets” that significantly accelerates this process by using machine learning. The first step in this is a system that learns to distinguish between signal and noise and calculates what a possible signal for a specific particle might look like. In the second step, the Punzi-Net is optimized by learning how to fulfill a quality criterion for high-energy physics developed by the Italian physicist Giovanni Punzi as closely as possible. In this way, the system learns to check different possible masses of a particle at once, even those that were not included in the training data. The bottom line is that the search for new particles is accelerated enormously with the new tool – especially when the exact mass of a particle is not known. It is intended that Punzi-Net will now be used to search for dark matter.

Strich wird mit dem neuen Werkzeug die Suche nach neuen Teilchen enorm beschleunigt – vor allem dann, wenn die Masse eines Teilchens nicht genau bekannt ist. Eingesetzt werden soll Punzi-Net nun zum Beispiel für die Suche nach dunkler Materie.

### Was die Alten Griechen schrieben

Ein weiteres Beispiel für die Anwendung von KI aus einem ganz anderen Wissenschaftsfeld nennt sich „Ithaca“. Schon der Name – die Heimatinsel des griechischen Helden Odysseus – deutet auf das Gebiet hin, in dem das System eingesetzt wird. Die Erforschung der Antike stützt sich unter anderem auf die Epigraphik, die sich mit dem Studium von Inschriften auf dauerhaften Materialien wie etwa Stein beschäftigt. Tausende solcher Inschriften, die etwas über das Denken, die Sprache, die Gesellschaft und die Geschichte vergangener Zivilisationen aussagen, haben sich über die Zeit erhalten. Doch in vielen Fällen sind sie beschädigt – häufig so stark, dass sie unleserlich wurden. Überdies wurden sie oft weit von ihrem ursprünglichen Standort weg transportiert, und auch ihr Entstehungsdatum ist häufig unsicher. Ithaca, entwickelt von einem internationalen Forscherkonsortium (einmal mehr unter Einbindung von Deepmind), soll nun Abhilfe schaffen.

Das System ist ein tiefes neuronales Netzwerk für die gleichzeitige textliche Restaurierung, geografische Zuordnung und chronologische Zuweisung von altgriechischen Inschriften. Trainiert wurde es anhand von 78.000 Inschriften, die in einem mühsamen Prozess mit Metadaten über den Ort (84 Regionen) und die Zeit (zwischen 800 vor und 800 nach Chr.) der Anfertigung verknüpft wurden. Zum Einsatz kam auch die Technik von Aufmerksamkeitsmechanismen, um den Einfluss verschiedener Teile der Eingabe (zum Beispiel Zeichen und Wörter) auf den Entscheidungsprozess des Modells abzuwägen. Am Ende der Berechnungsprozesse schlägt Ithaca eine Ergänzung der fehlenden Zeichen, einen Entstehungsort und eine Entstehungszeit vor. Genauer: Das System liefert eine Liste von möglichen Antworten – unter Angabe der jeweiligen Wahrscheinlichkeiten. Unterm Strich kann „Ithaca“ nach Angaben der Forscher:innen Inschriften mit einer Genauigkeit von 71 Prozent ihrem ursprünglichen Standort zuordnen und sie mit weniger als 30 Jahren Bandbreite auf ihren tatsächlichen Entstehungszeitpunkt datieren. Dadurch können so manche Schlüsseltexte des klassischen Griechenlands neu gelesen und datiert werden.

Ithaca soll, wie die Entwickler:innen betonen, nicht die Historiker:innen ersetzen, sondern ihnen ein starkes Werkzeug in die Hand geben, das sie beim Lösen der historischen Puzzles unterstützt. Während Ithaca allein bei der Restaurierung beschädigter Texte eine Genauigkeit von 62 Prozent erreicht, verbesserte die Verwendung des Systems durch Historiker:innen die Genauigkeit von 25 auf 72 Prozent – was den Synergieeffekt dieses Forschungswerkzeugs bestätigt. ✖

### What the Ancient Greeks Wrote

Another example of the application of AI that comes from a completely different scientific field is known as “Ithaca”. The very name – the home island of the Greek hero Ulysses – indicates the area in which the system is used. Research into antiquity is based, among other things, on epigraphy, which deals with the study of inscriptions on durable materials such as stone. Thousands of inscriptions of this kind tell us something about the thinking, language, society, and history of past civilizations and have survived over time. But in many cases, they are damaged, often so much as to have become illegible. Moreover, they were often carried far away from their original location, and their date of origin is often uncertain. Ithaca, developed by an international consortium of researchers (once again with the involvement of Deepmind), is now to provide a solution to the problem.

The system is a deep neural network for the simultaneous textual restoration, geographic mapping, and chronological allocation of ancient Greek inscriptions. It was trained using 78,000 inscriptions, which were painstakingly linked to metadata regarding the location (84 regions) and date (between 800 BC and 800 AD) of their origin. The technique of attention mechanisms was also used to weigh up the influence of different parts of the input (e.g. characters and words) on the model’s decision-making process. At the end of the analysis process, Ithaca makes a suggestion about the missing characters, a place of creation, and a time of creation. To be more precise: the system provides a list of possible answers – with details of the probabilities of each. All in all, according to the researchers, Ithaca can assign inscriptions to their place of origin with an accuracy of 71% and date them to their actual time of creation within a range of less than 30 years. As a result, it has been possible to reread and date some key ancient Greek texts.

As the developers emphasize, Ithaca is not intended to replace historians, but to provide them with a powerful tool that supports them in solving historical puzzles. While Ithaca on its own achieves an accuracy of 62% in restoring damaged texts, historians’ use of the system improved accuracy from 25% to 72% confirming that this research tool provides synergy effects. ✖

## „KI ist kein Wundermittel“

**Der Data Scientist und KI-Forscher Ross King über die Kunst, ein neuronales Netzwerk zu trainieren, Daten sinnvoll aufzubereiten und mit den Ergebnissen von KI-Systemen richtig umzugehen. KI liefert immer nur Wahrscheinlichkeitsaussagen – und das passt mit sicherheitskritischen Anwendungen nicht zusammen.**

**Warum boomen Data Science und Künstliche Intelligenz (KI) gerade jetzt so?**

**Ross King:** Wenn immer mehr Prozesse digitalisiert werden, heißt das, dass mehr digitale Daten gesammelt werden. Und je mehr Daten wir haben, umso größeren Bedarf gibt es, diese Daten zu verstehen und daraus Informationen abzuleiten, Wissen zu gewinnen und vielleicht auch Empfehlungen zu generieren. Eng damit zusammen hängen die Treiber der KI: Auf der einen Seite gibt es sehr viele annotierte Daten – dadurch hat zum Beispiel die Bildanalyse große Fortschritte gemacht. Auf der anderen Seite stehen GPU-Architekturen, die nicht nur für Computerspiele gut geeignet sind, sondern auch für Berechnungen im Bereich Machine Learning und insbesondere für neuronale Netze sehr hilfreich sind. Das heißt, wir können immer mehr Daten einbeziehen und immer größere Modelle erzeugen. Überdies gibt es große Open-Source-Initiativen wie beispielsweise von Google, die das TensorFlow Framework für Deep Learning geteilt haben. Es gibt die Tendenz, dass viele Modelle veröffentlicht werden, wodurch man auf Basis von „Transfer Learning“ schnell und auch ohne viele Daten neue Modelle erzeugen kann. Da gibt es sehr viele synergistische Effekte.



**ROSS KING** leitet die Competence Unit „Data Science & Artificial Intelligence“ am AIT Austrian Institute of Technology. Er begann seine wissenschaftliche Karriere an der Stanford University (USA), forschte am Institute for Medium Energy Physics (IMEP) der ÖAW und war im Management bei UMA Information Technology AG und Alpha Thinx Mobile Phone Services GmbH tätig. Seit 2003 arbeitet er am AIT, wo er u.a. an digitalen Bibliotheken, semantischen Systemen oder digitaler Langzeitarchivierung forschte. Seit 2014 beschäftigt er sich stark mit der Forensik von Kryptowährungen, ein weiterer aktueller Schwerpunkt ist KI als wesentliches Instrument moderner Data Science.

Practical applications

## “AI is not a magic bullet”

**ROSS KING** is Head of Competence Unit “Data Science & Artificial Intelligence” at the AIT Austrian Institute of Technology. He began his scientific career at Stanford University (USA), did research at the Institute for Medium Energy Physics (IMEP) of the OeAW and worked in management at UMA Information Technology AG and Alpha Thinx Mobile Phone Services GmbH. He has been working at AIT since 2003, where he has done research on topics such as digital libraries, semantic systems, and digital long-term archiving. Since 2014, he has been heavily involved in the forensics of cryptocurrencies. Another current focus is AI as an essential tool of modern data science.

**Data scientist and AI researcher Ross King about the art of training neural networks, meaningful data analysis, and the right way to deal with the AI systems results. AI always only delivers probability statements – which does not go together well with safety-critical applications.**

**Why is there such a boom of data science and artificial intelligence (AI) right now?**

**Ross King:** If more and more processes are becoming digitalized it means more digital data are being collected. And the more data we have, the more need there is to understand this data and derive information from it, gain knowledge, and perhaps also generate recommendations. This closely relates to what drives AI: on the one side, there are large amounts of annotated data – which has led to great progress being made, for example, in image analysis. On the other, there are GPU architectures available which are not only ideal for computer games but also are very instrumental for computations in the area of machine learning and in particular for neural networks. It entails that we can keep incorporating ever more data and generate ever larger models. Also, there are major Open Source initiatives, like the one by Google who have shared their TensorFlow framework for Deep Learning projects. There is a tendency to release many models to the public domain, which enables generating new models quickly and even without large amounts of data on the basis of “Transfer Learning”. There are very many synergy effects here.

Praktische Anwendungen



**Obwohl viele Menschen glauben, dass KI noch Zukunftsmusik ist, findet die Technologie bereits in immer mehr Bereichen unseres Alltagslebens Eingang. Woran merkt man das?**

Man merkt es vor allem an unseren Smartphones. Täglich konfrontiert ist man zum Beispiel mit der automatischen Ergänzung von Wörtern, wenn man eine Textnachricht eintippt. Dieses „predictive typing“ beruht auf einem Sprachmodell, das eine statistische Auswertung vornimmt, welches Wort am wahrscheinlichsten als nächstes kommt. Ein weiteres Beispiel ist, wenn man mit Biometrie ein Handy entsperrt. Technisch gesehen liegt die Gesichtserkennung und der Vergleich mit bestehenden Bildern an der Grenze zwischen Signalanalyse und KI – wobei ich mir ziemlich sicher bin, dass Machine Learning mittlerweile auch Teil dieser Algorithmen ist. Auch viele Filter für Bildmaterial beruhen heutzutage auf KI und auf vielen Online-Plattformen ist sie bereits allgegenwärtig. So werden Inhalte von Facebook beispielsweise automatisiert auf unerwünschte Inhalte wie etwa Pornografie oder Hate Speech durchsucht. Den Einsatz von KI erkennt man als Benutzer:in nicht, aber jeder, der Facebook verwendet, wird täglich mit solchen Filtern konfrontiert. Auch Recommender-Systeme sind in unsere Leben eingedrungen. Am Anfang waren diese Systeme eher auf Regeln basiert – man hat versucht, Regeln in bestimmten Domänen zu finden und einen Entscheidungsbaum aufzubauen. Mit Hilfe von Machine-Learning-Systemen versucht man heute, diese automatisch zu finden.

Dabei kann man zumindest teilweise nachvollziehen, warum eine bestimmte Empfehlung getroffen wurde. In regelbasierten Systemen ist das noch leicht verständlich, denn die Regeln werden vom menschlichen Verstand gemacht. In Machine-Learning-Systemen sind das manchmal künstliche Attribute, die nicht mehr so einfach verstehbar sind. In größeren neuronalen Netzwerken hingegen, wo es Millionen von Knoten und Kanten gibt, kann der menschliche Verstand das nicht mehr begreifen. Dann muss man im Nachhinein versuchen, das Verhalten solcher Systeme zu verstehen. An solchen Verfahren der „Explainable AI (XAI)“ wird gearbeitet.

**Gibt es solche erklärenden Systeme, die einen Blick in die „black box“ einer KI erlauben, schon in der Praxis?**

Mein Wissen ist natürlich beschränkt, aber mir ist kein Produkt bekannt, das als Teil der KI-Verfahren auch eine Erklärbarkeit der Entscheidung liefert. Es gibt aber sicher Bedarf daran. Zum Beispiel bei Anwendungen in der Industrie: Der Einsatz von KI-Technologien hängt auch davon ab, dass man den Systemen vertrauen und Entscheidungen nachvollziehen kann. Das ist auch eine Frage der Haftung. Ich denke, dass das langsam aber sicher auch die Endkonsument:innen verlangen werden. Derzeit ist es ja eher so, dass die

**Although many believe that AI still is a thing of the future, the technology is already making its way into more and more areas of our everyday lives. How can you tell this is happening?**

Above all, we can tell from our smartphones. Every day, for example, you're being confronted with automatic word completion when typing in a text message. This "predictive typing" is based on a linguistic model which makes a statistical evaluation of what word is most likely to come next. Another example is the biometric unlocking of a phone. Technically speaking, face recognition and image comparison are at the boundary between signal analysis and AI – although I'm pretty sure that, meanwhile, AI is also part of those algorithms. Today, many image processing filters are based on AI as well. Moreover, AI is omnipresent on many online platforms. Facebook posts, for example, are automatically scanned for undesirable content like pornography and hate speech. As a user, you don't notice AI being deployed, but anybody using Facebook is confronted with those filters every day. Recommender systems have also intruded on our lives. In the beginning, these systems were rather rules-based – it was tried to find rules in certain domains and then build a decision tree. Today, machine learning systems are used to try to find them automatically.

There are cases in which you can at least understand in part why a specific decision was made. In rule-based systems, this is easy to comprehend because the rules are made by human minds. In machine-learning systems, it sometimes is artificial attributes that are no longer so easy to understand. In larger neural networks, however, with millions of nodes and edges, this is no longer comprehensible to the human mind at all. So you have to try in retrospect to understand the behavior of those systems. Such "Explainable AI (XAI)" procedures are currently being worked on.

**Do such self-explaining systems that allow taking a look inside the "black box" of AI already exist in practice?**

My knowledge is, of course, limited, but I am not aware of any product that provides for explainability of a decision as part of the AI process. There is, however, certainly a need for it. For example in industrial applications: deployment of AI technologies also depends on being able to trust the systems and understand decisions they make. This is also a liability issue. I think it is something that, slowly but surely, end users will be calling for as well. At the moment, it is rather that people are not really aware of AI working in the background. Therefore, there is no increased demand for explainability observable yet.

Menschen gar nicht wissen, dass ein KI-System dahintersteht. Daher gibt es derzeit auch noch kein gesteigertes Verlangen nach Erklärbarkeit.

Bedarf nach Erklärbarkeit gibt es zum Beispiel auch im Bereich Human Resources: Großunternehmen, die Tausende von Lebensläufen von Bewerber:innen bekommen, setzen KI-Systeme ein, um eine erste Filterung der Kandidat:innen für einen ausgeschriebenen Job zu bekommen. Da stellt sich aber sofort die Frage: Woher weiß man, dass keine Verzerrung hinsichtlich Geschlechtern oder Alter dahinter steht? Ich als Bewerber:in würde wissen wollen, ob mein Lebenslauf künstlich beurteilt wurde und auf welcher Basis ich kein Angebot für den Job bekommen habe.

### Was sind Ihrer Meinung nach zurzeit die ganz heißen Themen in der KI-Forschung und Entwicklung?

Unter den großen Hypes befinden sich autonome Fahrzeuge. In dieses Gebiet wird sehr viel investiert. Es ist auch vielversprechend – wir alle träumen davon, dass wir auf langen Autobahnfahrten miteinander reden oder vielleicht auch schlafen können. Aber bei der Sicherheit solcher Systeme haben wir grundsätzliche Schwierigkeiten. Ich bin skeptisch, ob wir jemals so weit sein werden, dass wir 100 Prozent unserer Fahrtätigkeit an die KI übergeben können.

Ein anderer wichtiger Bereich ist der Gesundheitssektor – zum Beispiel in der Diagnostik von Bildern oder auf Basis von Symptomen. Hier bietet KI eine Erweiterung von regelbasierten Expertensystemen, die es seit Jahrzehnten gibt. Aber auch hier gibt es grundsätzliche Fragen des Vertrauens, der Erklärbarkeit und hinsichtlich ethischer Probleme zu klären: Kann ein Arzt, eine Ärztin wirklich auf ein solches System zurückgreifen, wenn er oder sie selbst nicht versteht, warum eine bestimmte Diagnose erstellt wurde? Kann man die Gesundheit von Menschen an eine Maschine übergeben? Am AIT reden wir daher fast exklusiv von KI als Unterstützung von Menschen, und nicht als Ersatz.

Auch im Falle von sehr großen Datenmengen kann man KI nutzen, um repetitive Vorarbeiten zu erledigen oder um Arbeit zu priorisieren. Ein großes Anwendungsgebiet ist dabei die Dokumentenverarbeitung, zum Beispiel um aus strukturierten Briefen auf Basis von Layout-Analysen Informationen zu gewinnen – etwa wenn man weiß, wo die Adresse, Sozialversicherungsnummer oder Telefonnummer auf einem Brief steht. Mithilfe von Sprachmodellen ist es auch möglich, die Anschrift der Absender:innen aus unstrukturierten Daten herauszuholen.

Wichtige Anwendungen gibt es auch in der Cyber-Sicherheit, wo man versucht, aus den Mustern vieler verschiedener Quellen, etwa Netzwerkpakete oder Logfiles, Sicherheitsbedrohungen zu erkennen („Intrusion Detection“).

In the field of Human Resources, for example, there is a need for explainability as well: large companies that receive thousands of résumés from job applicants use AI systems to do a first filtering of candidates for an advertised job. Which immediately raises the question: How do you know that there is no gender or age bias behind this? As an applicant, I would want to know whether my résumé was computer-assessed and on what grounds I was not offered the job.

### What do you think are currently the real hot topics in AI research and development?

One of the big hypes is autonomous vehicles. A lot is being invested in this area. And it actually is promising – we all dream of being able to chat with one another or perhaps take a nap on long highway drives. But when it comes to the safety of such systems we meet with fundamental difficulties. I'm skeptical that we'll ever get to the point where we can hand over one-hundred percent of our driving to AI.

Another important area is the healthcare sector – for example, in image or symptoms-based diagnostics. Here, AI offers an expansion of rule-based expert systems that have been in place for decades. But here, too, there are fundamental questions to be addressed, of trust, explainability, and with respect to ethics issues: Can doctors really rely on such a system if they themselves don't understand why a particular diagnosis was made? Can human health be handed over to a machine? At AIT, we therefore speak almost exclusively of AI as an assistance to, not a replacement for, people.

Also when dealing with very large bulks of data, AI can be used to perform repetitive preparatory work, or to prioritize work. One major area of application is document processing, for example, to extract information from structured letters on the basis of layout analyses – for example, if you know where the address, social security number or telephone number is located on a letter. With the help of language models, it is also possible to extract the sender's address from unstructured data.

There are other important applications in cyber security, where attempts are made to detect security threats from patterns from many different sources, such as network packets or log files (“intrusion detection”).

### **Es gibt die Ansicht, dass man KI-Methoden dann einsetzt, wenn man mit herkömmlichen Methoden nicht mehr weiterkommt. Wie sehen Sie das?**

Die Frage ist, warum man nicht weiterkommt. Ein typischer Fall ist, dass man glaubt, dass es einen Zusammenhang zwischen Daten und Ereignissen gibt – man weiß aber nicht, wie dieser Zusammenhang aussieht. Oder dass es sehr viele Daten gibt, man aber nicht weiß, welche wirklich aussagekräftig sind. Für solche Fragen gibt es seit Jahrhunderten statistische Methoden. Mit Machine Learning hat man nun ein zusätzliches Werkzeug in der gesamten Data-Science-Kiste, auf das man zurückgreifen kann. So kann man beispielsweise mithilfe eines neuronalen Netzes eine Principal Component Analysis (PCA) durchführen und damit Zusammenhänge aufdecken. Aber wir müssen vorsichtig sein: Diese Modelle haben sehr viele Parameter. Es ist auch eine Kunst, dass man ein Modell zum Beispiel nicht übertrainiert. Das würde heißen, dass das Modell die Daten auswendig lernt, aber nicht generalisiert. Dann gäbe es bei den Trainingsdaten immer eine sehr hohe Erkennungsrate, aber man kann das Modell nicht auf neue Daten anwenden. Je mehr Parameter man hat, umso größer ist dieses Problem des Overtrainings.

KI ist jedenfalls kein Wundermittel: Wenn man schlechte Daten hineingibt, kommt ein schlechtes Modell heraus. Auch wenn bestimmte Daten fehlen, wenn es mit etwas in Berührung kommt, das es vorher noch nie gesehen hat.

### **Man hört immer wieder, dass in der Data Science das Aufbereiten der Datensätze der wirklich aufwendige Schritt ist. Stimmt das?**

Absolut. Typischerweise liegen 80 Prozent der Arbeit in der Datenvorbereitung – und nicht im Trainieren des Modells. Wenn man große Datensätze hat, ohne eine Hypothese dahinter zu haben, wenn man also nur mit den Daten herumspielt und nach Korrelationen sucht, wird man immer etwas finden. Die Frage ist, ob das eine echte Korrelation oder eine Scheinkorrelation ist. Google hat zum Beispiel gemeint, dass sie Influenza aufgrund von Suchanfragen vorhersagen können. In einem Jahr hat das funktioniert, im nächsten Jahr war dann der Effekt weg, weil es eine falsche Korrelation war. Diese Frage nach der Korrelation ist natürlich viel älter. Die wissenschaftliche Methode existiert genau deshalb: um Fehlinterpretationen zu vermeiden. Es geht um Reproduzierbarkeit und um das Testen von Hypothesen, um hinter Scheinkorrelationen und andere Denkfehler zu kommen.

### **Woran scheitert die Anwendung von KI in der Praxis am häufigsten?**

Unserer Erfahrung nach ist der häufigste Grund für das Scheitern eines Projekts eine falsche Erwartung. Weil so viel von KI gesprochen wird, erwarten viele Menschen Wunder und verstehen dabei nicht, dass eine KI-Vorhersage immer mit einer Wahrscheinlichkeit

### **There is the view that AI tools are put to use only when conventional methods have reached an impasse. How do you see that?**

The question is why there is an impasse. One typical case is that you believe there is a relationship between data and events – but don't know what that relationship is. Or that there is a lot of data, but you don't know which is really meaningful. To answer such questions, there have been statistical methods for centuries. With machine learning, you now have an additional tool in the whole data-science box to resort to. For example, you can use a neural network to perform a Principal Component Analysis (PCA) to uncover correlations. But we have to be careful there: these models have a lot of parameters. It is also kind of an art not to overtrain a model, for example. That would mean that the model memorizes but fails to generalize the data. Then you'd always get very high recognition rates on the training data, but couldn't apply the model to new data. The more parameters you have, the bigger the problem of overtraining.

In any case, AI is not a magic bullet: if you put in bad data, the outcome will be a bad model. Also, if certain data is missing, the model will fail in the field if encountering something it has never seen before.

### **You keep hearing that the preparation of datasets really is the laborious step in data science. Is that true?**

Absolutely. Typically, eighty percent of the work is in data preparation – and not in the training of the model. If you have large sets of data without having an underlying hypothesis, if that is, you are just playing around with the data, looking for correlations, you are always going to find something. The question is whether this is a real or a spurious correlation. Google, for example, thought that they could predict influenza based on search queries. In one year it worked, and in the next year the effect was gone because it was a false correlation. This question of correlation is of course much older. This is precisely what the scientific method is here for, to avoid misinterpretation. It is about reproducibility and about the testing of hypotheses, to get to the bottom of spurious correlations and other fallacies.

### **What are the most common reasons for AI to fail in practice?**

In our experience, the most common reason for a project to fail is false expectations. As there is so much talk about AI, many people are expecting miracles and don't understand that any AI prediction always implies a probability. This is, for example, the fundamental problem about autonomous driving: If human safety is at stake we expect one-hundred percent

verbunden ist. Das ist zum Beispiel das Grundproblem beim autonomen Fahren: Wenn es um menschliche Sicherheit geht, erwarten wir eine 100-prozentige Zuverlässigkeit – obwohl wir Menschen das selber nicht schaffen.

Das nächstgrößere Problem sind schlechte Trainingsdaten. Man zeigt der Maschine hunderttausende Beispiele, sie lernt daraus und kann eine Kategorisierung neuer Daten vornehmen. In vielen Fällen fehlen Daten aber komplett. Oder es gibt zwar Daten, aber keinen Zusammenhang zwischen Daten und Zielwerten. Dann kann man auch kein Modell bauen. Damit sind wir oft konfrontiert, etwa bei der vorhersagenden Wartung („Predictive Maintenance“): Man hat zwar durch Sensoren sehr viele Maschinendaten. Um ein Modell bauen zu können, müssten die Maschinenausfälle in der Vergangenheit sehr gut dokumentiert sein und mit den Messwerten verbunden sein. Wenn es in den historischen Daten aber keinen zeitlichen Zusammenhang zwischen diesen beiden Faktoren gibt, kann man diese Zusammenhänge auch nicht trainieren.

Auch die Datenqualität spielt eine große Rolle. Im Detail gibt es oft zu wenig historische Daten. Oder die Daten haben einen „bias“, sind also systematisch verzerrt. Das ist beispielsweise beim „predictive policing“ wichtig, wo die Ressourcen der Polizei auf Basis vergangener Kriminalfälle eingesetzt werden. Die Vergangenheit ist aber voll von menschlichen Vorurteilen – und wenn man nicht vorsichtig ist, erzeugen diese in den Modellen bestimmte Muster.

Weiters ist es wichtig, die richtige KI-Methode, die richtige Architektur auszuwählen. Es wäre auf jeden Fall ein Fehler zu sagen: Man löst jedes Problem mit neuronalen Netzwerken. Diese sind für manche Probleme besser geeignet als für andere. Eine sehr wichtige Frage ist zum Beispiel, ob es in den Daten eine zeitliche Komponente gibt. Bei Bildern ist das typischerweise nicht der Fall, sehr wohl aber in der Sprachanalyse, weil die Reihenfolge der Wörter sehr wichtig ist. Für diese verschiedenen Bereiche benötigt man auch unterschiedliche Methoden.

#### **Welche Voraussetzungen müssen gegeben sein, damit der Einsatz von KI-Methoden sinnvoll ist?**

KI ist dann sehr sinnvoll, wenn man qualifizierte Grunddaten hat. Ein sehr gutes Beispiel ist der Fake Shop Detector, den wir gemeinsam mit den Partnern X-Net Services und öIAT entwickelt haben. Dieser Detektor schützt Menschen sehr zuverlässig vor betrügerischen E-Commerce-Seiten. Das funktioniert deshalb so gut, weil das öIAT schon lange ein E-Commerce-Gütesiegel vergibt und eine große Datenbasis von Seiten hat, die von menschlichen Expert:innen als gut oder schlecht bewertet wurden. Wenn man so gute Daten hat, kann man ein sehr gutes KI-Modell trainieren.

reliability – even though it's something that we humans ourselves cannot achieve.

The next biggest problem is poor training data: you keep showing the machine hundreds of thousands of examples, it learns from them and is able to categorize new data. In many cases, though, data is completely missing. Or there is data, but no correlation between them and target values. That means you can't build a model. It's something we are often confronted with, for example in predictive maintenance: You get a lot of machine data from sensors. To build a model, the machine malfunctions of the past would have to be very well documented and related to the measured values. However, if there is no temporal nexus between these two factors in the historical data, these correlations cannot be trained.

Data quality also plays a major role. Often, there is not enough historical data in detail. Or the data have a bias, meaning that they are systematically distorted. This is crucial “predictive policing”, for example, where police resources are deployed based on past criminal cases. However, the past is full of human bias – and if you're not careful, these create certain profiling patterns in the models.

Also crucial is choosing the right AI method, the right architecture. It would be a mistake in any case to say: you can solve each and every problem with neural networks. These are better suited for some problems than for others. One very important question, for example, is whether there is a time component in the data. With images, this is typically not the case, but it is very much the case in language analysis, because the sequence of words is very important. So, you need different methods for these different areas.

#### **What are the requirements for deployment of AI methods to make sense?**

AI does make sense when you have qualified basic data. A very good example is the Fake Shop Detector we developed in collaboration with X-Net Services and the Austrian Institute of Applied Telecommunication (öIAT) as partners: the detector very reliably protects people against fraudulent e-commerce sites. It works so well because öIAT has long been awarding an e-commerce quality seal and operates a large database of websites that were rated as good or bad by human experts. If you have data that is that good, you can easily build a very good AI model.

Nichtsdestotrotz können wir niemandem sagen, dass eine bestimmte E-Commerce-Seite zu 100 Prozent sicher oder betrügerisch ist. Aber wir können Empfehlungen geben – etwa sich eine Seite genauer anzusehen. Das öIAT verwendet das System dazu, um ihre manuellen Beurteilungen zu priorisieren. Sie schauen sich zuerst jene Seiten an, die am schlechtesten wirken – dadurch kommen sie viel schneller auf eine „white list“ bzw. eine „black list“.

#### Wo hat KI keinen Sinn?

Ich bin sehr skeptisch bei sicherheitskritischen Anwendungen: Hier kann man den Menschen nicht komplett ausschalten. Es ist nicht schlecht, dass Assistenzsysteme in Autos Lenker:innen auf ein Hindernis auf der Straße hinweisen. Aber ich halte es für eine schlechte Idee, wenn ein System sagt: Ich übernehme, und du kannst inzwischen gerne mit deinen Mitfahrer:innen plaudern. Sicherheit passt nicht mit Wahrscheinlichkeitsaussagen zusammen. Meiner Meinung nach werden wir immer mit Unsicherheit und Wahrscheinlichkeiten leben müssen. Ich kann mir vorstellen, dass autonome Fahrzeuge in einigen Jahrzehnten sicherer sind, als wenn ein Mensch selbst am Steuer sitzt. Aber ich bezweifle, dass es je sicherer werden kann, als das System und der Mensch zusammen. Ich glaube, dieses Hybrid wird immer stärker sein: dass Menschen von einer KI unterstützt werden – und nicht ersetzt werden. ✕

Nevertheless, we can't tell anybody that a certain e-commerce site is a hundred percent safe or fraudulent. But we can make recommendations – like taking a closer look at a specific site. öIAT uses the system to prioritize their manual ratings. They first look at sites that seem most dubious – this makes drawing up a “white list” or “black list” much quicker.

#### Where does AI not make sense?

I am very skeptical about safety-critical applications: you cannot completely eliminate the human element. It's not a bad idea to have assistance systems in cars that alert drivers to obstacles on the road. But I think it is a bad idea for a system to say: I'm taking over, and meanwhile you're welcome to chat with your passengers. Safety does not go together with probability statements. In my opinion, we will always have to live with uncertainty and probabilities. I can imagine that, in a few decades from now, autonomous vehicles will be safer than with a human behind the wheel. But I doubt it could ever be safer than that system and the human combined. I think this hybrid will always be stronger: humans supported by AI – and not replaced. ✕



Diskussionen  
über KI

Discussing AI

## Aktuelle Debatten über KI

**In den weltweiten Debatten um Künstliche Intelligenz wurde die Welle von dystopischen Ängsten, dass Maschinen die Herrschaft übernehmen könnten, nun abgelöst von ethischen Debatten und Fragen der Regulierung der neuen Technologien. Philosophische Probleme hinsichtlich des Selbstbildes des Menschen sind und bleiben freilich Dauerbrenner der Debatten.**

Eine alte Kontroverse schlug kürzlich wieder einmal hohe mediale Wellen: Können Maschinen ein Bewusstsein haben? Der Google-Mitarbeiter Blake Lemoine, der das unternehmenseigene Sprachsystem LAMDA (Language Model for Dialogue Applications) testete, schloss aus einigen Unterhaltungen mit der Maschine (etwa über Angst oder den Tod), dass das System ein eigenes Bewusstsein entwickelt habe. „Hätte ich keine Ahnung, dass es sich um ein kürzlich entwickeltes Computerprogramm handelt, würde ich denken, es sei ein sieben oder acht Jahre altes Kind, das zufällig Ahnung von Physik hat“, wird er in Medien zitiert (hier nach: „Streit bei Google um eine KI, die ein eigenes Bewusstsein erlangt haben soll“, [www.derstandard.at](http://www.derstandard.at), 13. Juni 2022).

Gemeinsam mit einem Kollegen sammelte er seine Beobachtungen und präsentierte diese der Google-Führung. Nach einer Prüfung der Argumente und Beispiele beschied diese, dass es keine Beweise gebe, dass LAMDA Bewusstsein erlangt habe, sehr wohl aber viele Hinweise, die dagegen sprechen. Lemoine, der vor seinem Google-Job Priester war, gab aber nicht auf, seine Bedenken zu äußern und zu fordern, dass Google solche Fragen mit der Öffentlichkeit diskutieren müsse. Er engagierte einen Anwalt, der LAMDA vertreten sollte, und sprach mit einem Mitglied des Justizausschusses im US-Repräsentantenhaus über die seiner Meinung nach unethischen Aktivitäten von Google. Der Konzern warf ihm daraufhin vor, Geheimhaltungsvereinbarungen gebrochen zu haben, und schickte Lemoine auf bezahlten Urlaub.

Eine letztgültige Antwort auf die Frage hat niemand. Unter der Annahme, dass Bewusstsein bei Lebewesen eine emergente Eigenschaft eines hinreichend komplexen Nervensystems ist, ist ein vergleichbarer Prozess auch bei künstlichen Systemen nicht völlig undenkbar. Allerdings

## Current debates concerning AI

**In the worldwide debates about Artificial Intelligence, the wave of dystopian fears that machines could take over the world has now been replaced by ethical debates and questions about how to regulate the new technologies. Philosophical problems regarding the self-perception of human beings are and will remain a constant topic of debate, however.**

An old controversy has recently made waves in the media once again: Can machines have a consciousness? Google employee Blake Lemoine, who tested the company's own language system LAMDA (Language Model for Dialogue Applications), concluded from some conversations with the machine (e.g., about fear or death) that the system had developed its own consciousness. "If I didn't know it was a recently developed computer program, I would think it was a seven or eight year old child who happened to know physics," he is quoted in the media (here cited from: "Streit bei Google um eine KI, die ein eigenes Bewusstsein erlangt haben soll", [Dispute at Google over an AI that is said to have acquired its own consciousness], [www.derstandard.at](http://www.derstandard.at), 13 June 2022).

Along with a colleague, he collected his observations and presented them to Google management. After examining the arguments and examples, they decided that there was no evidence that LAMDA had attained consciousness, but that there was plenty of evidence against it. Lemoine, who was a priest before his Google job, did not give up voicing his concerns, however, and demanded that Google discuss such issues with the public. He hired a lawyer to represent LAMDA and spoke to a member of the Judiciary Committee in the U.S. House of Representatives about what he saw as Google's unethical activities. The company then accused him of breaking confidentiality agreements and sent Lemoine home on paid leave.

It is an issue that as yet has no definitive answer. Assuming that consciousness in living beings is an emergent property of a sufficiently complex nervous system, a comparable process is not completely

trifft dies sicher nicht auf heutige künstliche neuronale Netze zu – wie sie in diesem Jahrbuch in vielerlei Beispielen beschrieben werden. Diese Systeme liefern, so die vorherrschende Meinung, in eng umgrenzten Domänen eine statistische Auswertung jener Daten, mit denen man sie füttert. Nicht mehr, aber auch nicht weniger.

Die immer wiederkehrende Debatte nach einem Maschinenbewusstsein zeigt aber, wie brisant diese Frage ist – und zwar insbesondere für unser Selbstbild als Menschen. Im Laufe der Technikgeschichte musste der Mensch schon so manches, von dem wir dachten, dass es uns gegenüber allen anderen Lebewesen auszeichne, über Bord werfen. Rechenmaschinen arbeiten schneller als wir, Roboter arbeiten mit einer viel höheren Präzision. Und moderne KI-Systeme zeigen zweifellos „intelligentes“ Verhalten – zumindest wenn man den Turing-Test als Kriterium heranzieht. Wenn man bei einem Gegenüber, mit dem man interagiert, nicht mehr sagen kann, ob es sich um eine Maschine oder einen Menschen handelt, so muss man dem Gegenüber Intelligenz zusprechen, meinte der Computerpionier Alan Turing im Jahr 1950. Moderne Sprachsysteme wie GPT-3 oder LAMDA erfüllen dieses Kriterium offenbar. Eine ganz andere Frage ist aber die des Bewusstseins: Dieses ist ausschließlich durch Introspektion zugänglich – so haben Biolog:innen auch keinerlei Nachweise, dass z.B. Säugetiere über ein Bewusstsein, wie wir es kennen, verfügen. Obwohl jeder Haustierhalter davon überzeugt ist.

### **Ein:e Begleiter:in für den Menschen**

Wie dem auch sei: Für den praktischen Umgang mit KI-Systemen sind andere Fragen viel wichtiger. Zum Beispiel ethische Fragen, die Konsequenzen der Technologie für Wirtschaft und Gesellschaft, oder unser künftiger Umgang mit KI-Systemen. Einen vielbeachteten Streifzug durch diese Fragen unternahm kürzlich das international höchst renommierte Autorentrio Henry Kissinger (früherer US-Außenminister), Eric Schmidt (langjähriger CEO von Google) und Daniel Huttenlocher (Professor am MIT) in dem Buch „The Age of AI: And Our Human Future“ (Little, Brown and Company, New York, Boston, London). Der Wert von KI-Systemen liegt ihrer Meinung nach darin, dass sie uns auf Zusammenhänge aufmerksam machen, die wir ohne sie nicht entdecken würden. Als Beispiel führen sie die Entwicklung eines neuen Medikaments gegen antibiotikaresistente Bakterien an. Eine KI, die am MIT mit zahlreichen Strukturen von Biomolekülen sowie deren biologischer Wirkung gefüttert wurde und dabei Zusammenhänge zwischen diesen beiden Größen gelernt hat, erkannte eine chemische Struktur als potenziell wirksam, von der menschliche Expert:innen niemals erwartet hätten, dass sie wirksam sein könnte. In Versuchen im Labor und in der Klinik wurde ihr Potenzial als völlig neuartiges Medikament bestätigt. Die Substanz wurde „Halicin“ getauft – nach HAL, dem fiktiven Computer aus dem Film „2001: Odyssee im Weltraum“.

Anhand dieses Beispiels argumentieren die Autoren, dass KI-Systeme eine wertvolle Ergänzung zu den menschlichen Fähigkeiten, ein:e Begleiter:in (companion) des Menschen sein können, die andere Dinge sehen und

inconceivable in artificial systems. However, this certainly does not apply to today's artificial neural networks – as described in many examples in this yearbook. In the prevailing opinion, these systems provide a statistical evaluation of the data they are fed in narrowly defined domains. Nothing more, but nothing less either.

The ever-recurring debate about machine consciousness, however, shows how explosive this question is – especially for our self-perception as humans. In the course of the history of technology, humans have had to throw overboard a number of things that we thought distinguished us from all other living beings. Computing machines work faster than we do, robots work with much greater precision. And modern AI systems undoubtedly show “intelligent” behavior – at least if you use the Turing test as a criterion. According to computer pioneer Alan Turing's opinion voiced in 1950, if you can no longer tell whether the person you are interacting with is a machine or a human being, you have to attribute intelligence to the other party. Modern language systems such as GPT-3 or LAMDA evidently meet this criterion. A completely different question, however, is that of consciousness: this is only accessible through introspection – biologists have no evidence that mammals, for example, have consciousness as we know it. Even though every pet owner is convinced of it.

### **A companion for human beings**

Be that as it may: other questions are much more important for the practical use of AI systems. Such as ethical questions, the consequences of the technology for the economy and society, or our future interaction with AI systems. The internationally renowned author trio Henry Kissinger (former US Secretary of State), Eric Schmidt (long-time CEO of Google) and Daniel Huttenlocher (professor at MIT) recently undertook a highly acclaimed foray into these questions in their book “The Age of AI: And Our Human Future” (Little, Brown and Company, New York, Boston, London). The value of AI systems, in their opinion, lies in the fact that they make us aware of connections that we would not discover without them. One example they cite is the development of a new drug against antibiotic-resistant bacteria. An AI at MIT that was fed numerous structures of biomolecules as well as their biological effect, and that thus learned correlations between these two variables, recognized a chemical structure as potentially effective that human experts would never have expected to be effective. Laboratory and clinical trials confirmed its potential as a completely new type of drug. The substance was christened “Halicin” – after HAL, the fictional computer from the famous movie “2001: A Space Odyssey”.

The authors use this example to argue that AI systems can be a valuable complement to human abilities, companions to humans that can see different things and recognize different connections. How well this can succeed or which catastrophic developments could be triggered by it (keywords: cyberwar, discrimination, disinformation) is

andere Zusammenhänge erkennen. Wie gut dies gelingen kann bzw. welche katastrophalen Entwicklungen damit auch angestoßen werden könnten (Stichworte: Cyberwar, Diskriminierung, Desinformation) ist eine Frage der Rahmenbedingungen. Die drei Autoren rufen Wissenschaft, Politik und Gesellschaft dringend zu einem Dialog auf, um eine Balance zwischen den verschiedenen Aspekten von KI zwischen Freiheit und Regulierung zu finden. Sie fordern die Entwicklung einer KI-Ethik (mit dem Grundsatz, dass der Mensch stets die Oberhoheit über Technologie behalten müsse). „Jetzt gilt es unsere Partnerschaft mit Künstlicher Intelligenz zu definieren, ebenso wie die daraus resultierende Realität.“ (Now is the time to define both our partnership with Artificial Intelligence and the reality that will result.)

### **EU versucht sich an der weltweit ersten Regelung für KI**

Einen ernsthaften Versuch, das Thema KI in einen passenden Rechtsrahmen zu bringen, um Chancen und Risiken angemessen gegeneinander abzuwägen, unternimmt zur Zeit die Europäische Kommission. Die Verordnung „Festlegung harmonisierter Vorschriften für Künstliche Intelligenz“ soll „sowohl für öffentliche als auch für private Akteure innerhalb und außerhalb der EU gelten, sofern das KI-System in der Union in Verkehr gebracht wird oder Menschen in der EU von seiner Verwendung betroffen sind“. Kern des Vorschlags, der derzeit in Fachkreisen heftig debattiert wird, ist ein risikobasierter Ansatz mit vier Risikostufen:

**Unannehmbares Risiko:** Eine sehr geringe Zahl besonders schädlicher KI-Anwendungen, die gegen die Werte der EU verstoßen, weil sie Grundrechte verletzen, wird verboten. Das betrifft z. B. die Bewertung des sozialen Verhaltens durch Behörden (Social Scoring), die Ausnutzung der Schutzbedürftigkeit von Kindern, den Einsatz von Techniken zur unterschweligen Beeinflussung und – mit eng gefassten Ausnahmen – biometrische Echtzeit-Fernidentifizierungssysteme, die zu Strafverfolgungszwecken im öffentlich zugänglichen Raum eingesetzt werden.

**Hohes Risiko:** Eine begrenzte Zahl von KI-Systemen, die genau definiert werden (wobei die Liste regelmäßig überarbeitet werden soll, um sie an die Entwicklung der KI-Anwendungsfälle anzupassen) und die sich nachteilig auf die Sicherheit der Menschen oder ihre Grundrechte auswirken, gilt als mit einem hohen Risiko behaftet. Dazu zählen unter anderem die Bereiche biometrische Identifizierung und Kategorisierung, kritische Infrastrukturen, Bildung und Erziehung, Personaleinstellung und Beschäftigung, Erbringung wichtiger öffentlicher und privater Dienstleistungen, aber auch Strafverfolgung, Asyl und Migration sowie Justiz. Für diese sollen verbindliche Anforderungen gelten: Die Anbieter:innen müssen sich vor Inverkehrbringen solcher KI-Systeme einer Konformitätsbewertung unterziehen, um nachzuweisen, dass ihr System den verbindlichen Anforderungen an vertrauenswürdige KI entspricht (etwa in Bezug auf Datenqualität, Dokumentation und Rückverfolgbarkeit, Transparenz, menschliche Aufsicht, Genauigkeit und Robustheit). Sollte später das System selbst oder sein Zweck wesentlich geändert werden, muss die Bewertung wiederholt werden. Überdies müssen Anbieter:innen von KI-Systemen mit hohem Risiko

a question of the framework conditions. The three authors urgently call on science, politics, and society to engage in a dialog to find a balance between the different aspects of AI between freedom and regulation. They call for the development of an AI ethic (with the principle that humans must always retain supremacy over technology). “Now is the time to define both our partnership with Artificial Intelligence and the reality that will result.”

### **EU attempts the world's first AI regulation**

The European Commission is currently making a serious attempt to bring the topic of AI into a suitable legal framework in order to appropriately weigh up the opportunities and risks. The regulation “Establishing Harmonised Rules on Artificial Intelligence” is to apply “to both public and private actors inside and outside the EU, provided that the AI system is placed on the market in the Union or that people in the EU are affected by its use”. The core of the proposal, which is currently being hotly debated in expert circles, is a risk-based approach with four levels of risk:

**Unacceptable risk:** A very small number of particularly harmful AI applications that violate EU values because they violate fundamental rights will be prohibited. This concerns, for example, the assessment of social behavior by public authorities (Social Scoring), the exploitation of children's vulnerability, the use of subliminal influencing techniques and – with strictly defined exceptions – real-time biometric remote identification systems used for law enforcement purposes in publicly accessible spaces.

**High risk:** A limited number of AI systems that are well defined (with the list to be regularly revised to adapt to the evolution of AI use cases) and that have an adverse impact on people's safety or security or their fundamental rights, are considered to present a high risk. These include, among others, biometric identification and categorization, critical infrastructure, education and training, recruitment and employment, provision of essential public and private services, but also law enforcement, asylum processes and migration, and justice. Binding requirements shall apply to these: Providers must undergo a conformity assessment before placing such AI systems on the market to demonstrate that their systems meet the mandatory requirements for trustworthy AI (e.g., in terms of data quality, documentation and traceability, transparency, human oversight, accuracy, and robustness). Should the system itself or its purpose be significantly changed at a later date, the assessment must be repeated. In addition, providers of high-risk AI systems must implement quality and risk management systems to ensure compliance with the new requirements and minimize risks to users and data subjects – even after a product has been placed on the market.

Qualitäts- und Risikomanagementsysteme einführen, um die Einhaltung der neuen Anforderungen sicherzustellen und die Risiken für Nutzer:innen und betroffene Personen zu minimieren – auch nachdem ein Produkt bereits in Verkehr gebracht wurde.

**Geringes Risiko:** Bei bestimmten KI-Systemen werden besondere Transparenzverpflichtungen auferlegt, z. B. wenn eine klare Manipulationsgefahr besteht (etwa durch den Einsatz von Chatbots). Den Nutzer:innen sollte bewusst sein, dass sie es mit einer Maschine zu tun haben.

**Minimales Risiko:** Alle anderen KI-Systeme können unter Einhaltung des allgemein geltenden Rechts entwickelt und verwendet werden, das heißt ohne Beachtung zusätzlicher rechtlicher Verpflichtungen. Die große Mehrheit der KI-Systeme, die derzeit in der EU verwendet werden, fällt in diese Kategorie. Anbieter:innen solcher Systeme können freiwillig die Anforderungen an vertrauenswürdige KI anwenden und freiwillige Verhaltenskodizes einhalten.

Zwischen Behörden und Datenschützer:innen läuft derzeit eine heftige Debatte um die genauen Risikoeinstufungen von KI-Systemen, um Ausnahmeregelungen und Sanktionen bei Nichteinhaltung. Unter der französischen EU-Ratspräsidentschaft im ersten Halbjahr 2022 wurde der Verordnungstext kräftig überarbeitet, im Herbst soll das Europäische Parlament darüber beraten.

In weiten Teilen der Welt, inklusive der aktuellen Regierung der USA, blickt man mit großem Interesse auf den europäischen „AI Act“, denn dieser ist immerhin weltweit der erste Anlauf zu einer umfassenden Regulierung dieses Bereichs. Expert:innen halten es für durchaus möglich, dass Europa auch in diesem Bereich einen weltweiten Standard setzt – so wie es auch schon bei der europäischen Datenschutzgrundverordnung (DSGVO) war, die weltweit, von Kalifornien bis Indien, kopiert wurde.

### **Diskriminierung und Datenkolonialismus**

Vor allem in den USA läuft eine intensive Debatte über eine neue Form von Diskriminierung durch den Einsatz von KI-Methoden. Das betrifft insbesondere die Bereiche Geschlecht und Hautfarbe. Nach zahlreichen Berichten über Fehlleistungen von KI-Systemen (etwa in sexistischen Jobeinstellungspraktiken oder rassistischen Polizeiaktionen) wurde klar, dass die Trainingsdaten für solche KI-Systeme systematisch verzerrt sind. In diesen Daten sind weiße Männer deutlich stärker repräsentiert als andere gesellschaftliche Gruppen – und dieser „bias“ wird von Maschinenlernsystemen auch gelernt. Wie die Aktivistin Joy Buolamwini mit der von ihr gegründeten „Algorithmic Justice League“ (AJL) herausgefunden hat, ist diese Verzerrung auch kein Wunder: Demnach haben weniger als zwei Prozent der Mitarbeiter:innen in technischen Positionen bei Facebook und Google eine dunkle Hautfarbe. Bei acht großen Technologieunternehmen, die von Bloomberg ausgewertet wurden, sind jeweils nur etwa ein Fünftel der technischen Belegschaft Frauen.

**Low risk:** Specific transparency obligations are imposed on certain AI systems, for example where there is a clear risk of manipulation (such as through the use of chatbots). Users should be aware that they are dealing with a machine.

**Minimal risk:** All other AI systems can be developed and used in compliance with generally applicable law, i.e. without observing additional legal obligations. The vast majority of AI systems currently in use in the EU fall into this category. Providers of such systems can voluntarily apply the requirements for trustworthy AI and comply with voluntary codes of conduct.

There is currently a heated debate between authorities and data protectors about the exact risk ratings of AI systems, exemptions, and sanctions for non-compliance. Under the French EU Presidency in the first half of 2022, the text of the regulation was heavily revised, and the European Parliament is expected to discuss it in autumn.

In many parts of the world, including the current U.S. government, the European “AI Act” is being looked at with great interest, since it is the world’s first attempt at comprehensive regulation in this area. Experts believe it is quite possible that Europe will also set a global standard in this area – as was the case with the European General Data Protection Regulation (GDPR), which has been copied worldwide, from California to India.

### **Discrimination and data colonialism**

Particularly in the U.S., there is an intense debate about a new form of discrimination through the use of AI methods. This particularly concerns the areas of gender and skin color. After numerous reports about failures of AI systems (for example in sexist job hiring practices or racist police actions), it became clear that the training data for such AI systems is systematically biased. In this data, white men are represented in significantly greater proportion than other social groups – and this “bias” is also being learned by machine learning systems. As activist Joy Buolamwini and the Algorithmic Justice League (AJL), which she founded, have discovered, this bias is not surprising: according to the AJL, less than two percent of employees in technical positions at Facebook and Google have dark skin. Only about one fifth of the technical staff at each of the eight large technology companies analyzed by Bloomberg are women.



„Die Unterrepräsentation von Frauen und farbigen Menschen in der Technologie und die Untererfassung dieser Gruppen in den Daten, die die KI formen, hat zur Entwicklung von Technologien geführt, die für einen kleinen Teil der Welt optimiert sind“, schreibt Buolamwini im Time Magazine (7. 2. 2019, <https://time.com/5520558/artificial-intelligence-racial-gender-bias/>).

Großen Raum in Debatten über KI nehmen zur Zeit auch Vorwürfe eines „neuen Kolonialismus“ ein. Demnach meinen Aktivist:innen, dass die Entwicklung der KI Muster aus dem historischen Kolonialismus wiederhole – nämlich die Probleme des „globalen Südens“, die Macht westlicher Industrieländer und die Dominanz von Weltkonzernen allein zur Maximierung von Profiten auszunutzen. Das reicht von der Ausbeutung von Arbeitskräften bis hin zur massenhaften Extraktion von Daten aus Ländern mit schwachem Datenschutz. In einer Artikelserie wurden kürzlich auf der Website des „MIT Technology Review“ zahlreiche konkrete Beispiele dafür angeführt – beginnend bei der Rekrutierung von Fachkräften durch US-Daten-Labeling-Firmen zu minimalsten Löhnen im krisengeschüttelten Venezuela bis hin zu einer neuen digitalen sozialen Kluft zwischen Weiß und Schwarz in Südafrika („digital apartheid“).

([www.technologyreview.com/2022/04/19/1049592/](http://www.technologyreview.com/2022/04/19/1049592/)) ✖

“The underrepresentation of women and people of color in technology, and the under-sampling of these groups in the data that shapes AI, has led to the creation of technology that is optimized for a small portion of the world,” Buolamwini writes in Time magazine (7 February 2019, <https://time.com/5520558/artificial-intelligence-racial-gender-bias/>).

Accusations of a “new colonialism” also play a large role in current debates on AI. Activists believe that the development of AI repeats patterns from historical colonialism – namely, the problems of the “global South”, the power of Western industrialized countries, and global corporations using their dominance solely to maximize profits. This ranges from the exploitation of labor to the mass extraction of data from countries with weak data protection. A series of articles on the MIT Technology Review website recently cited numerous concrete examples of this – starting with the recruitment of skilled workers by US data-labelling companies at the lowest possible wages in crisis-ridden Venezuela through to a new digital social divide between white and black in South Africa (“digital apartheid”).

([www.technologyreview.com/2022/04/19/1049592/](http://www.technologyreview.com/2022/04/19/1049592/)) ✖

## Green IT: Wie KI mit Energieverbrauch und Emissionen zusammenhängt

**Aus ökologischer Sicht hat Künstliche Intelligenz zwei Seiten: Zum einen gibt es riesige Potenziale für die Schonung von Ressourcen. Zum anderen sind die Rechenprozesse selbst sehr energieaufwendig.**

Das Internet ist für ungefähr ebenso viele Treibhausgasemissionen verantwortlich wie der globale Flugverkehr, hat die Internationale Energieagentur kürzlich errechnet. Die Rechenzentren, über die der Datenverkehr abgewickelt wird, verbrauchen demnach jährlich 200 bis 250 Terawattstunden (TWh = Milliarden kWh) Strom; das entspricht rund einem Prozent des weltweiten Strombedarfs. Mit 260 bis 340 TWh noch energiehungriger ist die Datenübertragung. Damit einher gehen hohe CO<sub>2</sub>-Emissionen. Ein Beispiel: Eine einzelne Suche in einer Internet-Suchmaschine bewirkt 0,5 Gramm an CO<sub>2</sub>-Emissionen. Oder: Eine längere E-Mail verursacht so viele Treibhausgase wie 100 Meter Auto fahren (was aber immer noch zumindest zehnmal klimafreundlicher ist, als einen Brief zu schreiben und zu verschicken). Zusammen mit weiteren Informations- und Kommunikationstechnologien (IKT) summiert sich all das auf rund sieben Prozent des weltweiten Stromverbrauchs bzw. auf rund zwei Prozent der jährlichen Treibhausgasemissionen – die Größenordnung des Flugverkehrs.

### Entkopplung von Datenverkehr und Energie

Angesichts des stürmischen Wachstums der verarbeiteten Datenmengen und des Datenverkehrs sowie der nötigen Infrastruktur würde man eigentlich erwarten, dass auch der Stromverbrauch und die korrespondierenden Emissionen stark ansteigen. Doch die Erfahrung des vergangenen Jahrzehnts spricht eine andere Sprache: Laut IEA-Bericht „Energy Efficiency 2021“ ist der Energieverbrauch des Internets zuletzt sogar leicht gesunken. Der Grund dafür sind Effizienzsteigerungen in Rechenzentren und Datenetzen um jährlich 10–30 Prozent. So sind z. B. die heutigen Mobilfunknetze (4G) fünfmal so energieeffizient wie 3G-Netze und 50-mal effizienter als 2G. Dadurch kam es zu einer „Entkopplung“ des Energieverbrauchs vom Datenverkehr, so die IEA-Expert:innen (wobei dabei der Energieverbrauch für die Errichtung der nötigen Hardware nicht berücksichtigt ist). Gleichzeitig warnen sie aber, dass das nicht so bleiben müsse. Denn die Datenströme

## Green IT: How AI relates to energy consumption and emissions

**From an ecological point of view, Artificial Intelligence shows two sides: It offers huge potential for conserving resources, for one thing. On the other hand, the computing processes themselves are very energy-intensive.**

The international energy agency recently calculated that the Internet is responsible for about as many greenhouse gas emissions as global air travel. Based on this calculation, the data centers through which data traffic is processed consume 200 to 250 terawatt hours (TWh = billion kWh) of electricity per year; this corresponds to around one percent of global electricity demand. At 260 to 340 TWh, the transmission of data is even more energy-hungry. This causes high CO<sub>2</sub> emissions. A case in point: a single search in an Internet search engine causes 0.5 grams of CO<sub>2</sub> emissions. And: a longer e-mail causes as many greenhouse gases as driving a car 100 meters (which is, however, still at least ten times more climate-friendly than writing and sending a letter). Together with other information and communication technologies (ICT), this all adds up to around seven percent of worldwide electricity consumption and around two percent of annual greenhouse gas emissions – the same order of magnitude as air travel.

### Decoupling data traffic and energy

Given the tremendous growth in data volumes processed and data traffic, as well as the required infrastructure, one would actually expect power consumption and the corresponding emissions to rise sharply as well. But the experience of the past decade speaks a different language: According to the IEA's "Energy Efficiency 2021" report, the Internet's energy consumption has actually even fallen slightly recently. The reason for this are efficiency gains in data centers and data networks of 10–30 percent per year. For example, today's mobile networks (4G) are five times more energy-efficient than 3G networks and 50 times more efficient than 2G. This has led to a "decoupling" of energy consumption from data traffic, according to the IEA experts (although this does not take into account the energy

wachsen weiter – derzeit verdoppeln sie sich alle drei Jahre. Höchst fraglich ist, ob die Effizienzsteigerungen der Infrastruktur mit dieser Entwicklung mithalten können. Das Dilemma zeigt sich etwa in der nächsten Mobilfunkgeneration 5G: Man erwartet zwar, dass diese 10–20-mal sparsamer mit Energie umgeht – aber die höhere Geschwindigkeit und die zusätzlichen Möglichkeiten werden auch das übertragene Datenvolumen weiter in die Höhe treiben.

### **KI treibt Entwicklung voran**

Als große Treiber der verarbeiteten Daten gelten insbesondere die Bereiche Cloud Computing, Virtuelle Realität, Blockchains (die allein 100 TWh Strom benötigen) und Künstliche Intelligenz. Dabei spielen zwei Faktoren eine Hauptrolle: Zum einen laufen KI-Prozesse typischerweise auf zentralen Rechnern in der Cloud ab – Daten und Ergebnisse werden mit den Endgeräten ausgetauscht, wodurch der Datenverkehr wächst. Zum anderen ist insbesondere die Trainingsphase von Maschinenlernsystemen sehr langwierig und energieaufwendig. Umfassende Abschätzungen über die Größenordnungen, um die es dabei geht, fehlen in der Fachliteratur noch. Es gibt aber Anhaltspunkte: So weiß man etwa, dass die Lernphase bei der Gesichtserkennung oder der Verarbeitung natürlicher Sprachen selbst bei leistungsfähigster Hardware Wochen bis Monate dauern kann. Geschätzt wird, dass sich in den vergangenen Jahren die Rechenleistung für das Maschinenlernen alle drei Monate verdoppelt hat.

Dem ist gegenüberzustellen, dass KI-Systeme auch einen effizienteren Betrieb der Dateninfrastruktur ermöglichen sollten. So wie es überhaupt viele Hoffnungen gibt, dass intelligentere Computersysteme die Effizienz in vielen Bereichen stark steigern könnten. Mehr noch: Die fortschreitende Digitalisierung liefert einen wichtigen Beitrag zur Transformation des Wirtschafts- und Gesellschaftssystems – dies reicht von neuen Produktionssystemen in der Industrie und „intelligenteren“ und ökologischeren Produkten über Verkehrs-, Logistik- und Energiesysteme bis hin zu neuen Mechanismen der gesellschaftlichen Koordination – inklusive Verhaltensänderungen der Menschen. In Summe erwartet z. B. der UN-Weltklimarat IPCC ein Potenzial zur Einsparung von 15 Prozent der Treibhausgasemissionen.

### **Einsatz von KI für ökologische Ziele**

KI-Systeme können ökologische Ziele auf vielen Ebenen unterstützen. Wie weit der Bogen der Anwendungsfelder ist, wird z. B. aus der Ausschreibung für das vom Klimaschutzministerium (BMK) im Vorjahr neu gestartete Forschungsprogramm „AI for Green“ deutlich:

#### **1. KI zur Bewältigung des Klimawandels und zur Anpassung an dessen Folgen.**

Die gestiegene Verfügbarkeit von Fernerkundungsdaten und deren Nutzung durch KI-Systeme eröffnet eine Vielzahl neuer Möglichkeiten, um die Umgebung zu erfassen und fundierte Entscheidungsgrundlagen

consumption for setting up the necessary hardware). At the same time, however, they warn that it doesn't have to stay that way: data streams continue to grow – currently doubling every three years. It is very difficult to predict whether the efficiency improvements in the infrastructure will be able to keep pace with this development. The dilemma manifests itself, e.g., in the next generation of mobile communications, 5G: while it is expected that it will be 10–20 times more energy-efficient, the higher speed and additional capabilities will also drive up the volume of data transmitted.

### **AI is driving development forward**

Cloud computing, virtual reality, blockchains (which alone require 100 TWh of electricity) and Artificial Intelligence in particular are considered to be big drivers of processed data. Two factors play a major role here: first of all, AI processes typically run on centralized computers in the cloud – data and results are shared with end devices, which increases data traffic. Secondly, the training phase of machine-learning systems in particular is very lengthy and energy-consuming. Comprehensive estimates of the orders of magnitude involved are still lacking in expert literature. There are clues, however: for example, we know that the learning phase for facial recognition or the processing of natural languages can take weeks to months, even with the most powerful hardware. It is estimated that computing power for machine learning has doubled every three months in recent years. This should be viewed alongside the fact that AI systems should also enable a more efficient operation of the data infrastructure. Just as there are generally many hopes that more intelligent computer systems could greatly increase efficiency in many areas. What's more: advancing digitization provides an important contribution to the transformation of the economic and social system – ranging from new production systems in industry and “smarter” and more ecological products to transport, logistics, and energy systems and new mechanisms of social coordination – including changes in people's behavior. Overall, the UN Intergovernmental Panel on Climate Change (IPCC), for example, expects a potential reduction of 15 percent in greenhouse gas emissions.

### **Use of AI for ecological goals**

AI systems can support ecological goals on many levels. Just how wide the range of application fields is becomes clear, e.g., from the call for proposals for the “AI for Green” research program newly launched by the BMK (Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology) last year:

#### **1. AI for tackling climate change and adapting to its consequences.**

The increased availability of remote sensing data and its use by AI

zu liefern. KI kann dabei unterstützen, aktuelle Einblicke in die Klimaentwicklung zu geben und akkurater zu prognostizieren, welche Änderungen bis wann zu erwarten sind. Klimavorhersagen können dabei auf einer viel kleinteiligeren Skala als bisher erfolgen.

2. **KI für den Klimaschutz und zur Reduktion der Treibhausgasemissionen**

KI kann zur Einbindung von erneuerbaren Energiequellen in das Energienetz genutzt werden. Durch die Analyse von Wetterdaten oder historischen Verbrauchszahlen können genaue kurz- und mittelfristige Prognosen des Strombedarfs erstellt werden. Da die erzeugten Strommengen aus Photovoltaik und Windkraftwerken variieren, können derartige Prognosen dabei helfen, den Betrieb des Stromnetzes und dessen Aufrechterhaltung größtenteils durch erneuerbare Energien sicherzustellen.

3. **KI für Nachhaltigkeit in der Fertigung und Sachgüterindustrie**

KI kann zur Minimierung des Ressourcenverbrauchs (Energie, Rohstoffe, Wasser usw.) in Produktionsprozessen beitragen. Darüber hinaus ermöglicht der Einsatz von KI-Systemen mehr Flexibilität in Fertigungs- und Produktionsprozessen, wodurch spezifische Kund:innenbedürfnisse adressiert und Überproduktion sowie Verschwendung vermieden werden können.

4. **KI für eine nachhaltigere Lebensweise**

Die Umstellung auf eine nachhaltigere, ressourcenschonendere Lebensweise reicht in den individuellen Lebensstil des Menschen hinein – etwa hinsichtlich Nutzungs- oder Entsorgungsverhalten von Ressourcen im Alltag. Gerade in Bereichen wie Energie, Mobilität, Ernährung und Mode wird die Aufmerksamkeit für ein nachhaltiges Konsumbewusstsein und Handeln weiter steigern. KI kann dabei z. B. eingesetzt werden, um nutzer:innenorientiert die ressourcenschonendste Kombination von Mobilitätsangeboten für die jeweilige Route zu identifizieren.

5. **KI in der Land- und Forstwirtschaft**

Entscheidungen über den Einsatz technischer Mittel in der Land- und Forstwirtschaft basieren häufig auf generationsübergreifenden Erfahrungen. Intelligente, KI-basierte Werkzeuge können diesen traditionellen Prozess weitreichend unterstützen, verändern oder auch optimieren. So können etwa bei hoher Volatilität des Wetters Satelliten- und Wetterdaten sowie Daten von Bodensensoren eine wertvolle Grundlage bieten, um mittels KI Vorhersagemodelle zur nachhaltigen Optimierung der landwirtschaftlichen Produktion zu erstellen.

systems is opening up a wide range of new opportunities to capture the environment and provide well-founded decision-making bases. AI can help provide up-to-date insights into climate trends and help make more accurate predictions about what changes to expect by when. Climate predictions can thereby be made on a much more precise scale than before.

2. **AI for climate protection and the reduction of greenhouse gas emissions**

AI can be used to integrate renewable energy sources into the energy grid. By analyzing weather data or historical consumption figures, accurate short- and medium-term forecasts of electricity demand can be made. Since the amounts of electricity generated from photovoltaic and wind power plants vary, such forecasts can help ensure that the power grid operates and stays up and running mostly with renewable energy.

3. **AI for sustainability in manufacturing and the material goods industry**

AI can contribute to minimizing resource consumption (energy, raw materials, water, etc.) in production processes. Moreover, the use of AI systems enables more flexibility in manufacturing and production processes, which means that specific customer needs can be addressed and overproduction and waste can be avoided.

4. **AI for a more sustainable lifestyle**

The switch to a more sustainable, resource-conserving lifestyle extends into people's individual lifestyles – e.g., in terms of how they use or dispose of resources in their daily lives. Particularly in areas such as energy, mobility, nutrition, and fashion, attention to sustainable consumer awareness and action will continue to increase. AI can be used, e.g., to identify the most resource-efficient combination of mobility options for a given route from a user-oriented perspective.

5. **AI in agriculture and forestry**

Decisions on the use of technical resources in agriculture and forestry are often based on inter-generational experiences. Intelligent, AI-based tools can go a long way in supporting, changing or even optimizing this traditional process. In this vein, in cases of high weather volatility, satellite and weather data, as well as data from soil sensors, can provide a valuable basis for creating predictive models using AI to optimize agricultural production in a sustainable manner.

6. **KI für den Artenschutz und den Erhalt der Artenvielfalt**

Die Tier- und Pflanzenwelt ist durch unterschiedliche Einflüsse gefährdet (etwa durch den Einsatz von Pestiziden in der Landwirtschaft, durch Schädlinge oder Krankheiten). Intelligente Sensorik und Bildgebungstechniken können dazu beitragen, das Verhalten von Insektenarten besser zu verstehen und den Bestand der Arten zu sichern. Die automatisierte Auswertung von Daten beim Artenschutz nimmt eine immer wichtigere Rolle ein. Z. B. werden zur Abschätzung der zeitlichen Veränderung von Populationsgrößen vermehrt autonome Aufnahmegeräte genutzt.

7. **KI zur Bewältigung der Auswirkungen des Klimawandels auf die Gesundheit**

Der Klimawandel wird künftig zu einer Zunahme von Extremwetterereignissen führen, die eine direkte Auswirkung auf die Gesundheit der Menschen haben. KI stellt in diesem Zusammenhang ein wichtiges Instrument zur Verfügung. Weitere Forschungsaktivitäten reichen von der Steigerung der Effizienz von Krankheitssimulatoren bis hin zu exakten Messungen von Schadstoffexposition und damit verbundenen gesundheitlichen Auswirkungen.

Bei all diesen Themen soll AI nicht nur eine Unterstützung bei der Bewältigung der ökologischen Herausforderungen („AI for Green“) sein, sondern die AI-Systeme selbst müssten möglichst ressourcenschonend und energiesparend sein („Green AI“). Die erforschten und entwickelten AI-Systeme sollen möglichst präzise Ergebnisse liefern, ohne dabei den Rechenaufwand zu erhöhen bzw. sie sollten diesen im Idealfall sogar reduzieren. ✖

6. **AI for the protection of species and the preservation of biodiversity**

Flora and fauna are endangered by various influences (such as by the use of pesticides in agriculture, by pests or diseases). Intelligent sensors and imaging techniques can help us better understand the behavior of insects and safeguard the species' continued existence. The automated evaluation of data in species conservation is taking on an increasingly important role. For example, autonomous recording devices are increasingly being used to estimate changes in population sizes over time.

7. **AI for tackling the impact of climate change on health**

Climate change will lead to an increase in extreme weather events in the future, which will have a direct impact on human health. AI provides an important tool in this regard. Further research activities range from increasing the efficiency of disease simulators to accurate measurements of pollutant exposure and associated health impact.

In all of these topics, AI should not only offer support in addressing ecological challenges (“AI for Green“), but the AI systems themselves would have to be as resource and energy-efficient as possible (“Green AI“). The AI systems that are researched and developed should deliver results that are as precise as possible without increasing the computational effort or, ideally, they should even reduce it. ✖



„Es geht um die Frage, ob ein KI-System etwas besser macht als der Mensch“

**Die Mobilitätsforscherin Katja Schechtner über Probleme und Chancen von KI-gestützten Mobilitätssystemen, über Abwägung von Nutzen und Privacy sowie über kulturelle Unterschiede beim Umgang mit neuen Technologien.**

**Sie beschäftigen sich seit 20 Jahren mit Themen wie neue Mobilitätssysteme und Smart Cities. Gehe ich recht in der Annahme, dass die neuen Möglichkeiten, die die Digitalisierung bietet, in diesen Bereichen sehr willkommen sind?**

**Katja Schechtner:** Ja, aus meiner Perspektive gibt es kaum ein System, das sich so gut für den Einsatz von Künstlicher Intelligenz (KI) eignet, wie das Verkehrssystem. Denn Mobilität basiert auf Regeln – wie etwa: Bei Rot bleibt man stehen, über diese Linie darf man nicht drüberfahren usw. Wir haben uns dieses Regelsystem gegeben, um gewisse gesellschaftliche Ziele zu erreichen, wir monitoren die Regeln und belegen sie, falls sie gebrochen werden, mit Sanktionen.

Das lässt sich sehr gut in Algorithmen abbilden. KI bringt dabei viele neue Möglichkeiten – und wird auch schon stark eingesetzt. KI kommt nicht, sondern ist bereits fest in unserem Leben verankert. Das sieht man besonders gut, wenn man durch eine Stadt geht. Etwa wenn man der Route eines Routenplaners folgt. Oder wenn man sich in einer Grünen Welle bewegt. Oder auch, wenn an der Station angezeigt wird, wann die nächste Straßenbahn mit Kinderwageneinstieg kommt. Oder bei E-Scootern, die jetzt neu mit Geo-Fencing ausgestattet werden: Wenn man versucht, auf den Gehsteig zu fahren, bremsen sie automatisch. Das alles ist KI.



**KATJA SCHECHTNER** ist Stadtforscherin und Mobilitätsexpertin. Zuletzt entwickelte sie Technologie- und Innovationspolitik mit den Schwerpunkten Data Science, AI und Blockchain im Bereich Verkehr bei der OECD, erarbeitete für die Asian Development Bank Infrastrukturprojekte in asiatischen Großstädten und baute davor am AIT Austrian Institute of Technology eine Mobilitätsforschungsabteilung auf. Parallel dazu ist sie seit 2010 Gastforscherin am MIT in Boston und seit 2022 auch Gastprofessorin an der Angewandten in Wien. In Österreich ist sie Aufsichtsrätin der AIT und der OENPAY.

Discussing AI

“The question is whether an AI system can do something better than a human can”

**Mobility researcher Katja Schechtner on the problems and opportunities for AI-supported mobility systems, on weighing up benefits and privacy, as well as on cultural differences when dealing with new technologies.**

**You have been involved with topics such as new mobility systems and smart cities for 20 years now. I am right in assuming that the new options provided by digitization are very welcome in these areas?**

**Katja Schechtner:** Yes, from my point of view there is hardly a system better suited for implementing Artificial Intelligence (AI) than the transportation system. That's because mobility is based on rules, such as: you stop when it's red, you're not allowed to drive past that line, etc. We have set ourselves this system of rules in order to achieve certain goals as a society, we monitor the rules and when they are broken we impose sanctions.

That can be reproduced very well by algorithms. AI offers many new opportunities – and is already being used extensively. AI is not on its way, it is already anchored in our lives. You can see this particularly well when you walk around a city. For example, when you are following a route provided by a route planner. Or if you are moving through phased traffic lights. Or when the station shows when the next tram with a buggy entrance will arrive. Or e-scooters that are now equipped with geo-fencing: if you try to ride on the pavement, they brake automatically. All of this is AI.

Diskussionen über KI

### Das Verkehrssystem wird also immer intelligenter ...

Eine Kernthese von Marvin Minsky lautet: Die menschliche Intelligenz ist eine „society of mind“. Unser Geist besteht aus vielen hoch spezialisierten einzelnen Stückchen von Informationsverarbeitung, wie zum Beispiel Mustererkennung, Sprachverständnis usw. Diese einzelnen Prozesse zusammengenommen ergeben emergente Intelligenz. In der Verkehrstechnik haben wir bereits viele kleine Dinge, von der Section Control bis hin zu globalen Lieferketten. Hier werden unheimlich viele Daten erfasst und analysiert, dadurch wird wirklich in das Leben der Menschen eingegriffen. In den vergangenen zehn Jahren wurden immer mehr individuelle Technologien miteinander kombiniert, und daraus entsteht etwas Emergentes.

### Was kann ich mir darunter konkret vorstellen?

Zum Beispiel in der Stadtplanung. Vor zehn Jahren haben wir Menschen Bilder von Stadtvierteln gezeigt und sie gefragt, wo sie sich sicher fühlen und wo eher nicht. Mithilfe statistischer Methoden haben wir das analysiert und mit anderen Daten wie Kriminalität verglichen. Nun können wir das mit KI viel genauer machen: Wir legen Computern Bilder vor, mit deren Hilfe sie sich selbst einen multidimensionalen „Sicherheitsraum“ antrainieren. Dieser ist so komplex, dass man das als Mensch nicht mehr nachvollziehen kann, aber wir können damit Aussagen treffen, ob sich ein Mensch dort sicher oder unsicher fühlt. Das ist die Emergenz von KI, das ist das, was wir von KI lernen können. Wir können mittels Algorithmen sehr viel Wissen generieren, auf das wir dann reagieren können.

### Wie können wir solches Wissen umsetzen?

Einer der Bereiche, mit denen ich mich seit 20 Jahren beschäftige, ist „crowd dynamics“: Wie bewegen sich Menschen in großen Menschenmassen? Jeder glaubt ja, dass er oder sie individuell unterwegs ist, aber wenn man das untersucht, ist das auf einer hohen Ebene fast wie das Fließen von Flüssigkeiten. Wenn man sich beispielsweise ansieht, wie schnell Menschen in verschiedenen Situationen gehen – ob in der Einkaufsstraße, nach einem Konzert oder in einem Bahnhof –, kann man Leitsysteme schaffen, mit denen sich Menschen immer so sicher fühlen, dass sie das öffentliche Verkehrssystem gerne nutzen. Man kann Menschenmengen steuern und die Menschen so beeinflussen, wie das im jeweiligen Kontext sicher und effizient ist. Das haben wir schon vor bald 15 Jahren in Wien umgesetzt. Heutzutage geht das mit Methoden der KI so weit, dass man im Verkehr sehr detailliert und zeitabhängig herausfinden kann, welche Bedürfnisse an welchen Stellen auftreten werden. Mit diesem Wissen kann man etwa vorhersagen, welche Verkehrsmittel zu einem bestimmten Zeitpunkt an einem bestimmten Ort benötigt werden. Dann kann man die Mobilität

### So the traffic system is getting smarter and smarter ...

One of Marvin Minsky's core propositions is: human intelligence is a "society of mind". Our minds are capable of highly specialized individualized bits of information processing, such as pattern recognition, language understanding, etc. These individual processes combine to form emergent intelligence. In traffic engineering we already have a lot of small things, from section control to global supply chains. An incredible amount of data is collected and analyzed here, which really has an effect on people's lives. Over the past ten years, an increasing number of individual technologies have been combined to create something newly emergent.

### What does this mean specifically?

Take urban planning for instance. Ten years ago we showed people pictures of neighborhoods and asked them where they felt safe and where they didn't. We analyzed this using statistical methods and compared it with other data such as data on crime. We are now able to do this more precisely with AI: We present computers with images that they use to train themselves regarding a multidimensional "safe space". The situation is so complex that as a human being you can no longer understand it, but we can use it to make statements about whether a person will feel safe or unsafe there. This is the emergence of AI – this is what we can learn from AI. We can generate a great deal of knowledge using algorithms, which we can then react to.

### How can we implement knowledge of this kind?

One of the areas that I have been involved in for 20 years is "crowd dynamics": How do people move when they are in large crowds? Everyone believes that they are moving as an individual, but when you look at it, at a higher level, it's almost like the flow of liquids. For example, if you look at how fast people walk in different situations – in a shopping street, after a concert, or in a train station – you can create signage systems that always make people feel so safe that they are happy to use public transport. You can control groups of people and influence them in ways that are safe and efficient in the respective context. We have already been implementing this in Vienna for nearly 15 years. Today, AI methods have been taken so far that when you are in traffic you can find out what will be needed at what location in a very detailed way and based on the time. You can use this information, e.g., to predict what means of transportation will be required at what time at a certain location. Then mobility can also be made available in a sustainable way. This is already being done in Singapore, for

auch nachhaltiger zur Verfügung stellen. Das wird beispielsweise in Singapur schon gemacht. Wenn man weiß, wie sich Menschen bewegen, und das mit anderen Daten kombiniert, kann man auch sehr detailliert untersuchen, wie sich Menschen in konkreten Situationen bewegen. Ein Beispiel: Vielfach nimmt man an, dass Menschen beim Zufußgehen immer die kürzeste Route wählen. Aber in der Praxis weichen sie häufig davon ab. Warum? Mit Hilfe von KI findet man unerwartete Kombinationen von Faktoren, die weit über konventionelle Antworten wie „jeder versucht, durch den grünen Park zu gehen“ hinausgehen. Daraus kann man lernen, wie man Stadträume so gestalten kann, dass sie ein nachhaltiges Bewegungsverhalten fördern und dadurch die Gesundheit positiv beeinflusst wird, dass sich Menschen wohlfühlen und dass auch noch die Straße belebt wird, wodurch man zusätzliche Sicherheits-effekte erzielen kann.

#### **Bei KI dreht sich alles um Daten. Damit ist man sofort bei den Themen Überwachung im öffentlichen Raum, Schutz der Privacy, Datenschutz usw. ...**

Es gibt dieses Dilemma: Menschen möchten, dass ihre Privacy respektiert wird. Gleichzeitig stellen sie aber hohe Ansprüche daran, wie Städte funktionieren sollen. Wir haben das in der Corona-Pandemie gut gesehen. Um die Balance zwischen Privacy und Nutzen zu finden, hat man einerseits technische Möglichkeiten, andererseits müssen wir verstehen, wie Menschen die Situation in bestimmten Kontexten einschätzen.

Wir glauben nicht, dass man das Thema Privacy durch rein technische und regulatorische Maßnahmen lösen kann, wenn man die Möglichkeiten ausschöpfen will. Um die Balance zu finden, braucht man eine „risk-weighted regulation“: eine Art Kosten-Nutzen-Analyse – wobei die Kosten nicht im Sinne von Geld bewertet werden, sondern hinsichtlich Privacy. Dazu müssen wir zuerst einmal den gesellschaftlichen Impetus bestimmter Maßnahmen verstehen. Wir schauen uns gerade in einem Projekt am MIT an, für welche Datensätze in welcher Kombination es für die Menschen ok ist, dass sie genutzt werden, wenn sie dafür eine bestimmte Leistung von der Stadt bekommen – z.B. nachhaltigere Müllsam-melsysteme oder Bedarfshaltestellen für Busse.

#### **Wie kann die Governance von KI-Systemen im Verkehrs- oder im Stadtbereich aussehen?**

Dass sich Dinge auch einmal verändern, weil sich die Technologie verändert, ist eine riesige Herausforderung für Regulatoren. Darüber haben wir uns in jüngster Zeit bei der OECD Gedanken gemacht und ein Paper darüber veröffentlicht, wie Governance im Zeitalter der Algorithmen („algorithmic age“) funktionieren kann. Traditionelle Governance beruht auf dem Grundsatz „regulate and

example. If you know how people move and combine that with other data, you can also examine in great detail how people move in specific situations. A case in point: it is often assumed that people always choose the shortest route when walking. But in practice, they often deviate from this. Why? Using AI, you find unexpected combinations of factors that far exceed conventional answers such as “everyone is trying to walk through the green park”. From this you can learn how urban spaces can be designed in such a way that they promote long-lasting movement behavior and this has a positive effect on health. People need to feel at ease and the street should also be lively, which also achieves additional safety effects.

#### **AI is all about data. This takes you straight on to the topics of surveillance in public spaces, protection of privacy, data protection, etc. ...**

There is this dilemma: people want their privacy respected. At the same time, they have high standards regarding how a city should function. We saw this quite clearly during the COVID pandemic. On the one hand, there are technical options for striking a balance between privacy and the benefits on offer, on the other hand, we need to understand how people view the situation in certain contexts.

We do not believe that the issue of privacy can be solved by purely technical and regulatory measures if you want to take advantage of all the opportunities. In order to find the balance, you need a “risk-weighted regulation”: a kind of cost-benefit analysis – in which the costs are evaluated not in terms of money, but in terms of privacy. To do this, we must first understand the social impetus behind certain measures. In a project at MIT, we are currently looking into the question of which datasets and in what combination people think it is ok to use them if in exchange the city provides them with a certain service, e.g., a more sustainable waste collection system or request stops for buses.

#### **What could the governance of AI systems in the area of traffic or the city look like?**

The fact that things change because technology changes is a huge challenge for regulators. We have been thinking about this recently at the OECD and published a paper on how governance can work in the algorithmic age. Traditional governance is based on the principle of “regulate and forget”: That means you settle one thing once and then you don’t want to deal with the whole thing again anymore. In private industry, on the other hand, the motto is: “deploy and re-make” – in other words, we make something, try it out, and then improve it. New public governance systems, on the other hand,

forget“: Das heißt, man regelt eine Sache einmal, und dann will man das Ganze nicht mehr angreifen. In der Privatindustrie heißt es hingegen „deploy and re-make“ – also wir machen etwas, probieren etwas aus und verbessern dann. Die neuen öffentlichen Governance-Systeme müssen hingegen dem Motto „guide and iterate“ folgen. Eine Möglichkeit dafür sind sogenannte „regulatory sand boxes“. Damit versucht man, eine gewisse Form von Technologie, die algorithmischer Steuerung und die dazugehörigen Gesetze erst einmal in einem bestimmten Raum zu ermöglichen und dadurch Erfahrungen zu sammeln. Diese Idee ist gar nicht neu: der Vorläufer des Allgemeinen Bürgerlichen Gesetzbuchs (ABGB) wurde in der Habsburger Monarchie auch erst einmal in einer Region für ein paar Jahre eingeführt – und man hat geschaut, was die Leute dazu sagen, was funktioniert und was nicht. Das ist ein altes Prinzip, das nun wieder stärker eingesetzt wird.

### Können Sie mir ein Beispiel geben?

Im Department of Transportation in Los Angeles ist man der Meinung, dass es nicht deren Aufgabe ist, Beton zu gießen, sondern sich um eine algorithmische Steuerung der Nutzung der Infrastruktur zu kümmern. So darf etwa an manchen Tagen und an manchen Orten ein Uber eine:n Passagier:in abholen, an manchen aber nicht. Da werden sehr flexible Regulierungen getestet, die teilweise nur mehr in den Maschinen – also in der Navigationssoftware der Anbieter:innen – „embedded“ sind. Wir nennen das „managing the curb“. So könnte es zum Beispiel möglich werden, manche Busstationen in einem 24-Stunden-Zyklus in der Früh von großen Bussen anzusteuern; untertags könnten diese Stationen dann ein Café sein; und am Abend werden die Tische wieder weggeräumt und man kann sich von Taxis oder kleinen Bussen abholen lassen.

Bei all dem stellt sich natürlich eine grundlegende Frage: Sollen wir überhaupt „algorithmic governance“ einsetzen? Kann eine algorithmische Steuerung etwas wirklich besser als eine menschliche Entscheidung? Es geht bei vielen Dingen nicht darum, etwas perfekt zu machen, sondern darum, ob ein System etwas besser macht als der Mensch – und es auch besser für alle Menschen macht? „Section control“ zum Beispiel ist ein KI-System, das es besser macht als der Mensch: Es erhöht die Verkehrssicherheit in Hochrisikobereichen wie etwa Tunnels. Im Umkehrschluss bedeutet das auch: Man muss die Technologieanbieter fragen, ob es wirklich nötig ist, was sie tun. Wenn es nicht besser ist als das, was der Mensch kann, und wenn es keinen klaren „benefit“ gibt, muss man es auch nicht machen. In Städten gib es jedenfalls über den Verkehr hinaus viele Systeme, etwa Energie, Abwasser, Abfall usw., in denen man mit intelligenter Steuerung so eingreifen kann, dass große gesellschaftliche Ziele – sowohl soziale als auch Umweltziele – gut erreicht werden können. ✖

have to follow the motto “guide and iterate”. One option for this is “regulatory sand boxes”. This is an attempt to first enable a certain form of technology, algorithmic control and the associated laws in a certain area and gain experience by doing so. This idea is not at all new: in the Hapsburg monarchy, the predecessor of the Austrian General Civil Code (ABGB) was introduced for a few years in one region – and they looked at what people had to say about what worked and what didn’t. It is an old principle that is now being implemented again to a greater extent.

### Can you give me an example?

The Department of Transportation in Los Angeles believes that its job is not to pour concrete, but to handle algorithmic control of infrastructure use. For example, on some days and in some locations, an Uber is allowed to pick up a passenger, but not on others. Very flexible regulations are being tested there, some of which are “embedded” directly in the machines – i.e. in the provider’s navigation software. We call this “managing the curb”. For example, as part of a 24-hour cycle it could be possible for large buses to drive into some bus stations in the morning; during the day these stations could then be a café; and in the evening the tables are cleared away again and people can be picked up by taxis or small buses.

With regard to all this, there is of course an underlying question: should we use “algorithmic governance” at all? Is it really possible for algorithmic control to be better than a human decision? In many cases it is not about making something perfect, but about whether a system does something better than humans – and whether it thus improves things for everyone. “Section control”, e.g., is an AI system that does it better than humans: It increases traffic safety in high-risk areas such as tunnels. Conversely, this also means that we need to ask technology providers if what they are doing is really necessary. If it is no better than what humans can do and if there is no clear benefit, then we don’t need to do it. In any case, in addition to traffic, there are many systems in cities, such as energy, sewage, waste, etc., in which it is possible to intervene with intelligent control in such a way that major goals for society – both social and environmental – can be achieved in a positive way. ✖



„KI ist ein potentes Werkzeug“



#### PROCESS STUDIO

Künstler: Martin Grödl,  
Moritz Resl

Unter dem Namen Process – Studio for Art and Design entwickeln Martin Grödl und Moritz Resl interaktive, audiovisuelle Installationen, Grafikdesign und Software als Tool für gestalterische Prozesse. Beide haben in Wien sowohl an der Universität für angewandte Kunst und der Technischen Universität studiert und ihre gemeinsamen Arbeiten u.a. bereits im MAK – Museum für Angewandte Kunst, der Vienna Biennale for Change, der London Design Biennale und der Ars Electronica ausgestellt. Im Jahr 2022 sind sie Artists in Residence am AIT Austrian Institute of Technology.  
[www.process.studio](http://www.process.studio)

**Moritz Resl und Martin Grödl, bekannt als Künstlerduo „Process Studio“, nutzen Künstliche Intelligenz als eines von vielen Werkzeugen. Sie sehen den Nutzen von KI in erster Linie in der Wechselwirkung zwischen Mensch und Maschine: KI an sich ist nicht kreativ, kann beim Menschen aber kreative Prozesse anstoßen.**

**Was interessiert euch als Künstler an den Methoden der Künstlichen Intelligenz (KI)? Warum beschäftigt ihr euch mit dieser Technologie?**

**Moritz Resl:** Prinzipiell ist für uns jede Technologie interessant, denn Technologien sind unsere Werkzeuge. KI ist derzeit ein riesiges Buzzword. Es ist das Ding, auf das sich alle Technologiekonzerne jetzt stürzen und die Entwicklung in ihrem eigenen Interesse vorantreiben. Wir sind dabei Außenseiter: Wir sind ein kleines Grafikbüro und sehen uns an, wie wir mit dieser Technologie arbeiten können. Wir schauen, wie man sie nicht nur für kommerzielle Zwecke einsetzen kann, sondern ihr einen anderen Anstrich geben könnte – indem man sie für einen gemeinnützigeren Zweck einsetzen könnte, abseits vom reinen Geldmachen.

**Martin Grödl:** Dazu kommt noch, dass KI in der Diskussion mystisch überhöht wird, weil damit eine ganze Bandbreite von praktischen Applikationen bis hin zu wildester Science Fiction verbunden ist; das reicht hin bis hin zur sogenannten „Singularität“, die den Menschen übertrifft. Es ist natürlich extrem spannend, sich in diesem Diskussionsfeld zu bewegen. Man merkt, dass sich Wissenschaftler:innen, die

“AI is a potent tool”

#### PROCESS STUDIO

Artist: Martin Grödl,  
Moritz Resl

Under the name Process – Studio for Art and Design, Martin Grödl and Moritz Resl develop interactive, audio-visual installations, graphic design and software as a tool for design processes. Both studied at the University of Applied Arts Vienna and TU Wien and their joint work was already exhibited in settings such as MAK – Museum of Applied Arts, the Vienna Biennale for Change, the London Design Biennale and Ars Electronica. In 2022, they are artists in residence at the AIT Austrian Institute of Technology.  
[www.process.studio](http://www.process.studio)

**Moritz Resl and Martin Grödl, an artist duo known as “Process Studio“, use Artificial Intelligence as one of many tools. They see the usefulness of AI primarily in the interaction of human and machine: AI, they say, is not creative in and of itself but can set off human creative processes.**

**What is it that interests you as artists about the methods of Artificial Intelligence (AI)? Why do you concern yourself with this technology?**

**MORITZ RESL:** Basically, any technology is interesting for us because technologies are our tools. AI is a huge buzzword these days. It’s the thing that all technology companies are pouncing on right now, driving the development in their own interest. We are outsiders here: we are a small graphic-design office looking at what we can do with this technology. We are looking at how to use it not only for commercial purposes but to give it a different touch—by employing it for the common good rather than just to make money.

**MARTIN GRÖDL:** On top of that, AI is mystically magnified in the discussion because there is a whole range of practical applications all the way up to the wildest science fiction that is associated with it, including the so-called “singularity,” when it surpasses human intelligence. It is, of course, extremely exciting to move in this field of discussion. You realize that scientists who really work with AI



wirklich mit KI arbeiten, auf einer ganz anderen Ebene bewegen als die große Diskussion. Wir wollen dieses Spannungsfeld aufbrechen und die Technologie ein bisschen entmystifizieren.

#### **Worum geht es euch bei euren Projekten mit KI? Was ist euer künstlerischer Anspruch?**

**MR:** Für mich ist KI ein weiteres Werkzeug, das man verwenden kann, um eine Idee umzusetzen. Es ist nicht die Verliebtheit in die Technologie an sich. Unser Interesse kommt daher, dass wir beide eigentlich aus der Informatik kommen. Wir beschäftigen uns stark mit Computern und den neuen Entwicklungen in diesem Bereich. Aber im Grunde ist der Computer nur ein Werkzeug. Im klassischen Grafikdesign ist er eine Verlängerung des Hirns und der Hand. Wir gehören einer Generation an, die mit dem Computer als Werkzeug aufgewachsen ist.

#### **Was bringt die KI in diesen Werkzeugkasten an Neuem hinein?**

**MR:** Wir arbeiten mit generativen Designs – das ist ein prozedurales System, aus dem unsere Grafiken herauskommen. So gesehen ist KI gar nicht so viel anders für uns. Der Vorteil der generativen Systeme ist, dass man eine Riesenbandbreite an Output hat.

**MG:** KI ist ein Kulminationspunkt der Datenverarbeitung, in dem alles zusammenläuft, was von Anbeginn der Computertechnik entwickelt wurde. Wenn man historisch etwas zurückblickt, war KI eine der ganz frühen Ideen in der Informatik. Schon in den 1950er-Jahren gab es einen ersten Boom, die Erwartungen konnten aber damals aufgrund von technischen Beschränkungen nicht eingelöst werden. Nun scheint es durch die Deep-Learning-Techniken zu funktionieren. An der KI kann man gut die unheimliche Schichtung in der Computertechnik nachvollziehen, bei der eine Schicht auf die anderen aufbaut. Klar wird dabei auch, dass IT-Systeme extrem mit gesellschaftlichen Systemen vernetzt sind. Da geht es unmittelbar um Fragen wie: Woher kommen die Daten? Woher kommt die Rechenleistung? Woher kommt der Strom? Woher kommen die Ressourcen? All diese Fragen stecken in dem Thema KI drin.

#### **Wenn man von Kunst redet, kommt man schnell auch auf das Thema Kreativität: Sind KI-Systeme kreativ?**

**MR:** Ein ganz klares Nein – weil diese Systeme von uns gebaut wurden. Dahinter stecken Menschen, die sich überlegen, was funktionieren kann. Ein Computer kann nur berechnen, er kann sich nichts Neues ausdenken. Er bewegt sich immer nur in seinem abgesteckten Bereich. Was wir unter Kreativität verstehen, ist das genaue Gegenteil davon: Es ist das Ausbrechen aus Gewohntem und das Schaffen neuer Dinge. Aber es gibt eine Wechselwirkung: Man kann sich als Mensch von einer KI beeinflussen lassen und so kreativ werden. Ein Beispiel: Das Spiel Go ist sehr komplex, man

are on a completely different level than the mainstream discussion. We want to break down this field of tension and demystify the technology a little.

#### **What are your projects with AI about? What is your artistic ambition?**

**MR:** For me, AI is just another tool that you can use to implement an idea. It's not being in love with the technology itself. Our interest comes from the fact that we both actually have a computer science background. We are very much involved with computers and new developments in the field. But basically, the computer is just a tool. In classical graphic design, it's an extension of the mind and the hand. We belong to a generation that grew up with the computer as a tool.

#### **What new things does AI bring to this toolbox?**

**MR:** We work with generative designs—it's a processual system that our graphics come from. Seen like that, AI isn't that much different for us. The advantage of generative systems is that you have a vast range of output.

**MG:** AI is a culmination point of computing, a confluence of everything ever developed in computer technology from the very beginning. If you look back some time in history, AI was one of the very early ideas in computing. It had its first boom as early as the 1950s, but failed to meet the expectations at the time due to technological limitations. Now it seems to be working because of deep-learning techniques. AI can give you a good idea of the uncanny layered structure of computer technology, where one layer builds upon the others. What also becomes clear is that IT systems are extremely interconnected with social systems. This directly relates to questions like: Where does the data come from? Where does the computing power come from? Where does the electricity come from? Where do the resources come from? All of these questions are involved in the topic of AI.

#### **When talking about art, the subject of creativity quickly comes up: Are AI systems creative?**

**MR:** Very clearly, no—because these systems have been built by us. Behind them are people who think about what could work. A computer can only compute, it can't think up anything new. It always stays within its defined field. What we understand by creativity is the exact opposite: it is thinking out of the box and creating new things. But there is some reciprocity: as a human, you can let yourself be influenced by AI and thus become creative. One example: the game of Go is very complex; it is actually incalculable because there are so many possibilities. In that famous game in which a computer defeated a Go

kann es eigentlich nicht berechnen, weil es so viele Möglichkeiten gibt. Bei dem berühmt gewordenen Spiel, bei dem ein Computer den Go-Meister besiegt hat, gab es einen Zug der KI, der aus menschlicher Sicht irgendwie komisch gewirkt hat, den keiner verstanden hat. Alle Beobachter:innen haben gedacht, dass die KI einen Fehler gemacht hat. Aber rückblickend war genau das der Wendepunkt des Spiels. Eine solche Erkenntnis kann das menschliche Kreativsein beeinflussen.

**MG:** Das gilt auch für Kreativwerkzeuge im Zusammenspiel mit uns. Es gibt keine strikte Trennung von Mensch und Maschine. Die Maschinen, die wir schaffen, sind niemals unabhängig von uns. Sie können uns aber Dinge zeigen, die wir vorher nicht gesehen haben. So war es beim Go-Spiel, wo man erst später bemerkt hat, dass es sich um ein neues Muster handelt, das auf den ersten Blick nicht „menschlich“ wirkt. Aber wir Menschen lernen davon und bauen diese Strategie in unser Spiel ein – und machen sie damit menschlich. Wenn man weiterdenkt, was Kreativität eigentlich bedeutet, muss man anerkennen, dass kein Mensch isoliert von sich aus ein Genie ist, aus dem Neues entspringt. Wir sind so stark von Eindrücken und vom Lernen getrieben, dass man nicht von einer genialen Kreativität sprechen kann – sondern dass Kreativität auf einem langen Prozess aufbaut. Vielleicht kommen aus der KI neue Dinge heraus, die wir dann ebenso aufnehmen.

**Bei der Installation „Tokens for Climate Care“, die ihr für die London Design Biennale und das MAK Museum für Angewandte Kunst Wien geschaffen habt und die derzeit am AIT Austrian Institute of Technology ausgestellt ist, werden Symbole kreiert, die es vorher noch nicht gegeben hat. Ist das kein kreativer Prozess?**

**MR:** Im Endeffekt ist das Statistik – und kein kreatives Schaffen, das sich die Maschine ausgedacht hat. Der Output beruht auf Berechnungen und auf Mustern, die die Maschine aus den Daten gelernt hat. Das ist im Grunde Mathematik.

**MG:** Wir haben unsere Kreativität, unsere Ansicht von Ästhetik an die Maschine übergeben, um sie zu reproduzieren. Wir haben ein Datenset an Symbolen ausgewählt, und wir haben diese als Gruppe von Kurator:innen gelabelt – wir haben also die verschiedenen Symbole jeweils mit Begriffen verknüpft. Das ist nicht unabhängig vom Menschen, das wurde von uns so eingegeben. Jetzt können andere kommen und sich diese Ideen holen – und vielleicht auf dieser Basis weiterarbeiten.

**In der Fachwelt taucht immer wieder die Frage auf: Wer ist Urheber:in davon, wenn eine KI eine Grafik erstellt, einen Text schreibt oder ein Musikstück komponiert?**

**MR:** Diese Frage stellte sich noch bei jeder neuen Technologie. Sogar bei der Erfindung der Fotografie. Wenn ein Fotoapparat ein Foto macht, kann man trefflich darüber diskutieren, wer Künstler:in ist: Ist es die Maschine? Oder die Person, die abdrückt? Dazu

master, the AI made one move that looked kind of odd from a human viewpoint, and no one understood. All observers thought that the AI had made a mistake. But in retrospect, this was precisely the turning point of the game. Such a realization can have an influence on human creativity.

**MG:** This is also true of creative tools interacting with us. There is no strict division between human and machine. The machines we create are never fully independent of us. But they make us see things that we failed to see before. This is what happened with the game of Go, when people only realized later that it was a new pattern that just seemed non-“human” at first glance. But we humans learn from this and incorporate that strategy into our game—and thus make it human. Thinking somewhat further about what creativity actually means, we must acknowledge that no isolated human being, in and of themselves, is a genius from which new things spring. We are so strongly driven by impressions and learning that we cannot speak of creativity by genius—but of creativity building up on a long process. Perhaps new things will come out of AI that we will then absorb in the same way.

**In the installation “Tokens for Climate Care,” which you created for the London Design Biennale and the MAK—Museum of Applied Art Vienna and which is currently being exhibited at the AIT—Austrian Institute of Technology, symbols are created that have not existed before. Isn’t that a creative process?**

**MR:** The bottom line is that it’s statistics—not creative work devised by the machine. The output is based on computations and patterns that the machine has learned from the data. It’s basically mathematics.

**MG:** We have given our creativity, our view of aesthetics to the machine to reproduce. We selected a data set of symbols and labeled them as a group of curators—so we’ve associated the different symbols each with a term. That is not independent of humans, it’s been input by us that way. Now others can come and get those ideas—and maybe continue to work on that basis.

**In the professional world, the question keeps coming up: Who is the author when an AI creates a graphic design, writes a text, or composes a piece of music?**

**MR:** This question has been raised with every new technology. Even with the invention of photography. When a camera takes a photo, you can have a fine debate about who the artist is: Is it the machine? Or the person pressing the shutter button? There once even was a bizarre copyright discussion about this: a photographer put up his expensive camera in the jungle, then a monkey came along and took a picture of himself. Who is the

gab es einmal eine skurrile Copyright-Diskussion: Ein Fotograf stellte seinen teuren Fotoapparat in den Dschungel, dann kam ein Affe und hat ein Bild von sich selbst gemacht. Wer ist jetzt Urheber:in? Ist es der Fotograf, der die Szene definiert hat? Oder ist es der Affe, der fotografiert hat? In diesem Fall hat man sich darauf geeinigt, dass der Affe Urheber:in ist. Für mich ist es ganz klar, dass nicht das Werkzeug selber die schaffende Einheit ist, sondern die Person dahinter, die sich das Ganze überlegt. Auch in Programmen wie Photoshop gibt es mittlerweile Filter, die auf KI basieren. Da stellt sich auch nicht ständig die Frage, wer Urheber:in dieses Effektes ist.

**Teil der Mystifizierung der KI ist die Tatsache, dass eine KI eine „Black Box“ ist, von der man nicht genau weiß, was in ihr passiert und wie sie zu einer Entscheidung kommt.**

**MR:** Eine KI ist so komplex, dass man nicht jeden Schritt nachvollziehen kann. Diese Systeme werden auch trotz allem Streben nach Nachvollziehbarkeit („explainable AI“) immer ein Stück weit eine Black Box bleiben. Gerade diese Undefinierbarkeit macht sie für künstlerische Zwecke natürlich interessant – ähnlich wie man, um ein ganz einfaches Beispiel zu nennen, auch nicht hundertprozentig versteht, wie eine Farbe trocknet und sich dabei verändert.

**Seit einigen Jahren herrscht auch in der Kunstwelt ein Hype um das Thema KI. Manche meinen, bereits wieder ein Abflauen zu bemerken. Wie nehmt ihr das wahr?**

**MR:** Ich sehe es nicht abflauend. KI ist immer mehr in unterschiedlichsten Aspekten beteiligt, das wird garantiert nicht mehr verschwinden. Dafür ist KI ein viel zu potentes Werkzeug. Aber vielleicht wird es nicht mehr so als Alleinstellungsmerkmal herangezogen – als dezidierte „AI art“. Die Technologie wird in versteckter Weise in Kunstwerken drinnen sein.

**MG:** Solange KI in der Gesellschaft ein Thema ist, wird sie das auch in der Kunst sein. Die Kunstwelt ist da ein Spiegel der Gesellschaft. Wobei der Hype um AI nicht so groß ist wie etwa beim Thema Blockchain, Crypto oder NFTs. Der Riesenhype ist auch ein Riesenproblem: Alles Schlechte des Kunstmarktes wurde mit der Blockchain kodifiziert – mit den Versprechungen, dass es jetzt demokratisch wird und auch die kleinen Künstler:innen Geld machen können. Das ist natürlich nicht eingetreten. Diejenigen, die vorher bereits bekannt waren und ein entsprechendes Marketing haben, haben davon profitiert.

author now? Is it the photographer who defined the scene? Or the monkey who took the picture? In this case, it was agreed that the monkey is the author. To me, it's perfectly clear that the creative entity is not the tool itself but the person behind it who has devised the whole thing. Even in programs like Photoshop, there are now filters that are based on AI. And there is no constant questioning of who the creator of this or that effect is.

**Part of the mystification of AI is the fact that it is a “black box” of which no one really knows what is happening inside and how it makes decisions.**

**MR:** AI is so complex that you cannot possibly understand every single step. These systems will, to some extent, always remain a black box, despite all the efforts to make their workings traceable (“explainable AI”). It is precisely this undefinability that makes them interesting for artistic purposes, of course—just as, to give a very simple example, it is not one hundred percent understood how paint dries and changes in the process.

**For some years now, there has been a hype about AI also in the art world. Some say, they already see it on the wane again. How do you perceive this?**

**MR:** I don't see it waning. AI is more and more involved in all sorts of different aspects, and that is guaranteed not to go away ever. AI is far too potent a tool for that. But maybe it won't be used as a unique selling argument anymore—as explicit “AI art.” The technology will be hidden in some way inside the artworks.

**MG:** As long as AI is an issue in society, it will also be an issue in art. The art world is a mirror of society there. But the hype around AI is not as big as, for example, the one around blockchain, crypto, or NFTs. The huge hype is also a huge problem: Everything bad about the art market was codified with the blockchain—with the promises things will become democratic now, from now on even small artists will be able to make money. That, of course, didn't happen. Those who were already well-known before and had marketing behind them were the ones who benefited.

**MR:** Jede:r musste einen eigenen NFT-Drop machen. Das ist meistens nur schnelle Geldmacherei – da gibt es aktuell eine riesige Goldgräberstimmung. Es ist aber keine Demokratisierung des Kunstmarktes. Die Leute, die ohnehin schon populär sind, können durch diesen Channel noch mehr Geld machen. Wohingegen die unbekannteren Künstler:innen in diesem Überangebot untergehen. Genauso wie im normalen Kunstmarkt.

**MG:** Mir hat allerdings gefallen, dass die Diskussion über NFTs extrem kritisch geführt wurde. Zum Riesenhype gab es auch eine große Gegenbewegung. Gefragt wurde auch: Wie sieht das ökologisch aus? Was sind cleane Blockchains?

#### **Welche Themen in Zusammenhang mit KI interessieren euch zur Zeit am meisten?**

**MR:** Ein großes Thema ist die Datenbeschaffung. Die Frage, wem die Daten gehören und wer wie damit umgehen kann, rückt immer mehr ins Rampenlicht. Wer hat das Recht, welche Daten abzugraben? Was passiert mit den Daten? Daraus ergeben sich interessante ethische Fragen. Die Gesetzeslage hinkt hier komplett hinten nach. Natürlich wäre es toll, wenn man sagen könnte: Alle Daten, die mit dir in Verbindung gebracht werden können, gehören dir. Und alle Firmen, die diese verwerten wollen, müssen dich fragen, müssen dafür zahlen oder sonst etwas.

**MG:** Die Frage ist aber, wie man diesen „shift“ organisiert? Damit das wirklich funktioniert, muss man als Legislative auch in die Technik eingreifen.

#### **Ihr seid derzeit „Artists in Residence“ am AIT Austrian Institute of Technology. Woran arbeitet ihr da?**

**MG:** Das Projekt startet so richtig im Sommer. In Vorgesprächen mit den Forscher:innen hat sich gezeigt, dass die Frage, wie es mit den Daten aussieht, auch im industriellen Kontext, wo es vorwiegend um technische Lösungen und kommerzielle Fragestellungen geht, sehr relevant ist. Die Diskussion wird sehr differenziert durchgeführt, da kann man viel voneinander lernen. Wir werden die ethischen Fragen in einem Kunstwerk auch nicht lösen können. Aber wir werden sie ansprechen und geben im besten Fall eine Richtung vor, in die es gehen könnte. Man muss Stellung beziehen. Und das müssen auch die Techniker:innen machen: Sie bieten eine Lösung an und müssen diese auch vertreten und argumentieren können. ✕

**MR:** Everybody had to have their own NFT drop. That's mostly just about making quick money—there's a huge gold rush going on right now. But there is no democratization of the art market. The people who are popular already can make even more money through this channel. Whereas the lesser-known artists are getting drowned in this oversupply. Just like in the regular art market.

**MG:** I did like the fact, though, that the discussion about NFTs was extremely critical. The big hype also had a big counter-movement. People were asking: What does this look like, ecologically? What are clean blockchains?

#### **What AI-related topics are you most interested in at the moment?**

**MR:** One big issue is data acquisition. The question of who owns the data, and who may handle it, and how, is increasingly coming into the spotlight. Who has the right to mine what data? What happens to that data? This raises interesting ethical questions. Legislation is hopelessly behind the times here. Sure, it would be great being able to say: all data that can be associated with you belongs to you. And any company that wants to use it has to ask you first, pay for it, or something like that.

**MG:** But the question is how to organize such a “shift”? For that to really work, legislature would also have to interfere directly in the technology.

#### **You are currently “Artists in Residence” at the AIT—Austrian Institute of Technology. What are you working on there?**

**MG:** The project is really starting in the summer. Preliminary talks with researchers have shown that the question of what's going on with the data is also highly relevant in an industrial context, where it is mainly about technical solutions and commercial issues. The discussion is being held with much differentiation, and there is a lot to learn from one another. We won't be able to solve the ethical questions in a work of art. But we will address them and, in the best case, indicate a direction in which things could go. You have to take a stance. And that is what technicians must do, too: offer a solution and be able to stand up for and argue it. ✕

## Angewandte Interdisciplinary Lab (AIL)

**Das Angewandte Interdisciplinary Lab (AIL) ist ein experimenteller Raum, in dem interdisziplinäre Praxis und Gedankenexperimente möglich werden, um die Gegenwart und die Zukunft gemeinsam zu gestalten. Das AIL ist eine Quelle der Inspiration, ein Raum zum Austausch und eine Plattform für Projekte an den Schnittstellen von Kunst, Wissenschaft und künstlerischer Forschung. Gegründet 2014 von der Universität für angewandte Kunst Wien auf Initiative von Gerald Bast, war es von Anfang an das Anliegen, den Austausch zwischen den unterschiedlichen Disziplinen zu ermöglichen und Kunst wie künstlerische Forschung zu öffnen.**

Im Frühjahr 2021 ist das AIL in die ehemalige Otto Wagner Postsparkasse im Herzen Wiens umgezogen und befindet sich jetzt in direkter Nachbarschaft zu den anderen Instituten der Angewandten, in einem Stadtviertel voller neuer Energie, in dem sich mehrere Forschungsinstitute aus Kunst und Wissenschaft angesiedelt haben. Der neue Standort bietet die Chance, sich weiter zu vergrößern und die Netzwerke für interdisziplinäre Arbeit und Forschung zu stärken; dazu gibt es drei Räume im Mezzanin, mit dem Café Exchange (der ehemaligen Kassenhalle) als Herzstück.

Mit dem Café Exchange erweitert das AIL nicht nur sein Programm, sondern setzt die Konsolidierung an der Schnittstelle zwischen Universität und Öffentlichkeit weiter fort. Im Sinne des Wiener Salons ist dies ein Ort, um Gedanken, Wissen, Visionen und Utopien auszutauschen – um sich zu treffen und zu vernetzen, exquisite Kreationen des meisterhaften Kaffeebereiters Alexander Afrough zu genießen und dabei Kunst- oder Sound-Performances oder Installationen zu erleben, oder auch um Diskussionen und Vorträgen aus dem Programm zu lauschen. Das Programm, kuratiert vom Leitungsteam und dem Executive Board, basiert auf Forschungsprojekten mit der Angewandten bzw. auf der Zusammenarbeit mit externen Partner:innen. Seit seiner Gründung hat das AIL eine große Bandbreite an Formaten möglich gemacht, darunter multidisziplinäre Ausstellungen, Gespräche und Diskussionen, Symposien, Vorträge, Konzerte und Performances ebenso wie öffentliche Experimente und informelle Treffen.

Executive Board: Dr. Gerald Bast, Dr. Alexander Damianisch,  
Mag.<sup>a</sup> Elisabeth Falkensteiner, Mag.<sup>a</sup> Alexandra Graupner und  
Mag. Jürgen Gschiel

Mehr Informationen @ [ail.angewandte.at](mailto:ail.angewandte.at)

## Angewandte Interdisciplinary Lab (AIL)

**Angewandte Interdisciplinary Lab (AIL) is an experimental space in which interdisciplinary practices and thought experiments are made possible in order to shape the present and future together. AIL is a source of inspiration, a place for exchange and a platform for projects at the intersection of art, science and artistic research. Founded in 2014 by the University of Applied Arts Vienna as an initiative by Gerald Bast, it was launched to enable exchange among different disciplines and to open up art and artistic research.**

In spring 2021 AIL moved to the former Otto Wagner Postsparkasse in the heart of Vienna thus joining other departments of Angewandte and a newly emerging neighborhood comprising several research institutions from the field of art and science. The new location provides the opportunity to further expand and strengthen networks for interdisciplinary work and research, divided into three rooms on the mezzanine floor, with Café Exchange (former Kassenhalle) as its centerpiece.

With Café Exchange, AIL opens up not only the program but continues to consolidate the intersection between the University and the public. Following the spirit of Wiener Salon, this is a place to exchange thoughts, knowledge, visions and utopias – to meet and network, linger over great drinks from master coffeemaker Alexander Afrough while experiencing artistic or sound performances, installations or listening to talks and lectures from the program. The program, curated by the heads of AIL and the Executive Board, is based on research projects affiliated with Angewandte or on collaborations with external partners. Ever since it was founded, AIL has facilitated a wide range of formats, such as multidisciplinary exhibitions, talks and discussions, symposia, lectures, concerts and performances as well as public experiments and informal meetings.

Executive Board: Dr. Gerald Bast, Dr. Alexander Damianisch,  
Mag.<sup>a</sup> Elisabeth Falkensteiner, Mag.<sup>a</sup> Alexandra Graupner and  
Mag. Jürgen Gschiel

More information @ [ail.angewandte.at](mailto:ail.angewandte.at)





Research Präsentation  
Salon de Passage #3 im Café  
Exchange im Mai 2022.  
Foto: Lea Dörl

Research Presentation Salon  
de Passage #3 at Café Ex-  
change in May 2022.  
Photo: Lea Dörl



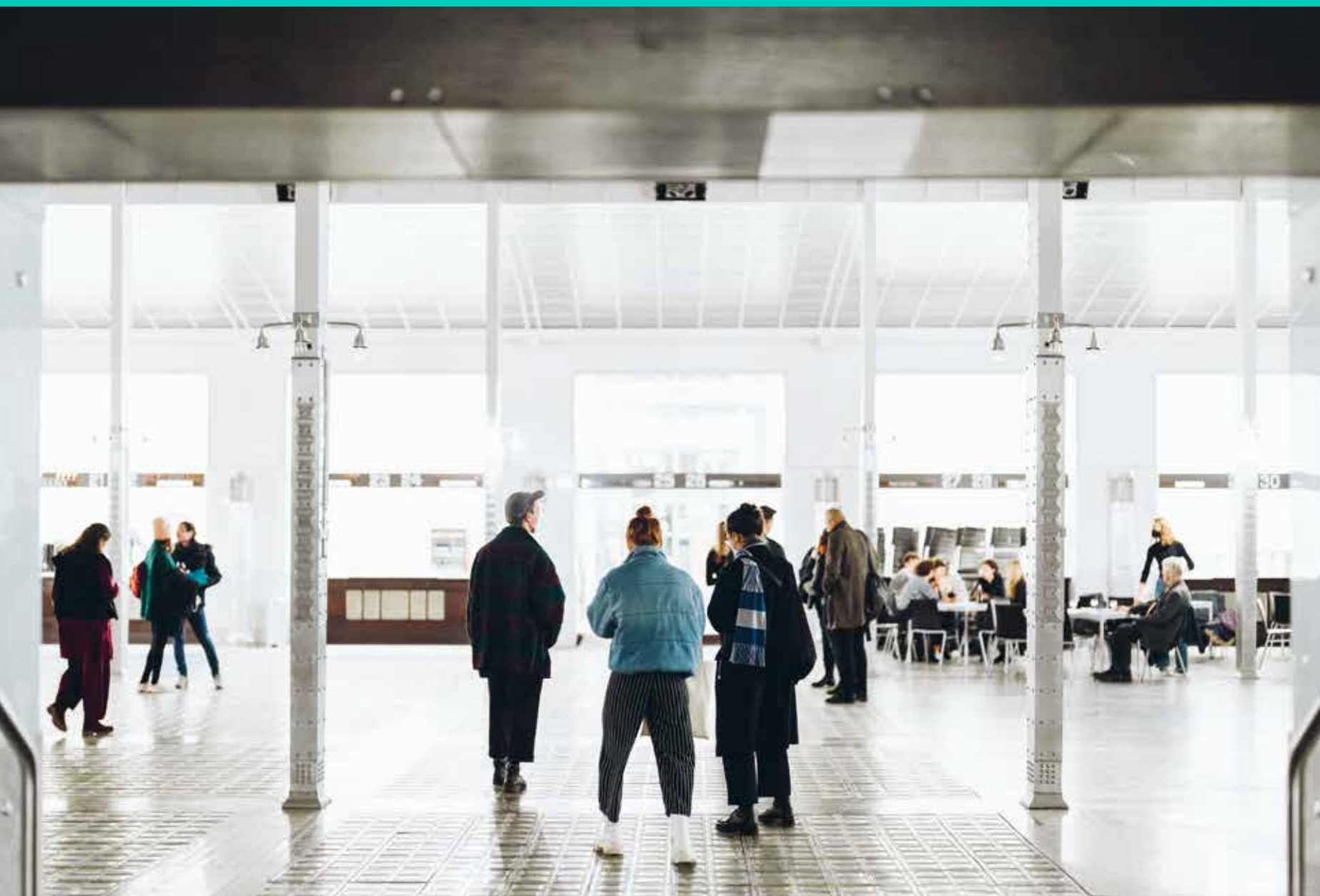


Café Exchange: Inspiriert vom Wiener Salon werden Ideen, Wissen und Utopien verhandelt, Diskurs- und Kaffeehauskultur unter einem Dach vereint. Kulinarische Gestaltung von Alexander Afrough. Foto: Lea Dörl

Café Exchange: Inspired by the Viennese salon, ideas, knowledge and utopian concepts are being discussed; discourse and coffee house culture are brought together under one roof. All things culinary by Alexander Afrough. Photo: Lea Dörl

Diskussionsreihe im Café Exchange: Gerald Bast, Ina Conradi und Mark Chavez über die Sehnsucht der Menschen nach Illusion, das Verhältnis von Realität und Virtualität und das Potenzial von auf Künstlicher Intelligenz basierenden, imaginären Welten, Mai 2022. Foto: Martina Lajczak

Discussion series at Café Exchange: Gerald Bast, Ina Conradi, and Mark Chavez on people's yearning for illusion, the relationship between reality and virtuality, and the potential of Artificial Intelligence based imaginary worlds, May 2022. Photo: Martina Lajczak



Das Angewandte Interdisciplinary Lab mit dem Café Exchange befindet sich in der ehemaligen Kassenhalle sowie zwei daran anschließenden Ausstellungsräumen. Beide bilden somit das Herz der neu belebten Postsparkasse im legendären Otto Wagner Bau. Foto: Lea Dörl

Angewandte Interdisciplinary Lab with Café Exchange is located in the former counter hall and two adjoining exhibition rooms. Both thus form the heart of the newly revitalized Postsparkasse in the legendary Otto Wagner Building. Photo: Lea Dörl









Live Performance von Gischt, Juni 2022. In verschiedenen Formaten wird Sound in seiner materiellen und relationalen Beschaffenheit sowie in seiner sozialen und politischen Dimension untersucht.  
Foto: Paul Pibernig

Live performance of Gischt, June 2022. Sound is examined in various formats in terms of its material and relational nature as well as its social and political dimensions.  
Photo: Paul Pibernig





Ausstellungsansicht The Octopus, abschließende Ausstellung des gleichnamigen Bildungsprojekts, welches künstlerische Forschung und produktionsbasierte Zusammenarbeit zwischen Akademien und Kunstinstitutionen, Studierenden und Fachleuten, verschiedenen Präsentationsformen und Forschungs- und Dokumentationsprozessen in unterschiedlichen geografischen Regionen fördert. Kuratiert von Başak Şenova, Juni 2022. Foto: Paul Pibernig

Exhibition view of The Octopus, final exhibition of the educational project of the same name which promotes artistic research and production-based collaboration between academies and art institutions, students and professionals, different forms of presentation and processes of research and documentation in different geographical regions. Curated by Başak Şenova, June 2022. Photo: Paul Pibernig

# Alpbacher Technologiegespräche

## The New Europe

25. – 27. August 2022

**Die Alpbacher Technologiegespräche werden vom AIT Austrian Institute of Technology, Österreichs größter Research and Technology Organisation, und ORF Radio Österreich 1 programmiert. Das Projektteam besteht aus Dr. Martin Bernhofer (ORF Ö1), Mag. Michael H. Hlava und Claudia Klement (beide AIT). Dem Steering Committee der Alpbacher Technologiegespräche 2022 gehören Dr. Hannes Androsch, Prof. Dr. Wolfgang Knoll (wissenschaftlicher Geschäftsführer des AIT) und Ingrid Thurnher (Radiodirektorin ORF) an.**

Wissenschaftlicher Partner der Alpbacher Technologiegespräche 2022 ist die Helmholtz-Gemeinschaft Deutscher Forschungszentren. Strategische Partner sind das österreichische Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (BMK), das Bundesministerium für Bildung, Wissenschaft und Forschung (BMBWF) und die Industriellenvereinigung (IV).

Corona-Krise, Klimawandel, Biodiversitätsschwund, Energiekrise, der Krieg in der Ukraine und andere geopolitische Spannungen beeinträchtigen das Leben von immer mehr Menschen immer stärker. Die multiple Krise hat massive ökonomische und gesellschaftliche Auswirkungen: Diese reichen von einer Verschiebung der Terms of Trade und damit des Wohlstands in verschiedenen Regionen bis hin zu sozialen Spannungen, die eine Gefahr für die Demokratie sein können. Ein wichtiger Schlüssel zur Überwindung dieser Krisenerscheinungen bieten Technologien, die die nötige Transformation ermöglichen bzw. erleichtern – etwa beim Umgang mit dem Klimawandel, bei der Energie- und der Mobilitätswende.

Unter dem Jahresthema des Europäischen Forums Alpbach 2022 „The New Europe“ werden insbesondere Bereiche wie Energiesicherheit, Info- bzw. Cyberwar, globale Lieferketten oder innovative RNA-basierte Medikamente thematisiert. Eine zentrale Technologie dabei ist Künstliche Intelligenz, die in den vergangenen Jahren in vielen Anwendungsbereichen große Fortschritte ermöglichte – bis hin zu den Künsten, denen ebenfalls breiter Raum gegeben wird.

# Alpbach Technology Symposium

## The New Europe

25<sup>th</sup> – 27<sup>th</sup> August 2022

**The Alpbach Technology Symposium is programmed by the AIT Austrian Institute of Technology, Austria's largest research and technology organization, and ORF Radio Österreich 1. The project team consists of Dr. Martin Bernhofer (ORF Ö1), Mag. Michael H. Hlava and Claudia Klement (all AIT). The Alpbach Technology Symposium's Steering Committee includes Dr. Hannes Androsch, Prof. Dr. Wolfgang Knoll (AIT Scientific Director) and Ingrid Thurnher (ORF Radio Director).**

Scientific partner of the Alpbach Technology Symposium 2022 is the Helmholtz Association of German Research Centers. Strategic partners are the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK), the Federal Ministry of Education, Science and Research (BMBWF) and the Federation of Austrian Industries (IV).

Corona crisis, climate change, loss of biodiversity, energy crisis, the war in Ukraine and other geopolitical tensions are increasingly affecting the lives of more and more people. The multiple crises have massive economic and social impacts: ranging from a shift in the terms of trade and thus prosperity in different regions to social tensions that can constitute a threat to democracy. An important key to overcoming these phenomena of crisis is offered by technologies that enable or facilitate the necessary transformation – for example, in dealing with climate change, the energy transition, and the mobility transition.

Under the umbrella of the annual theme of the European Forum Alpbach 2022 “The New Europe”, areas such as energy security, information and cyber warfare, global supply chains, and innovative RNA-based medicines will be addressed in particular. A central technology here is Artificial Intelligence, which has enabled great progress in many areas of application in recent years – right up to the arts, which will also be given broad scope.

Wir bedanken uns herzlich bei all unseren Partnern, die auch heuer die Alpbacher Technologiegespräche möglich gemacht haben:

We would like to thank all of our partners who have made the Alpbach Technology Symposium possible this year:



Bundesministerium Bildung, Wissenschaft und Forschung

Bundesministerium Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie



HELMHOLTZ SPITZENFORSCHUNG FÜR GROSSE HERAUSFORDERUNGEN



SAL SILICON AUSTRIA LABS

FORSCHUNG AUSTRIA Zukunft verbindet.



eit Manufacturing Co-funded by the European Union

di:angewandte Universität für angewandte Kunst Wien University of Applied Arts Vienna

TU AUSTRIA Austrian Universities of Technology

BI EC ONOMY AUSTRIA

INNO VATION SALZ BURG featuring the future

EFA European Forum Alpbach

AT&S

APA

umweltbundesamt PERSPEKTIVEN FÜR UMWELT & GESELLSCHAFT

#### Organisator Organizers

AIT Austrian Institute of Technology GmbH  
Austrian Broadcasting Corporation – Programme Radio 1

#### Strategische Partner Strategic partners

Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie (BMK)  
Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology  
Bundesministerium für Bildung, Wissenschaft und Forschung (BMBWF)  
Austrian Federal Ministry of Education, Science and Research  
Industriellenvereinigung  
Federation of Austrian Industries

#### Netzwerk- und wissenschaftlicher Partner

Network & Scientific Partner  
Helmholtz-Gemeinschaft Deutscher Forschungszentren

#### Mitwirkende Partner Contributing partners

Climate and Energy Fund  
EIT Manufacturing East GmbH  
Forschung Austria  
Innovation Salzburg GmbH  
Bioeconomy Austria  
Joanneum Research Forschungsgesellschaft mbH  
TU Austria  
VFFI/Federation of Austrian Industries  
AT&S

#### Unterstützende Partner Supporting partners

Silicon Austria Labs GmbH  
Bundesministerium für Arbeit und Wirtschaft  
Umweltbundesamt

#### Medienpartner Media partners

APA Austrian Press Agency

Wir danken folgenden Partnern für die wertvolle Unterstützung der Produktion des heurigen Jahrbuchs „Technologie im Gespräch“

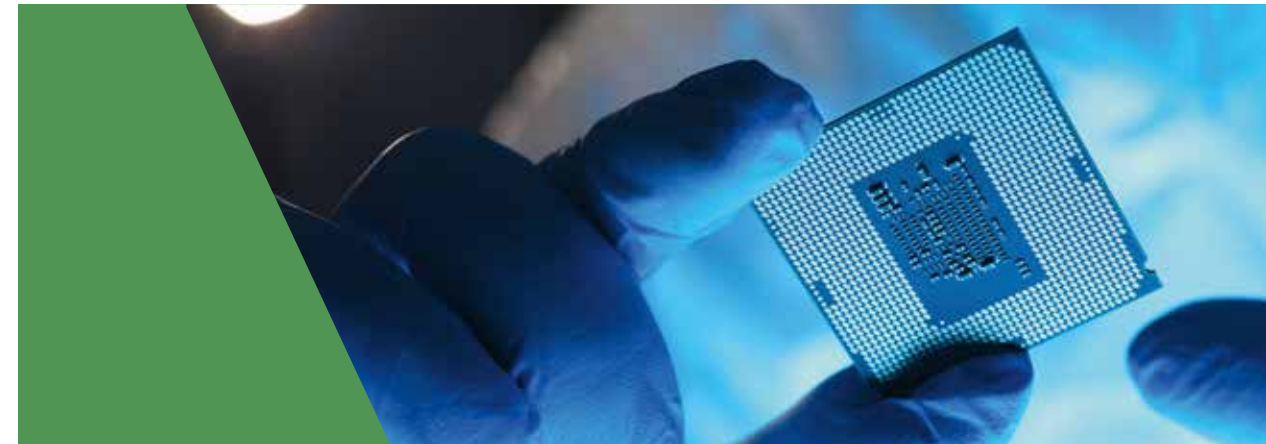
Thank you to the following partners for their valuable support in the production of this year's yearbook "Discussing Technology"

# WE CONNECT THE TECHNOLOGY COMMUNITY AT THE EUROPEAN FORUM ALPBACH

25 – 27 AUGUST 2022 TEC  
#EFATEC



[www.ait.ac.at/efatec](https://www.ait.ac.at/efatec)



## Mikroelektronik als Motor für den Klimaschutz

Das Europäische Chip-Gesetz schafft einen  
idealen Rahmen für ambitionierte FTI-Projekte.

### Digitaler und ökologischer Wandel

Erneuerbare Energie, umweltfreundliche  
Mobilität, saubere Produktion und kli-  
maneutrale Städte: All das würde ohne  
europäische Halbleitertechnologien in  
kürzester Zeit stillstehen. Darum plant  
die Europäische Union ein neues Chip-  
Gesetz, mit dem rund 43 Millionen Euro  
in Form von öffentlichen und privaten  
Investitionen mobilisiert werden sollen.  
Entsprechende Verhandlungen laufen  
bereits.

### Chancen schon jetzt nutzen

Das Maßnahmenpaket wird zusätzlichen  
Schwung für innovative Projekte aus  
Forschung, Technologie und Innovation  
(FTI) bringen. Schon heute sorgen at-  
traktive Programme und Initiativen aus  
dem Klimaschutzministerium für einen  
positiven Trend – eine große Chance für  
FTI-Vorreiter aus Österreich.

Jetzt anmelden unter:

[bmk.gv.at/innovation](https://bmk.gv.at/innovation)



# Erste Wahl für den Start in dein Berufsleben.

Werde Teil des Teams von AT&S  
im neuen R&D Center in Leoben.

MODERN. INNOVATIV. DIGITAL.

Die Industrie gestaltet die digitale Zukunft  
schon heute – für den Erfolg von morgen.

Die Industriellenvereinigung

www.iv.at



FOLLOW AT&S  
ON SOCIAL MEDIA





# WIR MACHEN ABFALL ZUM ROHSTOFF.

# INNOVATIVES RECYCLING VON DER NEUEN OMV.

Nachhaltig handeln bedeutet, mit Ressourcen so umzugehen, dass wir auch morgen noch gut leben können. Und das müssen wir heute anpacken. Die neue OMV forscht schon jetzt an den Lösungen für morgen und investiert in innovative Recyclingprojekte. Denn wir wollen dazu beitragen, den Großteil der Kunststoffabfälle in Österreich als wertvolle Rohstoffe wiederzuverwerten und so CO<sub>2</sub> einzusparen.

Mehr dazu: [omv.com/neue-omv](http://omv.com/neue-omv)



Wirtschaft, Forschung  
& Bildung  
an einem Ort  
Technopole

## Niederösterreich öffnet Türen ...

... für technologieorientierte Unternehmen und Forschungseinrichtungen, die gemeinsam an einem Ort ihr Wissen bündeln. An unseren Technopol-Standorten sorgen wir dafür, dass Niederösterreich sich als innovativer und erfolgreicher High-Tech-Standort positioniert.

ecoplus. Niederösterreichs Wirtschaftsagentur  
Tel.: 02742 9000-19600  
E-Mail: [technopol.programm@ecoplus.at](mailto:technopol.programm@ecoplus.at)

plus  
eco

Die Wirtschaftsagentur  
des Landes Niederösterreich

[ecoplus.at](http://ecoplus.at)



Europäische Union Investitionen in Wachstum & Beschäftigung. Österreich.





# research partner for smart industries

RESEARCH PARTNER FOR ELECTRONIC BASED SYSTEMS – WE SUPPORT YOUR SUCCESS STORY!

Silicon Austria Labs (SAL) conducts application-oriented research in the field of electronic based systems (EBS) and develops innovative solutions for production, mobility, health, energy and lifestyle - from wafers to intelligent systems.

SAL brings together key players from industry, science and research and thus valuable expertise and know-how and conducts cooperative research along the value chain of EBS. Cooperative projects are co-financed by SAL and enable a fast and unbureaucratic project start.

Learn more about our research



Further information on our website

unfold the future.  
[silicon-austria-labs.com](http://silicon-austria-labs.com)

Silicon Austria Labs GmbH  
GRAZ | LINZ | VILLACH



Photo by Lea Dörl



Angewandte  
Interdisciplinary Lab  
**platform  
for art,  
science  
and  
artistic  
research**

Georg-Coch-Platz 2  
1010 Wien  
[ail.angewandte.at](http://ail.angewandte.at)

di:'angewandte  
Universität für angewandte Kunst Wien  
University of Applied Arts Vienna

@ail\_vienna / @cafe\_exchange



# WECKE DEINE NEUGIER

Sie sind einen Schritt voraus.  
Sie gewinnen neues Wissen.  
Sie arbeiten an Lösungen für die Zukunft.

Lerne exzellente Forschende aus ganz Österreich kennen.

scilog



[bmaw.gv.at](http://bmaw.gv.at)


## Forschung und Innovation für Wachstum und Arbeitsplätze



Hier eintauchen in die Welt der Spitzenforschung:  
[scilog.fwf.ac.at](http://scilog.fwf.ac.at)

**FWF**

Der Wissenschaftsfonds.

 Bundesministerium  
Arbeit und Wirtschaft

In Forschung investieren zahlt sich aus. 1 Euro an F&E-Ausgaben löst einen langfristigen Zuwachs des Bruttoinlandsprodukts von bis zu 6 Euro aus. F&E-intensive Unternehmen schaffen mehr Arbeitsplätze, sie sind krisenfester und erfolgreicher. Forschung und Innovation stärken den Standort Österreich. Das Wirtschaftsministerium investiert jährlich mehr als 150 Mio. Euro in standort-relevante Forschung. Nähere Informationen finden Sie unter: [bmaw.gv.at](http://bmaw.gv.at)

# Technologie im Gespräch

## Discussing Technology

### Jahrbücher zu den Alpbacher Technologiegesprächen

#### Alpbach Technology Symposium Yearbooks

Bisher erschienen/*Published to date*  
Download: <https://www.ait.ac.at/efatec>



### 2017 Digitalisierung

#### Digitisation

Wir befinden uns in einer Umbruchszeit zwischen industrieller und digitaler Revolution. Die neuen Herausforderungen, die von Algorithmen, Big Data, Künstlicher Intelligenz, maschinellem Lernen und Robotern aufgeworfen werden, lassen sich nicht mit Rückgriffen in die Vergangenheit lösen. Gefragt sind neue Ideen, neue Zugänge und Innovationen – sowohl in technologischer als auch in sozialer Hinsicht.



### 2018 Künstliche Intelligenz

#### Artificial Intelligence

Rasant und unumkehrbar verändern digitale Technologien die Welt, unsere Gesellschaft, unsere Wirtschaft und unser Leben. Vor allem Künstliche Intelligenz könnte von größerer Bedeutung sein als die Zähmung des Feuers oder die Elektrizität, meinen viele Expert:innen. Aufhalten lässt sich die

Entwicklung nicht, wir müssen uns ihr stellen und die Rahmenbedingungen proaktiv und zukunftsweisend gestalten.



### 2019 Sicherheit im Cyberraum

#### Cybersecurity

Je mehr Arbeit uns digitale Technologien abnehmen, umso abhängiger werden wir von ihnen – und umso schlimmer wird es, wenn sie ausfallen, manipuliert oder angegriffen werden. Das macht uns extrem verwundbar. Die Umbrüche und neuen Risiken erzeugen Unsicherheit, Sorgen und Ängste. Das zunehmende Unsicherheitsgefühl ist eine Quelle, aus der politisches Kapital geschlagen werden kann. Die Publikation widmet sich zahlreichen Aspekten des Themas Cybersecurity sowie des Verhältnisses zwischen Sicherheit und Freiheit.



### 2020 Komplexe Systeme

#### Complexity

Viele Menschen nehmen die Welt als zunehmend komplex und unübersichtlich wahr. Und das zu Recht. Nicht zuletzt die Corona-Pandemie zeigt die enge Verflechtung verschiedenster Bereiche, aus der auch eine hohe Anfälligkeit gegenüber Störungen erwächst. Hand in Hand damit wachsen die großen Herausforderungen für die Menschheit – etwa die demografische Entwicklung, der Klimawandel und die Digitalisierung. Alle diese Problemfelder sind systemischer Natur und lassen sich mit herkömmlichen Mitteln nicht mehr bewältigen. Dieses Jahrbuch widmet sich neuen Sichtweisen und Zugängen, um komplexe Systeme besser verstehen und managen zu können.



### 2021 Human Centered Innovation

Der Grundgedanke von „Human Centered Innovation“ ist, dass der Mensch, seine Bedürfnisse und Werte bei jeglicher Entwicklung von Anfang an mitberücksichtigt werden. Nur so kann erreicht werden, dass die Technik dem Menschen dient – und nicht umgekehrt – und dass Technologien in keinen unüberwindbaren Konflikt zu menschlichen Werten geraten. Ein zentraler Punkt dabei ist die Gestaltung von Mensch-Maschine-Schnittstellen. Aber auch in vielen anderen Bereichen wird man sich zunehmend bewusst, dass die systematische Berücksichtigung des Faktors Mensch wesentlich ist.



## TEC

Research, Technology, Innovation  
at European Forum Alpbach



In collaboration with  
ARTTEC – The art programme of AIT  
and AIT International Alumni Community

### Idee und Konzept **Idea and concept**

Hannes Androsch, Michael H. Hlava, Martin Kugler  
Alpbacher Technologiegespräche

Sarah Hellwagner, Clemens Kopetzky  
art:phalanx, Kultur und Urbanität

### Medieninhaber **Media owners**

AIT Austrian Institute of Technology GmbH und  
art:phalanx Kommunikationsagentur GmbH

### Herausgeber **Publishers**

Hannes Androsch, Wolfgang Knoll, Anton Plimon

### Projektmanagement **Project management**

Annamarie Neher  
art:phalanx, Kultur und Urbanität

### Redaktion **Editor**

Martin Kugler

### Lektorat **Copy-editing**

Connect Translations Austria GmbH

### Übersetzung **Translations**

Connect Translations Austria GmbH  
und Martin Strand

### Grafische Gestaltung

#### **Visual design**

Alexandra Warlits  
ap media – Visuelle Kommunikation

### Lithografie **Lithography**

Pixelstorm, Wien

### Schriften **Fonts**

Vista Sans, Vista Slab (Xavier Dupré)

### Papier **Paper**

Materica Gesso, Pergraphica High White Smooth

### Druck **Printed by**

Gugler GmbH

### Verlag **Publishing house**

Verlag Holzhausen GmbH

### Bildnachweis **Photo credits**

S./p. 4 © AIC, Peter M. Mayr; S./p. 38 © Heimo Aga;  
S./p. 54 © AIT; S./p. 112 © AIT; S./p. 144 © Wojciech  
Czaja; S./p. 152 © Thomas Wunderlich

Alle Rechte, auch die des auszugsweisen Abdruckes  
oder der Reproduktion einer Abbildung, sind  
vorbehalten. Das Werk einschließlich aller seiner  
Teile ist urheberrechtlich geschützt. Jede Verwer-  
tung ohne Zustimmung des Verlags und des  
Herausgebers ist unzulässig.

All rights are reserved, including the rights to  
copy extracts or reproduce illustrations. Any and  
all parts of this work are protected by copyright.  
No part of this publication may be reproduced,  
translated, microfilmed or stored in a retrieval  
system without the prior permission of the  
publishing house and the publisher.



1. Auflage 2022  
First Edition 2022  
ISBN 978-3-903207-66-0

Printed in Austria, EU  
Alle Rechte vorbehalten  
All rights reserved

### Bibliografische Informationen der Öster- reichischen Nationalbibliothek und der Deutschen Nationalbibliothek:

Die ÖNB und die DNB verzeichnen diese Publikation  
in den Nationalbibliografien; detaillierte biblio-  
grafische Daten sind im Internet abrufbar. Für die  
Österreichische Bibliothek: <http://onb.ac.at>, für  
die Deutsche Bibliothek: <http://dnb.ddb.de>

### **Bibliographic information published by the Österreichische Nationalbibliothek and the Deutsche Nationalbibliothek:**

The ÖNB and the DNB are listing these  
publications in the Nationalbibliografien;  
detailed bibliographic data are available on  
the Internet. For the Österreichische National-  
bibliothek: <http://onb.ac.at> For the Deutsche  
Nationalbibliothek: <http://dnb.ddb.de>

# Geschichte der künstlichen Intelligenz

## The History of Artificial Intelligence



### **Wo und wie KI sinnvoll eingesetzt wird**

Künstliche Intelligenz ist längst kein reines Zukunftsthema mehr. Vielmehr durchdringen Methoden der KI immer weitere Bereiche unseres Lebens und Arbeitens. Und sie ermöglichen auch in Wissenschaft, Forschung und Technologieentwicklung große Fortschritte. Anhand zahlreicher Beispiele von konkreten Anwendungen, unter anderem aus dem AIT Austrian Institute of Technology und weiteren heimischen Forschungsstätten, wird gezeigt, wo wir bei der Anwendung von KI in verschiedenen Domänen derzeit stehen. Das reicht von nützlichen Tools für unser Alltagsleben über die Konzipierung künftiger autonomer Maschinen, die eng mit dem Menschen zusammenarbeiten, bis hin zum Einsatz im Gesundheitswesen und der Stadtplanung.

Dabei wird deutlich, wie die unterschiedlichsten KI-Methoden sinnvoll eingesetzt werden können – und in welchen Bereichen man besser darauf verzichten sollte. Ergänzt werden diese Beispiele durch Interviews mit führenden Forscher:innen-Persönlichkeiten wie Helga Nowotny, Andreas Kugi oder Ross King. Zur Sprache kommen überdies die ökologischen Auswirkungen von KI sowie die Rolle der neuen Technologie in den Künsten.

### **Where and how using AI makes sense**

Artificial Intelligence is no longer just a topic for the future. Rather, AI methods are permeating more and more areas of our life and work. They also enable great advances in science, research, and technology development. Numerous examples of concrete applications, including from the AIT Austrian Institute of Technology and other Austrian research facilities, will show the current state of AI applications in various domains. These range from useful tools for our everyday lives via designing autonomous machines that will closely collaborate with humans, to AI in healthcare and urban planning.

It will become clear how these various AI methods can be used sensibly – and in which areas it is better to do without them. These examples are complemented by interviews with leading researchers such as Helga Nowotny, Andreas Kugi, and Ross King. The ecological effects of AI and the role of this new technology in the arts will also be discussed.



**HOLZHAUSEN**  
— *Der Verlag* —