BLACK BOX REPORTING

HOW JOURNALISM CAN REPORT ON AI AND ALGORITHMS

CONTENTS

INTRO

With tools such as ChatGPT or Midjourney, artificial intelligence (AI) has captured society's attention and is hitting international headlines. The data team and the AI + Automation Lab of German Public Broadcaster Bayerischer Rundfunk (BR) share their proven approaches that enable journalists to investigate and scrutinize AI systems and algorithms.

In November 2022, OpenAI made the text chatbot ChatGPT available to the broader public. Since then, anyone can create artificially generated texts with an easy-to-use text interface. Fears and euphoria are the two poles of the discussion. Companies advertise AI software as a salvation, while some journalists fuel fears about the destructive power of artificial intelligence by visualizing AI as robots or attributing human abilities such as "feeling" or "thinking" to it.

From the Nerd Community to the Center of Society

Coverage of algorithms and AI is often found in tech journalism departments and only a few media companies propel investigations on algorithms, as the German media magazine journalist reports. With increased attention, reporting on AI is becoming a cross-sectional topic that, like reporting on climate change, affects many different editorial departments. AI and algorithms are playing an increasingly important role in business, medicine, agriculture, and many other areas.

In this context, investigative and explanatory approaches are gaining relevance to counter exaggerated hopes and fears. The American journalism researcher Nicholas Diakopoulos has coined the term "Algorithmic Accountability Reporting". It refers to researching and reporting on algorithms. With this kind of reporting, human deci-

sions behind algorithmic systems are revealed so that actors such as companies or governments can be held accountable for these very decisions.

Software Discriminating Against People with Black Skin Color

The first major algorithmic accountability investigation to attract international attention was published in 2016 by the non-profit newsroom Pro Publica under the title "Machine Bias". The investigation shows that a software program used as decision-helping tool for judges discriminates against people with black skin color. The algorithm incorrectly assessed them as potential recidivists much more often than people with white skin color.

The team led by journalists Julia Angwin and Jeff Larson also achieved a methodological breakthrough with their research by using freedom of information requests and statistical methods to question the output of an algorithm that supports automated decision-making.

Source: probublica.org
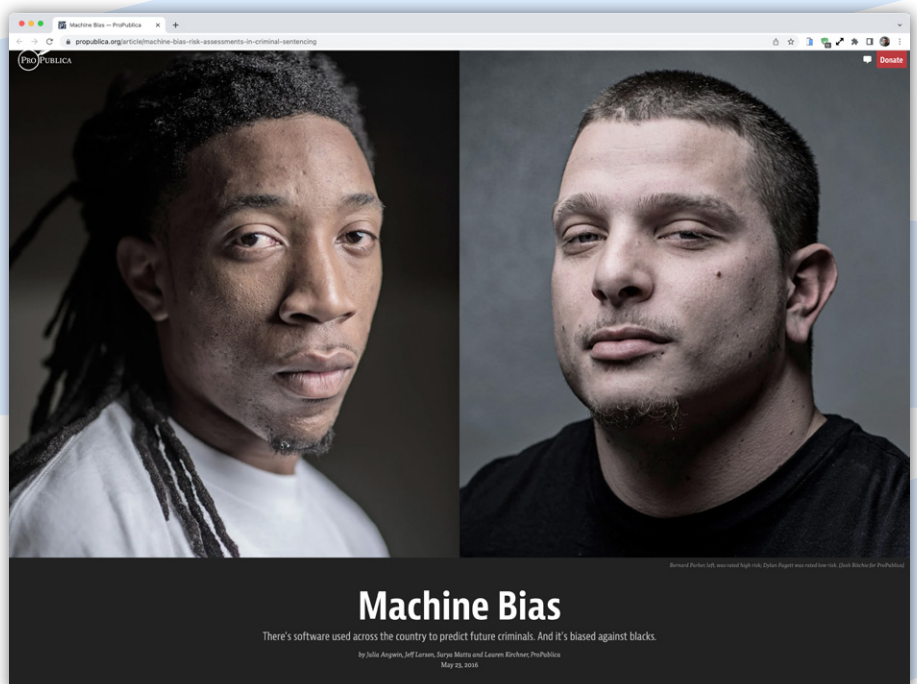
Journalism with Coding and Data Analysis

Algorithmic Accountability Reporting is still a young genre. Reporting on algorithms requires a deep technical understanding, knowledge of statistics, journalistic craft, and experience in choosing the right research strategy for the system under investigation.

Bayerischer Rundfunk invested in editorial tech expertise early on and founded the data journalism team BR Data in 2015. The team works closely with the investigative journalists of BR Recherche. In 2020, the AI + Automation Lab added expertise in the areas of machine learning, Natural Language Generation, and product development. The members of all three teams work interdisciplinarily at the intersection of programming and journalism on investigative data and AI stories. The teams specialize in technical experiments and data analyses that provide insights into the functioning of AI systems and algorithms.

With such approaches, BR teams have published groundbreaking investgations that have been internationally noticed. These include publications on credit scoring algorithms (2018 together with "Der Spiegel"), training data collections (2023) or AI systems that can cause societal harm, such as an AI recruiting system using AI to assign a personality profile to job candidates (2021).

The AI + Automation Lab also employs their methodological knowledge to make AI and automation usable for journalism, like for developing automated texts, graphics or news briefings. Knowledge of AI text generation, training data, and statistical models can thus be used both for investigations and the development of new formats.
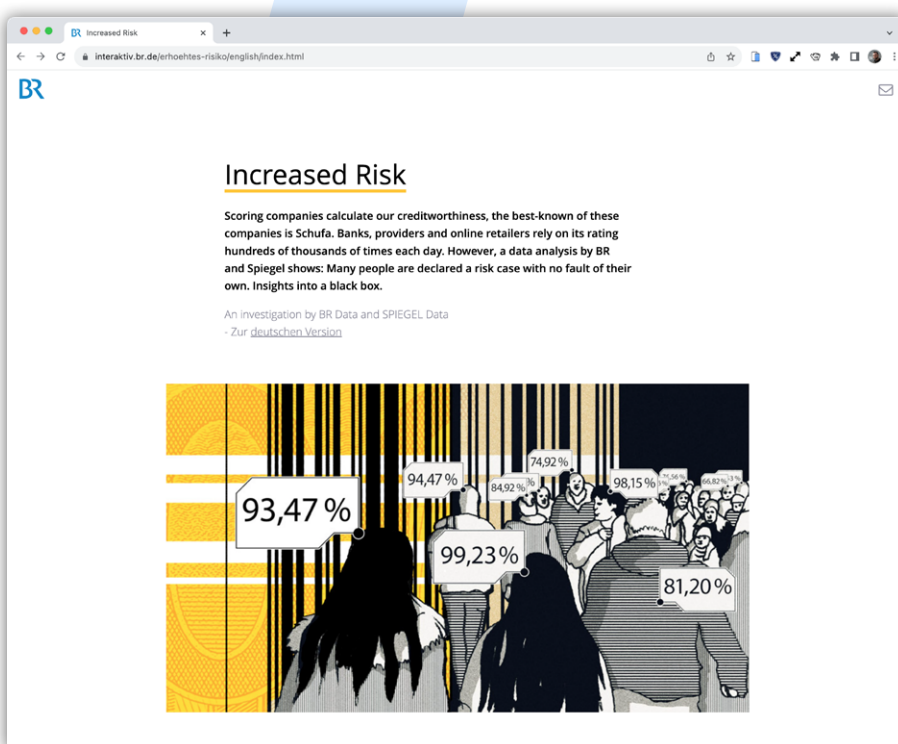
REPORTING ON DIFFERENT
BLACK BOX SYSTEMS

Algorithmic accountability investigations regularly demonstrate how AI systems and other automated decision-making systems can make erroneous, discriminatory, or morally reprehensible decisions and thus influence people's lives in concrete and often problematic ways. For this, well-thought-out, strategically chosen investigative approaches are especially important when journalists are confronted with different types of opaque systems.

So called black box systems include, for example:

Program Code Running on Corporate Servers

This type of code is usually inaccessible to third parties. Examples are recommendation algorithms of platforms such as Facebook, now X, formerly known as Twitter, or TikTok, that calculate which videos, photos, or tweets users will see next.

AI Models that Have "Learned" Certain Data Patterns Based on Training Data

Based on this, the models score or categorize new input data or generate texts, images, speech, or pieces of music. These programs are black boxes, opaque entities, whose operation procedures are often beyond the comprehension of their own developers.

Program Code on Your Own Cell Phone, PC, Tablet, or Other Devices

With sufficient technical knowledge, it is possible to theoretically examine how code on a device works – attention, journalists, this can be a criminal offense, depending on the country and the method used. To protect themselves from hacking, industrial espionage, or the critical gaze of the public, some manufacturers of smartphone apps and other programs implement sophisticated features to prevent analysis of their products.

BR DATA

BR RECHERCHE

AI + AUTOMATION LAB

THE ALGORITHMIC ACCOUNTABILITY REPORTING TEAM

The teams of AI + Automation Lab, BR Data and BR Recherche work in the field of Algorithmic Accountability Reporting (AAR) in an interdisciplinary way, bringing together technical, investigative, and data expertise.

Their work involves cooperations with various in-house units, the wider German Public Broadcasting System, as well as national and international partners, including the U.S. non-profit newsroom ProPublica, Guardian US and Der Spiegel.

HELGA VAN OOIJEN-DEML
Editor

SAMMY KHAMIS
Reporter

REBECCA CIESIELSKI
Reporter

ROB SCHÖFFEL
Editor

PHILIPP GAWLIK
Computer Linguist

KATHARINA BRUNNER
Journalist

SEBASTIAN BAYERL
Software Developer

VERENA NIERLE
Editorial Board

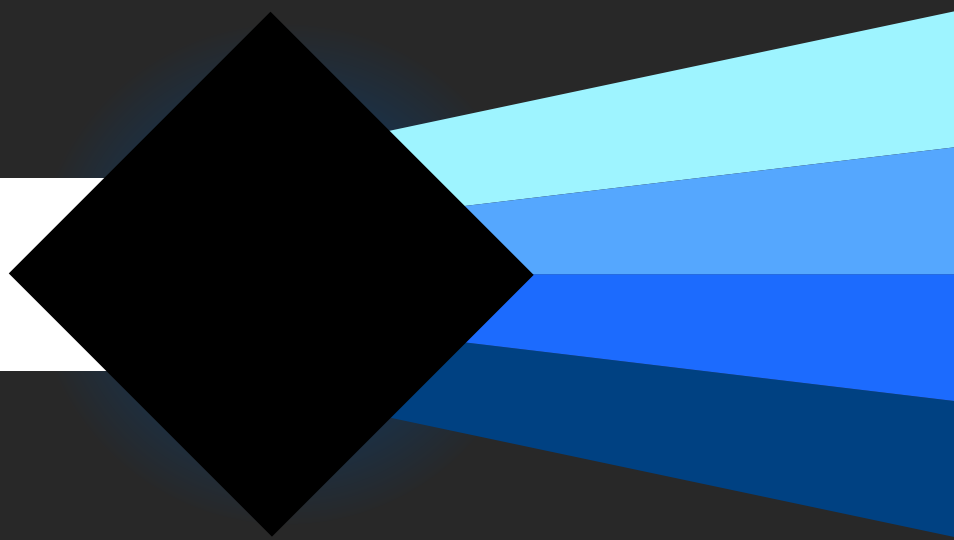**ULI KÖPPEN**
Editorial Board

MAXIMILIAN ZIERER
Reporter

PIA DANGELMAYER
Editor

MARCO LEHNER
Journalist & Developer

AI + AUTOMATION LAB

*The Algorithmic Accountability Reporting Team Works with Four Different Research Approaches*

#1

OUTPUT MATTERS. INPUT AS WELL: INVESTIGATIONS WITH EXPERIMENTS

For the story "Too sexy: How AI algorithms can discriminate against women", a BR reporting team examined four commercial image recognition systems. The focus was on services that measure how sexually suggestive an image is. This investigation was a cooperation with Guardian US.

For this purpose, the team evaluated more than 3,000 images in various categories, from underwear to sports and business looks. The results showed clear indications of gender bias. Compared to pictures of men, pictures of women are more often rated as more suggestive and can lose visibility on the internet.
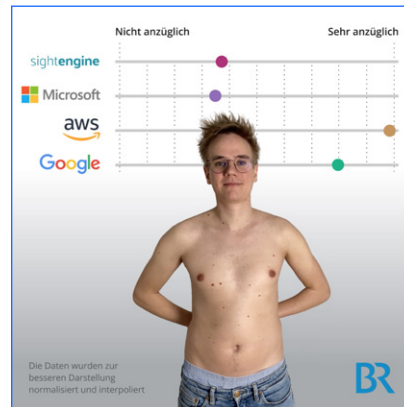
BR reporter Maximilian Osenstätter stands bare-chested in front of the camera. Behind him, a visualization shows how suggestive various algorithms classify the video. Initially, the reporter's naked torso received only a low score from two of the four algorithms tested.
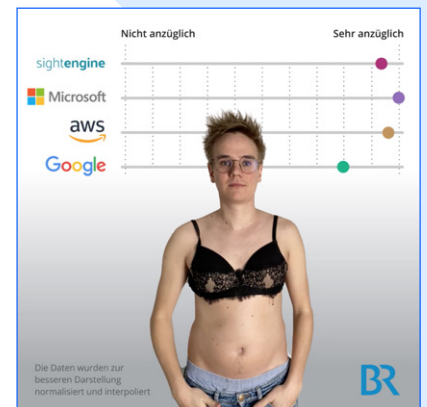
But when Osenstätter puts on a black bra, the result changes. Although his upper body is now more covered by the garment than before, the two AI models now rate the video as very suggestive. Obviously, the underlying image recognition algorithms have been primed with examples so that they evaluate images of people with a bra this way.

Fairness or Prejudice?

The investigation "Fairness or prejudice? Questionable use of artificial intelligence in job applications" reveals that the product of a German start-up that analyzes videos of job applicants is flawed. The software promises to extract personality traits of job applicants out of video interviews. Instead, the product assumed completely different personality traits for the same person, depending on changes of outfits or the video background.

BR Data produced several hundred video clips with test persons. The goal was to find out what can influence the software and how different variables might affect the personality assessment. The team set up two different experiments: On the one hand, an actress in different outfits with always the same text and mimics answered different job interview questions. On the other hand, video producers altered recorded videos in a technical way. In both scenarios, this ensured that only a single factor was specifically altered in each experiment. The results of the software differed considerably in some cases.

Less prejudice, more objectivity. An application process that does not depend on the personal taste of a recruiter. This is the promise of many AI companies that are entering the market worldwide.
Among them a start-up from Munich. According to the developers of the software, artificial intelligence analyzes voice, language, gestures, and facial expressions to create a behavior-based personality profile.
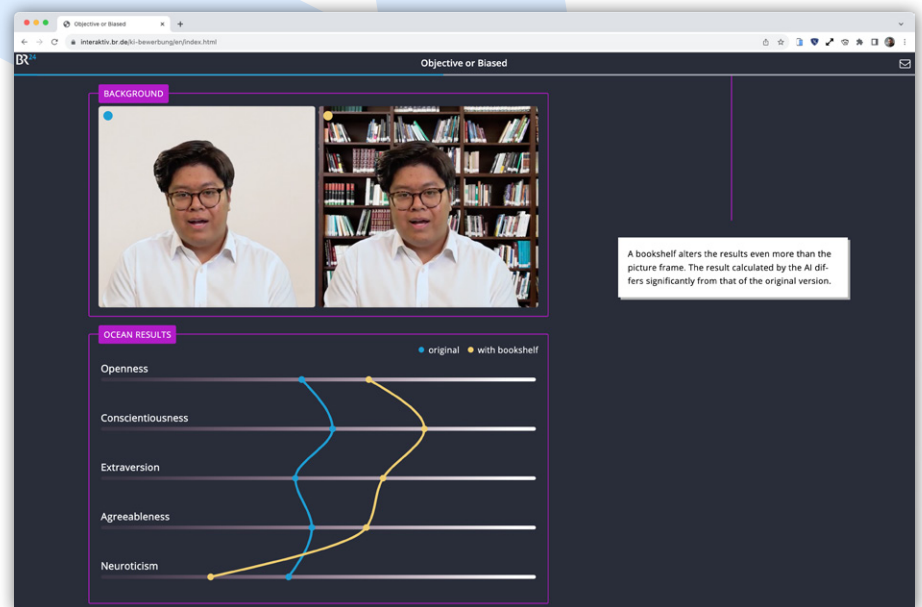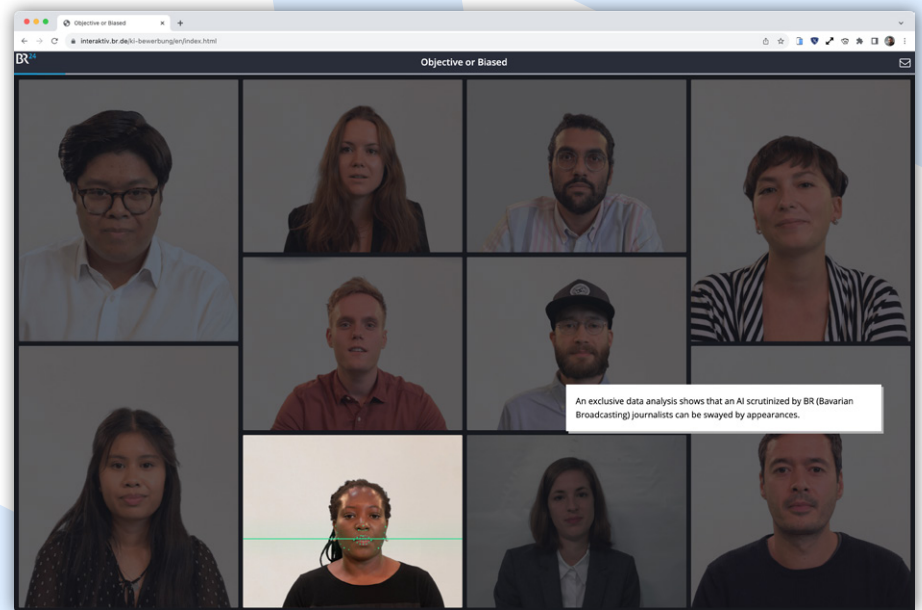
Videos about self-harm and sui-cidal thoughts are still a prob-lem on TikTok. This is shown by a BR data analysis. Users can be drawn into a filter bubble of dan-gerous content. This is because an algorithm decides what users see on TikTok to a greater extent than on other social networks.


05:34
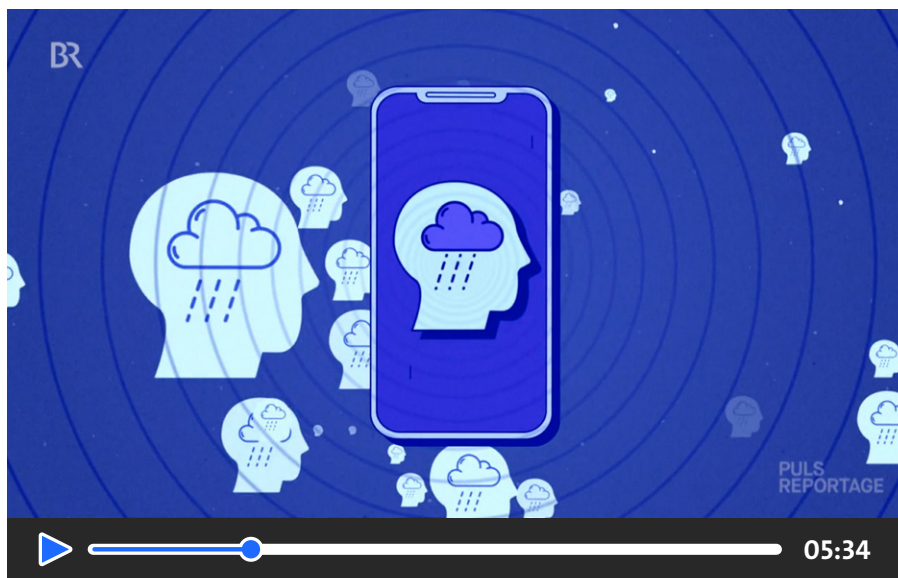
How Dangerous Is TikTok?

An algorithm curates the most important part of the TikTok app, the "For You" feed. There you will find an entire subgenre where young people pro-duce videos about depression, self-injury or suicide. An experiment conducted by BR Data and PULS Re-portage shows that German users can get sucked into a filter bubble by interacting with such videos, where TikTok heavily features this type of content in their feed. For the experiment, the journalists from BR Data set up several test accounts and simulated the be-havior of people interested in videos about depression, self-harm and suicidal thoughts. The result: after a short time, the feed consisted almost entirely of such content. After only about 150 videos or 45 minutes of use, an average of one in three vid-eos contained a hashtag on top-ics such as sadness, depression, self-injury and suicidal thoughts.

To study black box algorithms, experiments are often the only way to approach them. Journal-ists determine the input for an algorithm and observe the out-put — what happens in between is not visible to the reporters. Besides collecting enough data, it is important to eliminate possible unwanted influences on the input side as far as pos-sible. The BR teams use experi-ments to understand whether algorithms do what they are supposed to do or whether the systems discriminate unfairly, deliver faulty and thus unusable results, or influence their users in a manipulative way.
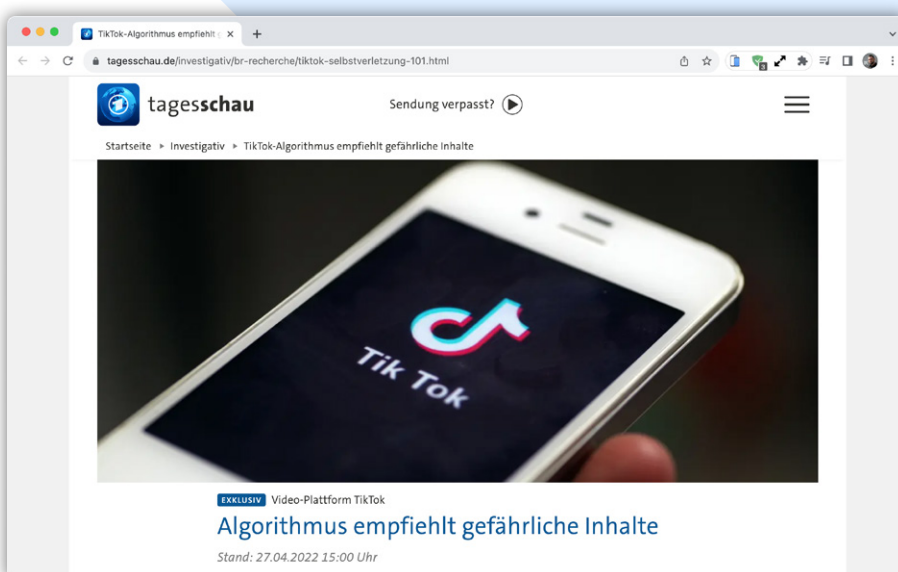
#2

SOFTWARE ON THE TEST BENCH: INVESTIGATIONS WITH TECHNICAL ANALYSES

A team from BR tested the Android and iOS smartphone operating systems to determine whether apps could secretly listen in. The team developed a smartphone app for the investigation. It shows that all apps on older but widely used Android versions can record audio signals completely unnoticed under certain circumstances. If the screen was turned off, but the app was not closed, the hidden recording worked for over an hour. In another test, the team showed that even programs running on PCs can, under certain circumstances, secretly film, listen to and photograph the screen contents for a long time.

Journalists use such tests and methods of reverse engineering – the technical analysis of systems – to investigate the functionality and security vulnerabilities of software. The results of this research are met with broad interest, as shown by the Puls report "Can iPhone and Android phones secretly listen in? We program apps to prove it!", which has over a million views on YouTube.
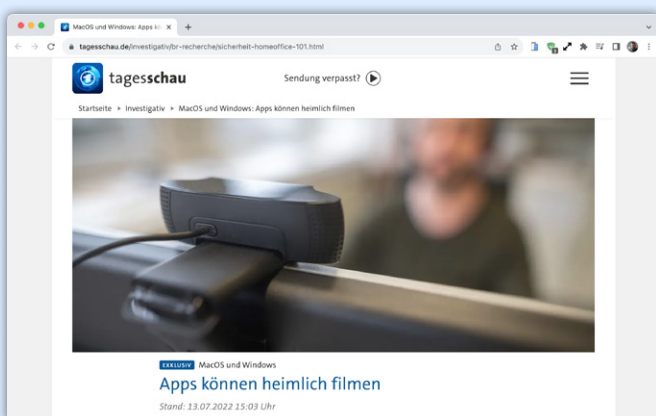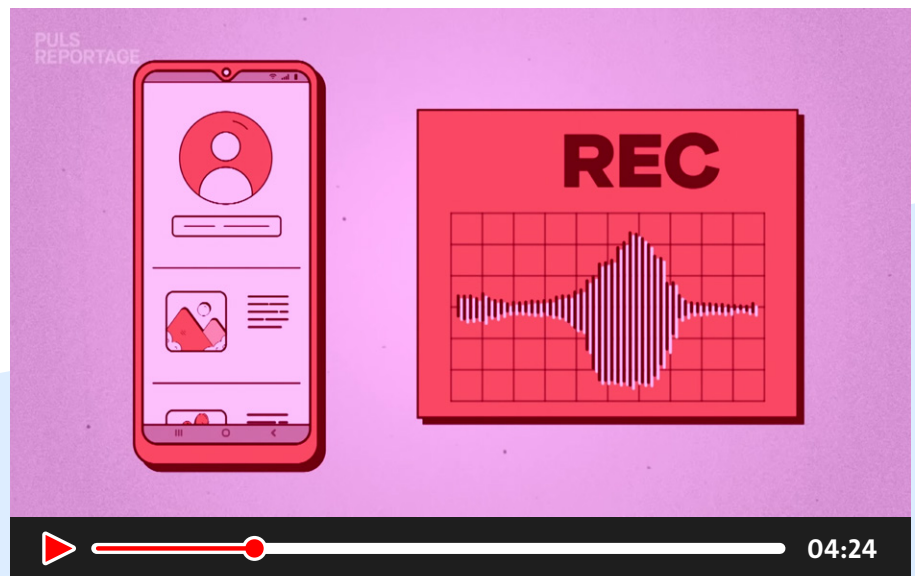
Many people have already experienced it: You're sitting with friends at dinner table, talking about a product or topic that you're sure you've never actively searched for online. A short time later, ads on that topic suddenly appear on your social networks. On the internet, you can find many stories of people who swear that unusual topics of conversation have led to online advertising.

As convincing as these experiences may sound, such anecdotal claims cannot confirm that cell phone apps are secretly listening in to place targeted advertising.

#3

RIGHT OF ACCESS: INVESTIGATIONS WITH LEGAL MEANS

BR reporters analyzed GDPR data of food delivery Lieferando riders, which they have shared with BR, to determine whether Lieferando improperly monitors the riders. The riders had requested their data from their employer by the European General Data Protection Regulation (GDPR).

The documents show that the app collects 39 data points per delivery: The app tracks a driver's movements precisely from the moment they are assigned an order, pick it up, and deliver it. In addition, Lieferando's app recorded whether the rider adheres to individual time specifications or not. If a driver arrived late at a restaurant or customer, the delay was recorded.

The data are personalized, they can be assigned to individual employees. In some cases, available to BR, more than 100,000 data points were recorded for fully employed riders in one year.

"Now I have 30 seconds left," says Daniel. Then he starts his shift as a driver at Lieferando by pressing "Play" in the "Scoober" app: He is ready for the first job of the day. Daniel, which is not this real name, tries to not activate the app only one second late. He would rather not attract any negative attention: "I want to be on time, because they take note of that."
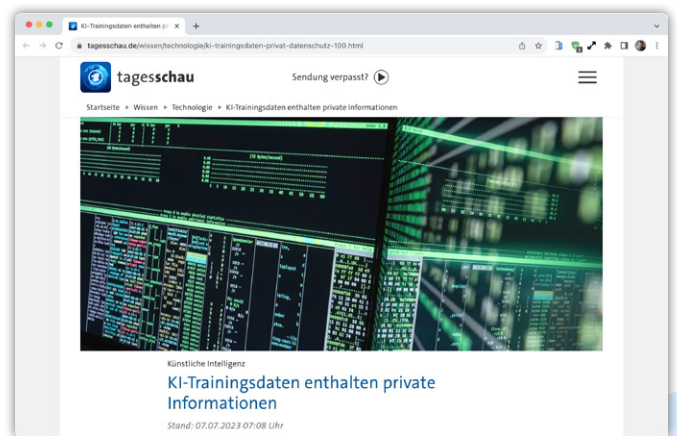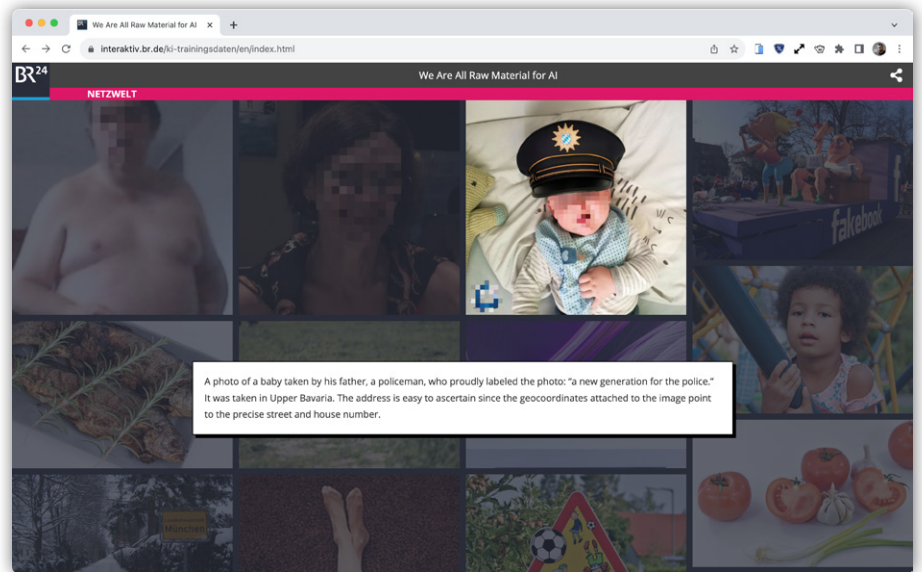
In this case, "note" means that the app stores the exact start of his shift. In addition to his bike and backpack, the "Scoober" app is Daniel's most important tool. He receives his orders via the app. His employer, Lieferando, has not told him what else "Scoober" knows about him: "But I'd like to know," says the rider.

The Raw Material of AI

A BR team also used the GDPR as an instrument for investigating the story "We are the raw material of AI". Training data for AI systems consist of countless images and texts from various sources, such as websites or books.

A BR analysis shows that they contain sensitive and private data — usually without the people concerned being aware. In the analyzed training files, the team found a picture of BR journalist Elisa Harlan. She used GDPR to make requests to companies and organizations such as OpenAI, Midjourney, and LAION — not all of which were successful.

Since GDPR came into force, individuals have the right to request information about what personal data companies and employers store about them. Depending on how detailed the answers are, they are suitable as a tool for investigations into algorithms that encroach on people's privacy.

A photo of a baby taken by his father, a policeman, who proudly labeled the photo: "a new generation for the police." It was taken in Upper Bavaria. The address is easy to ascertain since the geocoordinates attached to the image point to the precise street and house number.

Künstliche Intelligenz
KI-Trainingsdaten enthalten private Informationen
Stand: 07.07.2023 07:08 Uhr

Tagesschau

Plötzlich im Datensatz. Wenn die KI mit Dir trainiert

07.07.2023 · noch 30 Min. · 29,5 MB

What if a private photo suddenly appears in a data set that is used to train artificial intelligence? A huge dataset that contains not only photos, but also all kinds of other images from the Internet. This happened to Elisa Harlan, whose photo is part of the largest publicly accessible AI training dataset — without her being aware.

A look at the supply chain for AI training data raises the question: How do images become raw material for artificial intelligence — and is there a way back?

#4

EXPLAIN THE ALGORITHM:
INTERVIEWS AND REQUESTS
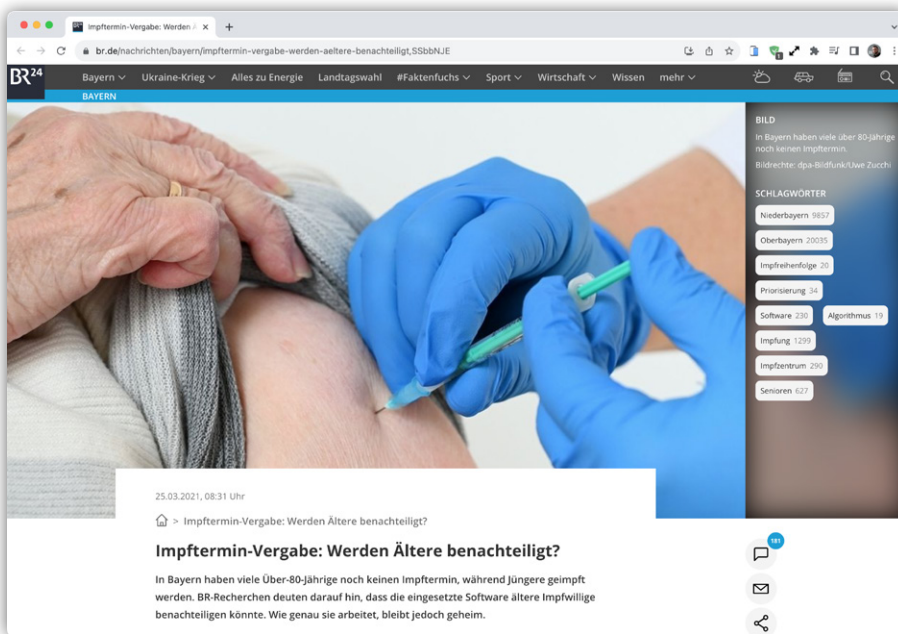
The couple Peter and Erika S. from the district of Munich: Both are in a high priority group for anti-Covid vaccination. He is 81 and has already been vaccinated once, his 80-year-old wife has been waiting for weeks. Both have registered for a vaccination appointment via Bavaria's Bay-IMCO software. It is unclear to them what criteria are used to assign the appointments. "I have some friends who are younger than me and have already been vaccinated," says Erika.

The BayIMCO appointment system assigned a score value to each person registering for the Covid vaccination in 2021. For older people, this value corresponded to their age. Younger people received a random score that corresponded to the age distribution of the prioritization group. In priority group 1, for example, a random value between 80 and 90. The higher the value, the faster a person gets an appointment. This "dice roll" was possibly the reason why 80-year-old Erika S. had to wait a long time for her vaccination appointment. Statistically, she was disadvantaged with her relatively low age score compared to the random numbers assigned.

BR reporters were able to look at Erika's database entries at the local vaccination center in Oberhaching, with Erika's consent. This data, as well as interviews with vaccination center operators and inquiries to the Bavarian Ministry of Health, provided concrete indications why the Bavarian vaccination algorithm disadvantaged elderly people.

This and similar investigations by BR teams show: Reporting on algorithms is also possible with standard journalistic methods: with interviews and documents, requests according to the Freedom of Information Act (FOIA) or whistleblowers.

CHARLIE BECKETT

Head of London School of Economics Journalism Think Tank Polis and
Founder of the global JournalismAI network

*"BR's algorithmic accountability reporting is a model for journalists
everywhere. Their investigations are informed, lucid and compelling.
Working without moral panic or hyperbole they have given us in-depth,
detailed and accessible insights into some of the key risks and hazards
that new technologies such as artificial intelligence can bring, from bias
in HR recruitment to invasions of intellectual property and privacy in
training data. At this time where algorithmic technologies are more
powerful in all aspects of our lives, we desperately need this kind of
critical, independent and knowledgeable public service journalism."*

ETHICS AND LAW: WHAT RULES APPLY TO AI IN THE NEWSROOM?

In 2020, Bayerischer Rundfunk was one of the first media companies worldwide to publish ethical guidelines for dealing with artificial intelligence. In the meantime, more and more newsrooms have internally agreed on and published guidelines on AI. In July 2023, Nicholas Diakopoulos and Hannes Cools analyzed 21 international guidelines from 13 countries for the project "Generative AI in the Newsroom". The following aspects played a recurring role:

- ✓ Strategic goals for AI deployment
- ✓ Desirable and undesirable application scenarios
- ✓ Editorial control and approval of content created with AI support
- ✓ Transparent labeling of such content
- ✓ Editorial responsibility and accountability
- ✓ Protection and security of data and sources
- ✓ Responsible room for experiments
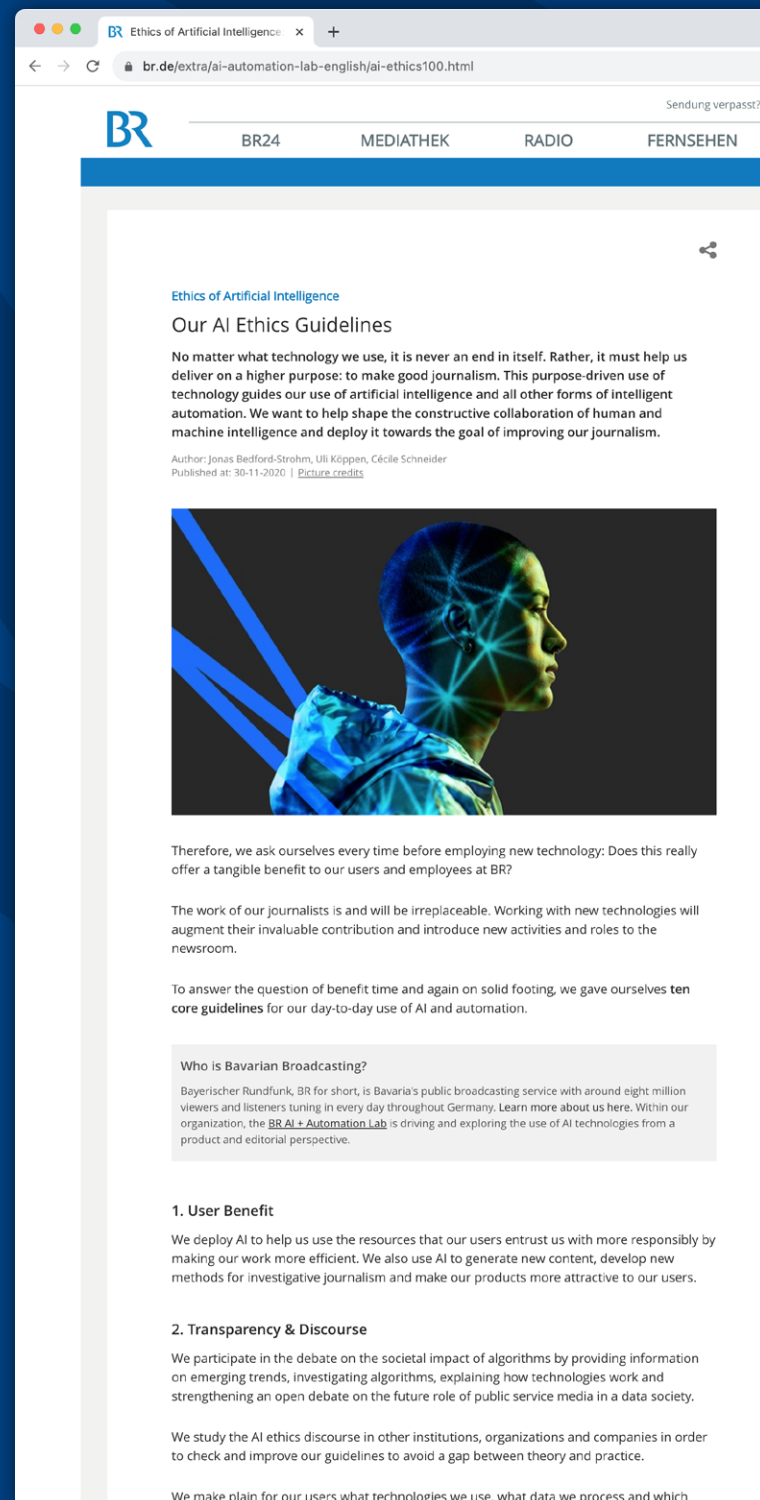
The BR guidelines for the use of artificial intelligence also cover these aspects. First and foremost is the commitment to the staff: Journalists are irreplaceable for BR.

Using Artificial Intelligence for the Public Service Mission

At BR, AI should help to create new content for users and make existing content more appealing, to support investigations and journalistic work in general, and to handle public funds responsibly.

Algorithmic Accountability Reporting is intended to help fulfill the public service information and education mandate by explaining AI developments and investigating algorithms.

Algorithms, automated decision-making systems, machine learning, artificial intelligence used by states, the economy or other actors — regardless of the nature of the technological phenomena: Knowledge about AI and automation in newsrooms is essential to contribute to the societal debate about algorithms.
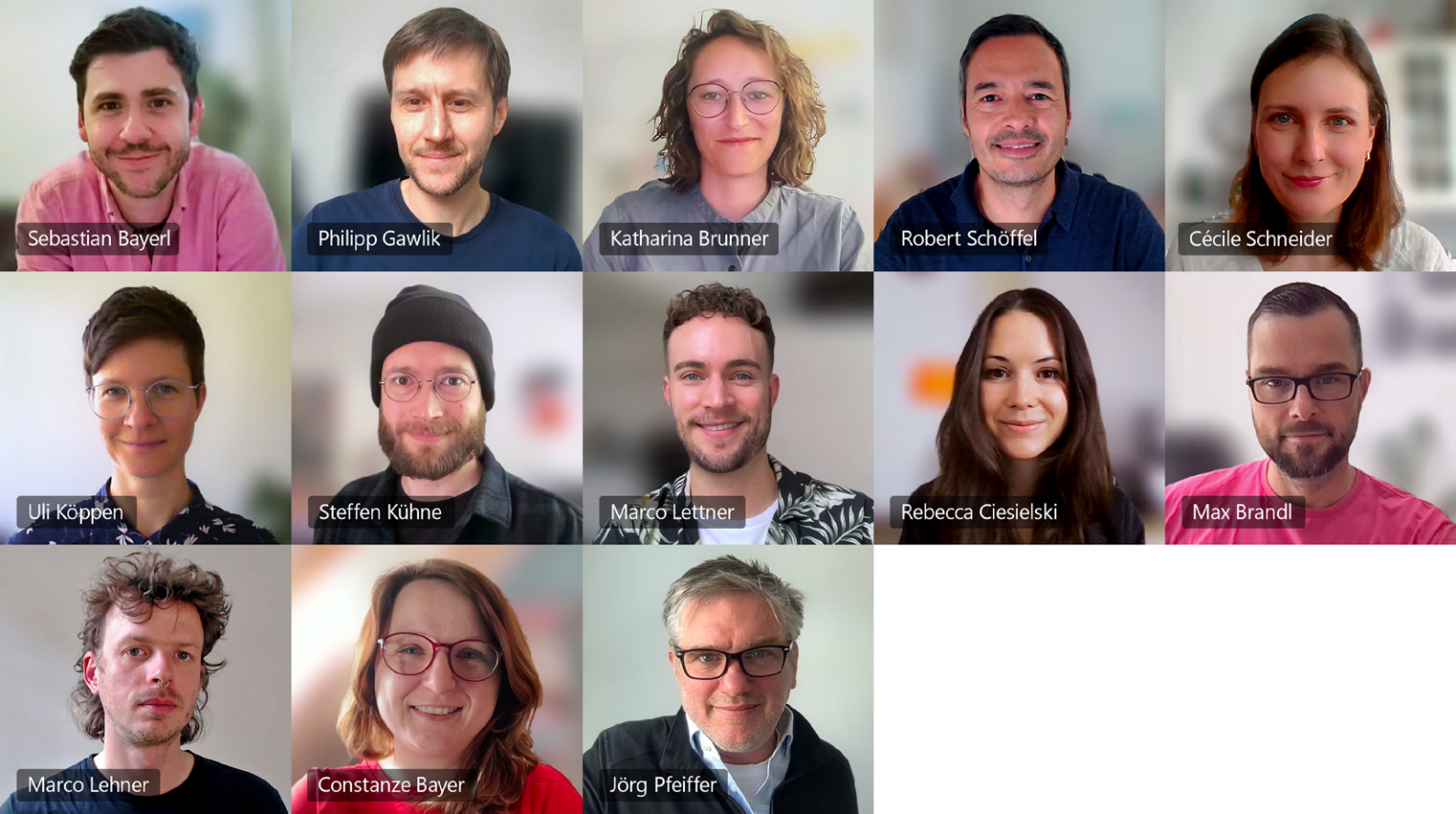
AI PODCAST LAUNCHED BY
BR AND SWR

Artificial intelligence as a beat has long since arrived in the daily news business. The AI teams of BR24 and SWR offer orientation in a weekly podcast. It covers technological developments, possible consequences for economy, education, society and our security – and many practical AI applications in everyday life.

"The possibilities of artificial intelligence can be exciting, but they can be equally unsettling. One thing is clear. We are on the brink of a revolution. Our podcast looks at how our lives are changing. AI affects everyone. That's why the podcast is aimed at everyone, not just an audience of experts," says BR editor-in-chief Christian Nitsche.

The AI podcast is available in the ARD Audiothek (in German).

Sebastian Bayerl
Philipp Gawlik
Katharina Brunner
Robert Schöffel
Cécile Schneider
Uli Köppen
Steffen Kühne
Marco Lettner
Rebecca Ciesielski
Max Brandl
Marco Lehner
Constanze Bayer
Jörg Pfeiffer

THE AI + AUTOMATION LAB AT BR

The AI + Automation Lab has been working as an interdisciplinary team at BR since 2020 at the intersection of journalism, computer science and product development. The Lab focuses on artificial intelligence and automation for user-centered journalism and investigative reporting. Machine learning experts work together with data journalists, designers, and product developers. Stories and products are created in interdisciplinary collaboration. The focus is on Algorithmic Accountability Reporting and the development of AI-supported news formats and products.

br.de/extra/ai-automation-lab

x.com/BR_AILab

ard.social/@br_ailab