
Constructive AI Compass

Knowledge, Tools and Tips for
Everyday Journalism

b

bonn institute

The Bonn Institute

Through practice-oriented research, publications, events and trainings, the Bonn Institute serves the media industry's increased need for networking and knowledge exchange on constructive and user-oriented journalism. By working at the intersection of research and practice, the institute provides media professionals with empirically proven tools that can be used every day.

Our goal is to help journalism develop so that it is centered on peoples' interests, which is especially important given growing social challenges. An internationally connected non-profit, the Bonn Institute combines journalistic expertise with insight from media studies, user research, psychology and mediation.

The Bonn Institute was founded in 2022 in partnership with Deutsche Welle, the Rheinische Post media group, RTL Deutschland and the Constructive Institute in Denmark. Ellen Heinrichs is the founder and CEO. The Board of Trustees includes Jay Rosen, Wolfgang Blau, Astrid Maier, Kübra Gümüşay and David Schraven, among others.



Constructive AI Compass

**Knowledge, Tools and Tips for
Everyday Journalism**

Contents

Foreword Page 3	Constructive Journalism	9
	Focus on solutions	10
Invitation Page 5	Multiple perspectives	11
	Constructive dialogue	11
Index and glossary Page 65		
Sources and literature Page 66		
The team Page 70		
Acknowledgements Page 72		
Publication details Page 73		
Five constructive prompts to get you started Back inside flap		
	Understanding AI	13
	Automated text	14
	The steps in AI's development	14
	Known AI models	22
	Using AI Responsibly	25
	Legal issues	26
	Journalistic diligence and transparency	29
	Internal guidelines	31
	Training AI models	34
	Environmental impacts	36
	Using AI Constructively	41
	Constructive prompt engineering	42
	Prompts for brainstorming constructive topics	49
	Prompting for many different perspectives	51
	Prompting and constructive interview questions	55
	From speech to text: Transcribing with AI	57
	Constructively tuning ChatGPT	59
	Can AI recognize synthetic content?	62



Foreword

A broad array of journalistic offerings and financially sustainable media companies are nothing less than the firewall of our democracy. As the Media Authority of North Rhine-Westphalia (Landesanstalt für Medien NRW), it is our task to ensure such a diversity of media providers and offerings is present in our state.

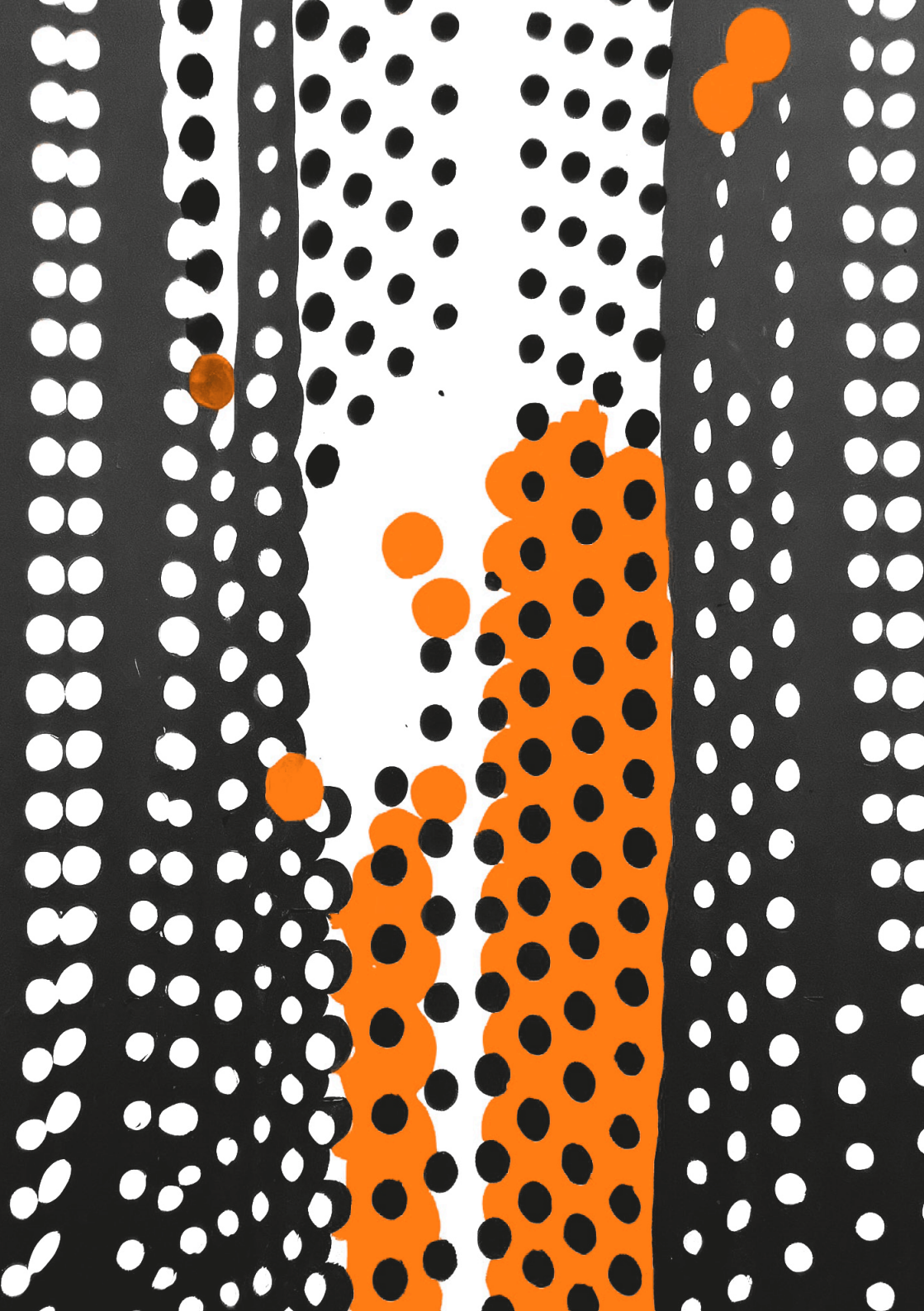
To maintain this going forward, we must work together with media companies and important stakeholders, like the Bonn Institute, to fight to renew journalism and jointly shape the future and competitive ability of media. AI will play a decisive role in enabling journalistic innovation in the coming years. It can contribute to a wide range of perspectives and information, which in turn facilitates open dialogue – as long as professional values, ethical standards and user needs guide its use.

The Bonn Institute's "Prompt Book," which has been made possible by support from the Journalism Lab of the Media Authority of North Rhine-Westphalia, provides knowledge transfer that can make a lasting difference. It will hopefully serve as a practical tool for many media professionals to constructively engage with AI.

Simone Jost-Westendorf

Head of Journalism Lab

Media Authority of North Rhine-Westphalia



Invitation

Are new technological developments destroying the value of human work? Could many people even lose their jobs? Such questions were already at the center of conflicts during the early industrial era, more than 200 years ago, when craft and trade guilds were deeply shaken by the possibility that machines could take over tasks that had previously been done by hand.

And such questions remain relevant today, for we have once again reached a point where a new type of machine is set to take over tasks that until now only humans have been able to do. Today, as before, all sorts of professionals, including we journalists, are asking themselves how this should be managed.

Following the widespread introduction of computers in the 1980s, many different technical innovations became part of the normal workday in journalism. Computers (first desktops and then portable models), the internet, mobile phones, social media and ever more different types of software and tools: All these things have changed journalistic research, production and distribution. Change has become a daily constant, making many things easier, and some things even possible in the first place. Yet it has also caused a certain level of exhaustion.

And now artificial intelligence has arrived, known in short as AI.

The implementation of steam engines in economic processes more than 200 years ago heralded an era of radical upheaval, later to be known as the Industrial Revolution. Contributing to this upheaval were the inventions that made use of the steam engine: Steamboats that revved up global trade; railroads that transformed how

people and goods were transported; steam-powered looms that revolutionized textile production; and many more inventions that drew on the new technology.

Two hundred years from now, how will we view the societal change unleashed by AI? What inventions will AI make possible?

This depends in part on how we confront and manage this new technology. Fearfully? Hopefully? Ambivalently? There are good reasons for each approach. We are aware of the danger posed to society by the misuse of AI (see e.g. Adeoso et al. 2024, p. 125 ff.).

At the same time, we also see the great potential these new developments hold.

In order for journalism to take advantage of AI's constructive possibilities, the technology must be applied in a way that offers additional value – for people, the media and democracy.

We believe two things are essential to achieving this: A solid understanding of the new technology, and inspiration for using it constructively, coupled with the relevant know-how. This publication aims to provide both. It specifically focuses on how to use text-generative language models such as ChatGPT in journalism. Programs that generate images and videos are not addressed; including such applications would go beyond the scope of this publication.

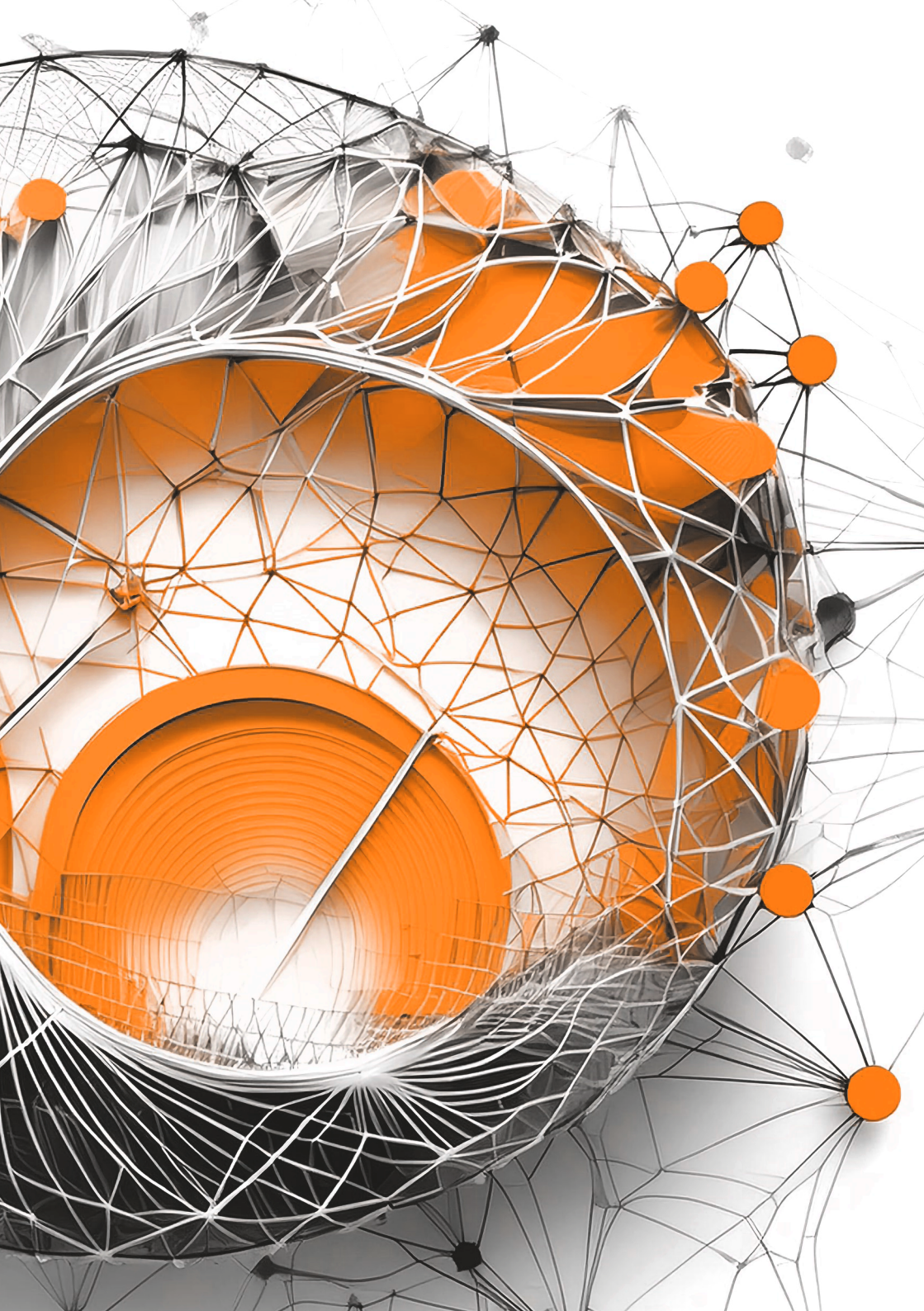
Our booklet is particularly aimed at journalists who want to use AI responsibly in their day-to-day. We invite you on a journey that brings together constructive journalism and AI. While this publication does not explicitly examine questions that management needs to consider in the context of AI, managers are also invited to join this expedition, as is anyone who is interested in the fundamentals of artificial intelligence and situations where it can be constructively applied.

After briefly visiting the three central elements of constructive journalism, we will dive into how AI works – in other words, into automated decision-making systems. We will then turn to the legal and ethical framework needed to use AI responsibly and look at the related questions we need to ask.

Things will then get practical with some constructive prompt engineering. We will discuss prompting basics as well as present some scenarios in which the three elements of constructive journalism also play a major role. We will also examine how AI can help with research.

The second half of the practical section is dedicated to GPTs (generative pre-trained transformers), or digital assistants for different tasks based on ChatGPT. Anyone who is interested can learn how to easily train their own digital assistant.

How you read this booklet is entirely up to you. This journey is not structured as a package vacation; the pages numbers do not prescribe the reading order. Instead, this booklet is designed so you can start with any chapter and chose where you go next. We hope you have fun as you explore, discover and experiment!



Constructive Journalism

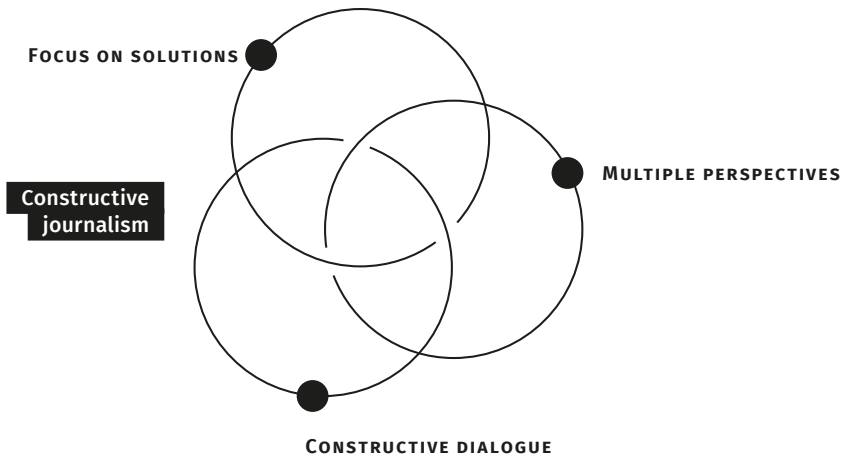
Diverse change processes, political and societal conflicts, disinformation, news avoidance and diminishing trust in the media all pose enormous challenges to the journalism industry.

The following question is therefore key: How does journalism need to change in order to remain relevant, sustainable and beneficial to people, media and democracy? Constructive journalism seeks to answer precisely this question. Studies show it has the potential to reduce news avoidance, counter societal polarization and provide media companies with new monetization opportunities, among other things.

It has three main elements: a focus on solutions, multiple perspectives and constructive dialogue.

Constructive journalism aims to provide media users with a future-oriented, fact-based, nuanced picture of reality. By researching possible solutions as thoroughly as problems, it counters a one-sided, negative worldview and bolsters feelings of agency among media users by illustrating various potential courses of action.

By purposefully emphasizing diversity and different perspectives, it reflects the world in all its complexity and counters oversimplification and polarization. And by redefining the role of journalists as moderators of a public constructive dialogue, it creates new possibilities for better conversations in our society.



Focus on solutions

Traditional journalism often ends once a problem has been adequately identified or described. Constructive journalism goes one step further. It places a special focus on possible solutions for societal problems, thereby broadening the scope of journalistic research. One of its central questions is, “Who does it better?”

Solutions journalism, as developed by the Solutions Journalism Network, includes four key elements that make it easier for journalists to produce high-quality, critical and constructive content. These elements are:

- Placing a possible solution front and center in a report;
- Providing evidence, either quantitative or qualitative, that shows the effectiveness of a possible solution;
- Ensuring a solution can be transferred elsewhere;
- Identifying the limits of a possible solution.

These elements prevent journalists from painting overly rosy pictures, engaging in greenwashing or getting trapped by other pitfalls. They can be used as a checklist for research and content creation.

Multiple perspectives

While solutions journalism primarily focuses on practical aspects, the idea of multiple perspectives also addresses systemic factors, whether this is staff makeup or one's own blind spots and prejudices and how to better understand them. Is my view distorted by a habitual focus on the negative? What interesting and relevant perspectives on a topic have I failed to include in my research? Are our teams diverse enough to make sure we accurately capture society in all its diversity?

Journalism that has a wealth of perspectives is journalism that is relevant because it is aware of the information needs of society's many different members.

Constructive dialogue

Constructive dialogue is the third central element of constructive journalism. Its key aspects include impartiality, empathetic interest and the ability to listen, as well as a focus on what unites (and not just what divides) and a forward-looking approach. This matters because journalists are needed not only as providers of relevant information and different perspectives but also as moderators between people and groups in society.

Constructive journalism is therefore an important tool for organizing human communication and understanding. Creating places of dialogue, including digital ones; bringing people of different backgrounds into conversation with each other; asking better questions; promoting and moderating peaceful, future-oriented debate on relevant social questions – all these things enable progress and can significantly strengthen social cohesion and democracy.



Under- standing AI

The origin of the term “artificial intelligence” is more banal than most think: It can be traced back to Stanford Professor John McCarthy. He invented it in 1955 to secure funding for a two-month research project to take place the following year. However, his funding proposal contained neither an explanation nor a formal definition of the term; in fact, it only used it as an umbrella term. “The study is to proceed on the basis of the conjecture that every aspect of learning and any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it,” McCarthy wrote in his proposal, before going on to list “some aspects of the artificial intelligence problem” (see McCarthy 1955 et al. 1955). That was it. In other words, artificial intelligence is a marketing term. McCarthy received the funding, and the world received the term. But it turns out it was not possible to precisely describe what intelligence is (see Brooks 2018 and Whittaker 2023).

This is the starting point for the current criticism of the term artificial intelligence (AI), as we still lack a general definition of intelligence. And if we don’t know what constitutes human intelligence, how are we supposed to define artificial intelligence? Yet even if the term isn’t totally clear, AI is already here. AI is not new. Neither are the related fields of research. What is new, however, is that since 2022, very user-friendly, generative AI programs have been publicly available. So it’s more than time to take a closer look at what’s behind it.

Automated text

Current discussions about how to use AI in journalism often mistakenly lump together two totally different approaches: automatic content creation on the one hand, and synthetic content generated by large language models on the other. However, it is crucially important that we journalists differentiate between these two things: With automatically created texts, we can understand where the result came from and also figure out how any errors arose, whereas we cannot yet do this with the results from language models. Why? With automated creation, an individual provides inputs that are specific enough to make results understandable. In contrast, language models generate content independently, meaning it's not possible to understand where it comes from.

Automated texts require two things: First, structured data, such as a table with weather data, election results or sport outcomes. Second, a program like Retresco or AX Semantics, in which rules are defined and text components prepared. These then guide the creation of automated texts.

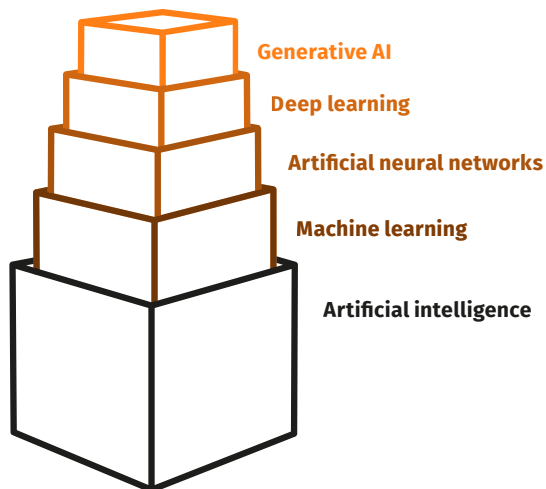
The Associated Press (AP) news agency has been using this technology since 2014 to produce texts about the quarterly reports of publicly listed companies. The list of users has since grown long. The German weekly magazine Der Spiegel used the technology during the federal election of 2021 to present data analysis from each of the nation's 299 election districts.

This type of text creation can save journalists a lot of work and open up new possibilities, but it isn't AI.

The steps in AI's development

To understand what's behind that which we refer to as artificial intelligence and how it works, we're going to lift up the hood and take a look at what is hidden underneath. What do terms like artificial neural networks, transformer or GPT refer to? This knowledge isn't yet part of our repertoire as media professionals. However, if we want to start using AI in our day-to-day work, we should have a basic understanding of it. Whoever drives a car doesn't have to be able to build or repair one. But being able to distinguish the motor from the gas tank from the exhaust is unquestionably 20th-century common knowledge. Twenty-first century common knowledge will soon include a basic understanding of AI technologies.

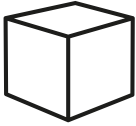
Steps in AI's development



So how are artificial intelligence, machine learning, artificial neural networks, deep learning and generative AI all related? How are they different? And what do you need to know about them?

To better understand all this, it can help to compare the technological developments to the stacking toys that many children play with. While different types exist, let's picture one consisting of wooden building blocks that are open on one side and that can be stacked into a tower in descending size order, with the largest block at the bottom and the smallest at the top. The taller the tower gets, the wobblier it becomes. The challenge for young kids is to figure out the correct order of the blocks, while babies might enjoy knocking the tower over. Parents like that the blocks can be placed nesting-style inside each other, saving space.

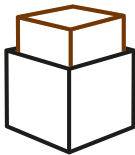
AI's development is similar to this stacking tower; generative AI didn't suddenly appear out of nowhere but rather came into being one step at a time, which each step building upon the previous one. Generative AI, the thing keeping us busy right now, is the most advanced technology in the realm of artificial intelligence and therefore the crowning block of our tower.



Artificial intelligence

The largest and first block in our imaginary stacking tower is artificial intelligence. This is what experts refer to when machines are capable of delivering rational automated results without the programming specifically dictating the way to an answer. The Fraunhofer Institute for Cognitive Systems (IKS 2024) describes it as follows: “Artificial intelligence imitates human cognitive abilities in that it is able to recognize and sort information from input data.”

In the course of AI history, researchers first tried to provide software with all the rules a system needed to be able to make its own decisions. This is known as an expert system. The world’s first chatbot, ELIZA, which was released in 1966 and imitated a psychotherapist, is one such example. Programs were also supposed to initially learn to process human language by following various inputted rules. Success with this was limited, however, as the rules governing language proved to be too complex. So researchers looked for new approaches. Machine learning plays a crucial role in today’s major AI applications.



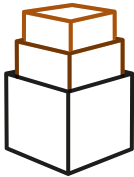
Machine learning

Machine learning is a subset of artificial intelligence and therefore the second block in our tower. It differs from the larger field of AI in that computers independently “learn” something from data rather than from rules. This starts with a learning algorithm being fed very large quantities of data. The algorithm analyzes all this data for patterns that it then uses to develop a statistical model. Sometimes human feedback guides these processes, but sometimes it doesn’t. Depending on the method, this is referred to as supervised or unsupervised learning. Programmers decide which learning method to use, and it is built into the learning algorithm.

The statistical model that the software develops during the learning process becomes the basis for its predictive abilities once the training process is over. This means that down the road, the software can incorporate previously unknown data into its model and generate similarly correct answers.

This is where things start to get a bit murky. The reason? “Programmers only provide the machine with a rough structure. The details are set by the learning method in conjunction with the data,” says computer engineering professor Katharina Zweig (2023). In other words: What we describe as machine learning is a process that even programmers no longer fully understand. They can see the statistical model that the software develops during its learning process, but they can only theorize as to how the system settled on it.

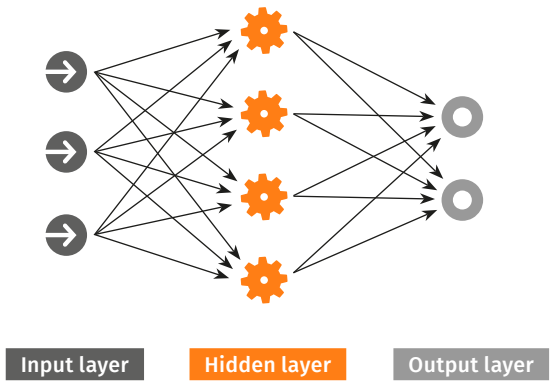
The theory of machine learning, just like the theory of artificial intelligence, is more than 70 years old. The first related research paper was written by Alan Turing in 1950, before the term “artificial intelligence” had even been created. In it, he posed the question of whether machines can think and how this could be tested. Research on this was initially confined to the theoretical realm, as two things needed to realize and test these ideas would be missing for a long time: data and computational power. Now and then there would be some notable successes, such as in 1996, when IBM’s chess-playing computer Deep Blue beat then-chess grandmaster, Gary Kasparov. More than 20 years later, in 2017, a computer named AlphaGo, which had been built by the Google subsidiary Deepmind, beat the world’s best player of the game Go, Ke Jie. However, a breakthrough in AI’s practical applicability had yet to take place.



Artificial neural networks

It would be cat videos, of all things, that would end up being a useful resource in AI’s further development. This brings us to the third step of development and the next block in our tower: artificial neural networks. This term refers to a special model of machine learning that was invented in 1943 (see Brooks 2018). Its breakthrough took place much later, however. Since then it has been frequently utilized in complex learning processes.

Single layer
neural network



Artificial neural networks are very sophisticated algorithms whose structures are modeled on the system of neurons and synapses in the human brain. Where the brain has neurons, an algorithm has various instructions for performing calculations – or, to put it more precisely, mathematical functions. These tell the software to bundle together certain inputs and compute them in a certain way. The results then become the inputs for the neurons at the next level.

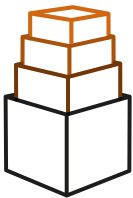
Just as synapses in the brain only pass on stimuli that exceed a certain intensity, the algorithm also determines the extent to which an individual result is passed on and incorporated into the next level. This means that not all results are treated the same way – comparable to what occurs in the human brain.

A simple artificial neural network has three layers: An input layer, a hidden layer and an output layer. The input layer consists of mathematical formulas designed by humans. They define the parameters for the machine and start the computational process. The results of this first layer are then further processed autonomously by the second layer. The computations for the hidden layer are predetermined.

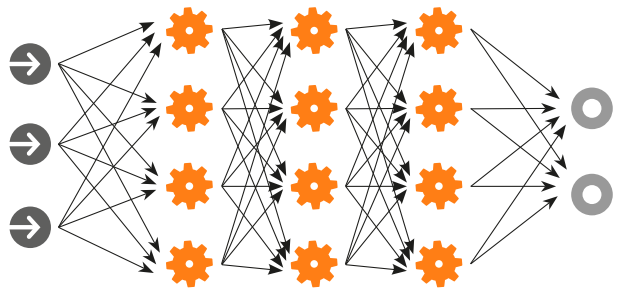
The machine’s “learning process” refers to how weights are set over the course of many repetitions. These weights affect how the results of one layer are incorporated into the calculations at the next layer. The second layer and all subsequent layers are hidden, however; they are where the algorithm performs further computations using the results of the respective previous layer and the weights it itself has developed. Programmers cannot see what takes place in these hidden layers. This is characteristic of artificial neural networks and the reason why AI is often described as a black box.

Deep learning

But what does this have to do with cat videos? As mentioned earlier, the mass quantity of data needed for AI’s development was simply not available for a long time. This changed in the early 2000s. After YouTube was founded in 2005, millions of videos were uploaded – a treasure trove of data that Andrew Ng took advantage of: Using a training dataset of more than ten million videos, he was able to train an artificial neural network to recognize cats in videos. He wanted the software to be able to learn what a cat looked like without him having previously shown it one (see DataScientest 2023). This was the start of a new era, that of deep learning, the fourth block in our stacking tower.



Deep artificial neural network



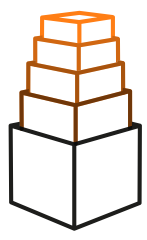
Input layer

Hidden layers

Output layer

Deep learning utilizes neural networks consisting of many layers. An artificial neural network that has at least two of these hidden layers is referred to as a deep neural network. However, according to the Fraunhofer IKS, it can have up to hundreds of thousands, or even millions of neural layers. Can we even picture this? Hundreds of thousands of back-to-back layers, with countless computations taking place at the same time on every one. Only the result is visible, and it astounds us: Cat recognized!

Generative AI



Even more astounding are the possibilities that were opened up by AI's next step of development – our tower's crowning block: generative AI. Ever since Open AI made ChatGPT publicly available in November 2022, it has been joined by a growing number of systems. Claude, Imagen, Sora and other tools can generate texts, images and videos, all in the blink of an eye.

But what exactly sets this type of AI apart from the other areas we have already looked at? The leading AI expert at The New York Times, Zach Seward, summarized it pretty succinctly: "If traditional machine learning is good at finding patterns in a mess of data, you might say that generative AI's superpower is creating patterns," he wrote (2023).

An AI application is typically designed to perform a specific task. This is the case for the therapy chatbot ELIZA as well as for AI that can recognize cats, and it also applies to generative AI. In such applications, large language models are trained to perform a very complex and far-reaching task: the processing of natural language.

Hence why they are called language models. Their statistical model is trained to predict the probability of word components and sentences, such as what is needed to complete the sequence “Twinkle twinkle little star”

Transformer

This is made possible at the current level of quality and speed by what is known as a transformer. ChatGPT even references the transformer in its name: GPT stands for generative pre-trained transformer – and that’s precisely what it is. “Transformer” refers to the type of learning algorithm. “Pre-trained” describes that the model needs to be trained before it can be put into action, though it can always be improved later. And “generative” describes the type of output – something newly created.

In technical terms, a transformer is a specially built deep learning architecture that is part of an artificial neural network. First presented at an industry conference in 2017, it has since revolutionized the processing of natural language. In comparison to numerical data, processing language presents a unique challenge since the same words, spelled the same way, can have different meanings depending on where in a sentence they are located. Put briefly, they are contextually dependent.

In order to process a text, the words must first be converted into numbers. Transformers are the first models that can numerically mark the positions of the words in a sentence, thus better grasping the context. Transformers also make it possible for the first time for algorithms to learn which pieces of inputted data are most relevant for a certain output (see Klofat 2023 and Clarke et al. 2023).

In the case of ChatGPT, we are dealing with a very large language model tasked with processing and generating language. Other systems operate similarly (see page 22).

OpenAI’s chatbot, ChatGPT, is based on models GPT 3.5, GPT 4 or GPT 4o, depending on user selection. These are referred to as foundation models because they can be further trained to execute different tasks.

So are large language models (LLM) and foundation models one and the same? Not exactly. The term foundation model does encompass large language models, but it also includes models that produce images or videos instead of text. Whereas language models are trained exclusively with texts, other foundation models, such as OpenAI’s DALL-E, are also trained with image data or, as is the case with Sora, with video data, too. Such models are therefore also able to generate images, audio or videos.

Strong and weak AI

Even if these systems seem extremely strong and powerful, such applications are nonetheless still considered weak AI, also known as narrow artificial intelligence. Weak AI is designed to perform a specific task that previously could only be executed by humans, such as generating texts or images. In contrast, computer science researchers use the term strong AI to refer to an application that can perform all the tasks a human is capable of performing. It is often colloquially referred to as artificial general intelligence (AGI), or super intelligence. Meta CEO Mark Zuckerberg has announced that work on super intelligence is currently underway. Independent scientists, however, believe the development of such applications remains a long way off (see Stieler/Zweig 2024).

Known AI models

Artificial intelligence company OpenAI was founded in 2015. It is behind **ChatGPT**, a text tool, as well as **DALL-E**, which generates images from written descriptions. These will soon be joined by **Sora**, a tool that generates videos from text prompts. ChatGPT 3.5 is a free version, while ChatGPT 4 is available through the ChatGPT Plus subscription plan. The results provided by the free version are markedly less eloquent and more error-prone than those from ChatGPT 4. A third model, ChatGPT 4o, has been available since May 2024. It can process text, video and audio.

Software provider Microsoft has invested more than \$13 billion in OpenAI since 2019, thereby securing access to its technology. Microsoft's current AI product, **Copilot**, consequently also utilizes ChatGPT. Copilot can be opened directly in Microsoft programs such as Word, Excel or PowerPoint. Microsoft also cooperates with Meta. In addition, Microsoft has been working on its own AI language model called **MAI-1**, which is intended to compete with the Google and OpenAI models.

In February 2023, **Facebook parent company Meta Platforms** released its language model, **Llama**. The acronym stands for Large Language Model Meta AI. The code used by the model to execute requests is publicly available. The follow-up model, Llama 2, was released in cooperation with Microsoft in July 2023. It, too, is open source. In addition to text and images, Llama 2 can also generate code. Meta introduced LLaMA 3 in April 2024.

In response to the release of ChatGPT, **Google** released its chatbot, **Bard**, in February 2023. In February 2024, the company changed the tool's name to **Gemini**. The models that make up the chatbot vary in size and capability, and some have fees. The current version can process various types of content such as text, images, audio and video. Overall, the functions are very similar to those of OpenAI. Gemini is also available as an app.

Anthropic was founded in 2021 by former OpenAI employees. The company is dedicated to developing safe and understandable AI systems. Its language model, **Claude**, can generate text and has a strong focus on AI security. Version Claude 3, which was released in March 2024, consists of three models that are tailored to different types of customers and uses. For a long time, Claude was only available in the US and the UK, but it has since been made available in the EU as well. While it primarily targets companies, Claude is also available for individual and general usage. In June 2024, Anthropic introduced Claude 3.5 Sonnet, a language model that can analyze images as well as texts.

Midjourney is an independent research lab in San Francisco. It was founded in July 2022 and specializes in generating images from written descriptions. The company's AI tool, also called Midjourney, is available via chatbot on the platform Discord for a fee. It allows users to generate images from written instructions that they input. It sets itself apart by being able to deliver specific artistic styles and visual interpretations, and produce highly photorealistic images.

Artificial intelligence company Stability AI, founded by Emad Mostaque in London in 2019, also primarily focuses on developing image-generating AI programs. Its flagship tool is **Stable Diffusion**, whose publicly available code was developed by a team at the University of Munich. Stability AI differs from DALL-E and Midjourney through its use of open-source code, which provides both programmers and businesses with easy access to it and lets them tailor the model to their specific needs. The tool can now also partially produce short video sequences and animated images.

The **German company Aleph Alpha** has been working on the large language model **Luminous** since 2019. It is currently still in development and not yet available to the broader public; instead, it is aimed primarily at businesses and organizations. Aleph Alpha's goal is to raise the security and application potential of AI in the European context by guaranteeing to abide by data protection regulations and offering custom solutions for local markets. Aleph Alpha also wants to make sure users can understand how the AI tool Luminous reaches its decisions. Its first customers include Bosch, SAP and numerous local governments.

Just over one year old, the **French startup Mistral** is currently the European AI scene's biggest hope. It was founded in April 2023 by three software developers: Guillaume Lample and Timothée Lacroix are former Meta employees who helped with the development of Llama while Arthur Mensch worked on Deepmind, the AI subsidiary of Google. Mistral is based in Paris, but most of its investment now comes from the US. Its investors include venture capital firm Andreessen Horowitz, software company Salesforce and Microsoft. Mistral's AI product is called **Le Chat**. The beta version is currently available for use upon registration. Similar to Aleph Alpha, Mistral principally targets companies and organizations. Its customers include the bank BNP Paribas and mobile service provider Orange.



Using AI Responsibly

In late 2023, the American startup Channel 1 released a video announcing its plan to enter the media market in 2024. It wants to be one of the first news broadcasters worldwide to exclusively use AI in its production processes. Humans will deliver the material and fact check; AI will take care of the rest. In the video, an artificially generated avatar explains what this means: “Everything you’ll see on Channel 1 ... uses AI to give you news the way you want it: personalized, localized and distilled.” The avatar then says, “Technology may be the tool, but journalism’s core values of integrity and accountability are at the heart of everything we do.”

Is this the future of journalism? Is this ethical – entirely, partially, or not at all? Is this the future we want?

Using AI responsibly and surviving in an AI work environment requires knowledge at many different levels. In addition to understanding how the technology works, one must also be familiar with the legal framework and media ethics. It also helps to understand how AI applications can reproduce biases and prejudices. And while it has largely yet to be discussed, the question of the environmental impact caused by AI’s training and use is another aspect we believe is important for the responsible management of this technology.

Legal issues

When it comes to AI, there are currently more unanswered legal questions than answered ones. Lawyers often focus on what still needs to be sorted out. However, currently existing laws already reveal some basic principles and trends. Below we provide a brief overview of some important aspects.

AI regulation

In March 2024, the European Parliament passed the world's first multinational AI law. Called the AI Act, it went into effect in all EU member states. However, its initial impact on day-to-day journalistic work will be negligible. For one, the law outlines various transitional periods that give everyone time to prepare to comply with the new rules. Furthermore, individual journalists and editors will likely have little contact with the EU regulation even once the transition periods end. This is because the law primarily targets developers, producers and providers of AI systems.

The law's regulatory approach centers on an AI risk classification system in which certain security precautions must be taken depending on the risk class. For journalism, questions regarding copyright law and data protection are currently more important.

Data protection and “media privilege”

Questions relating to data protection arise as soon as one enters a prompt containing personal data. Data protection may also be an issue when AI is used in other day-to-day journalistic activities. The General Data Protection Regulation (GSGVO) protects personal data, or all information relating to a person that can be linked directly to their name or associated with them. Individuals who collect or process such data therefore have various legal responsibilities.

However, an interesting exception may apply in journalism. Known as “media privilege,” the terms refers to the legal understanding that data protection requirements largely do not apply to journalistic work, from research to publication. The reason? Without media privilege, unrestricted media coverage would be jeopardized or even impossible in certain instances. One example is investigative research, if data protection made it necessary to obtain consent from the person concerned.

Media privilege is conditional, however, and it only applies when personal data is used or processed for journalistic purposes. And despite this exception for data processing, other applicable rules such as privacy law must still be upheld. Media privilege also does not apply beyond actual journalistic work. For instance, it does not cover the distribution of content, such as sending a newsletter that contains personalized offers.

In short, as long as media privilege can be invoked when researching and creating content, data protection should not pose a fundamental challenge to journalists wanting to use AI in their daily work.

It remains to be seen how this exception will apply in the future once AI is used more extensively in journalism. Presumably, this will also depend on the extent to which AI's use actually serves journalistic purposes.

Copyright and the question of human creativity

Whoever holds the copyright to a piece of work has the right decide if, where and how it is published, and at what price. Generative AI programs have created new challenges with respect to this. A later chapter, "Training AI models," will look at how journalistic content is used to train AI. But the practical application also raises new questions: Is it possible to violate copyright by publishing AI-generated content? And what about a case where AI helped produce content? Is such content still copyright protected to the benefit of the creator?

Such questions revolve around two legal issues: Was it created by a human, and is it really creative? Only if both questions can be answered affirmatively does copyright law actually apply.

Since copyright only protects human creations, content like texts, photos or videos that have been automatically and autonomously generated by AI are not protected by copyright. However, if the machine spits out a specific text or photo bearing little changes from the copyright-protected original, then the copyright remains with the person who created the original work and royalties are therefore due. This situation occurs relatively infrequently, but it is nevertheless important to remain aware and discerning.

The following practical question will become increasingly relevant: If I create a piece of content using AI, can I register it with a copyright collective or collecting society? This is only possible if you are the creator.

If AI was used solely as a tool or assistant, or for support during the creative process, and the actual creative work was carried out by a human, then the content can, in principle, be protected by copyright.

But there is also a second stipulation – regardless of whether AI comes into play. It's referred to as the “threshold of originality,” a term coined by lawyers. This is present when a text, graphic, video or another piece of content is deemed sufficiently unique and creative. A piece of news that merely repeats facts without any creative elements does not fulfill this criterion.

But when is the threshold of originality met so that copyright actually applies? There is no precise definition, as the threshold is fluid and variable. It can only be determined in individual cases relating to specific journalistic content. This being said, there is a general rule: The more a human has intellectually and creatively shaped the content, the more likely it is to be protected by copyright.

Journalistic diligence and transparency

“AI journalism goes awry when it’s unchecked, lazy, selfish, dishonest and opaque,” Zach Seward, the leading AI expert at The New York Times, wrote in an essay (2024). So what basic framework is necessary for AI to be used responsibly in journalism? “It’s got to be vetted, in a rigorous way,” Seward continued. “The idea should be motivated by what’s best for readers.” And more than anything else, he added, the basic principles of journalism must apply.

The two key principles in this context are journalistic due diligence and transparency. Even though the term due diligence sounds rather open-ended and undefined, violating it can have legal consequences. However, if it can be proved that due diligence was upheld, cease and desist orders, damage compensation claims or even criminal prosecution will be dismissed.

As the custodian of ethical standards in Germany, the German Press Council has been examining this topic. It wrote in its 2023 annual report: “The press code already requires newsrooms and desks to uphold professional ethical standards regardless of whether a text has been written by a human or by a machine” (German Press Council 2024). For this reason, and in light of current AI developments, the report argues that there is currently no need to adjust the press code to, for instance, make it mandatory to label AI content. Nevertheless, the report adds that AI-generated content should not mislead readers.

It also points out how labels are already required for AI-generated images. Images that have been generated by AI fall under the rule governing stock photos and must therefore be labeled as such. In contrast, the Swiss Press Council (see 2024) has already expanded its code of conduct to include guidelines on artificial intelligence.

One media organization at the forefront of AI is the Cologne-based publishing house M. DuMont Schauberg. In summer 2023, word got around that it had recently published texts by a certain Klara Indernach, or KI – the German abbreviation for artificial intelligence (Künstliche Intelligenz). “The Express [a DuMont regional tabloid] acted in part as if KI were a human,” read a September 2023 editorial in the German daily taz. Ethical or not? Navigating issues of transparency will prove tricky in many cases. What should be declared? Where? Is it enough for there to be a notice somewhere on the website stating that AI has been used to produce texts (similar to how author bios are available by clicking away from an article) rather than on the AI-generated text itself?

In the case of Klara Indernach, the Express was reprimanded by the German Press Council – but not for how it created the text, nor for where it placed the related transparency notice. Instead, it was for failing to indicate that the photo of the supposed author had been generated by Midjourney.

In 2023, the German Press Council issued its first ever reprimand in relation to an AI-generated text. The German weekly magazine Die Aktuelle had published what they presented as an interview with the former Formula One racing driver Michael Schumacher, describing it on the cover as his “first interview” since his tragic skiing accident in 2013. The interview was identified as AI-generated toward the end of the text, but this was deemed insufficient. The Press Council described it as having “seriously mislead readers” and considered it a possible violation of Schumacher’s personal rights.

Aside from its landmark calls on missing image labels and the Schumacher interview, the German Press Council has issued no further statements. A spokesperson also said there had been no other noteworthy complaints through May 2024. A working group is currently discussing how to proceed in the future. Swiss journalist and communications researcher Colin Porlezza (2024) considers this response too tentative and points to guidelines that have been developed by the Council of Europe.

In “Guidelines on the Responsible Implementation of Artificial Intelligence Systems in Journalism” (see Council of Europe 2023), the Council of Europe highlights the opportunities and potential for innovation that AI offers journalism while simultaneously providing recommendations on how media professionals, organizations, technology providers and national governments should deal with the issue. The guidelines include a reminder that the right to freedom of expression, enshrined in Article 10 of the European Convention on Human Rights, comes with responsibilities toward society. “Implementing and using journalistic AI systems in accordance with values is a difficult task, and there are often no ready-made answers as the true challenge sits in the operationalisation,” the guidelines state.

Internal guidelines

So how should news desks and departments use AI? And what is considered unethical? The latest AI tools have made these questions even more loaded and relevant. This is why the international non-governmental organization Reporters Without Borders, which promotes and defends press freedom, and 16 partner organizations published the “Paris Charter on AI and Journalism” on November 10, 2023.

The Paris Charter (ROG 2023) views artificial intelligence as a tool with an important and unique characteristic – the ability to make its own decisions, thereby steering public attention. Use of AI must therefore be constantly questioned, monitored and made transparent, the charter argues. As Nobel Peace Prize winner and charter commission chairperson Maria Ressa said at the document’s signing, “Technological innovation doesn’t inherently lead to progress: It must be steered by ethics to truly benefit humanity.”

The Charter lays out ten principles that should govern the use of artificial intelligence in journalism; the principles are aimed at media outlets and journalists alike. For instance, it calls on media outlets to “prioritize human agency” and “ensure content origin and traceability.” Furthermore, “Media outlets are always accountable for the content they publish.”

The Charter also outlines how both media outlets and individual journalists must “[draw] a clean line between authentic and synthetic content” and “should play an active role in the governance of AI systems.” Their actions must always be guided by journalism’s main mission – to ensure the right of all people to high-quality, trustworthy information.

Reporters Without Border’s Paris Charter was not the first of its kind; many media organizations had already created their own codes of conduct. One of the first German outlets to do so was Bayerischer Rundfunk, the regional public broadcaster for the state of Bavaria. It issued its own AI guidelines in November 2020 (see BR 2020). Many other outlets followed suit starting in 2023. A global comparative study from September 2023 (Becker/Simon/Crum 2023) shows that these organizations share a similar understanding of some key points related to the responsible usage of AI.

For example, they all agree on transparency, responsibility for AI and the need to monitor AI-generated content and decisions. The media organizations examined in the study all also explicitly stated that they do not want to replace journalists with AI. Some of them additionally highlighted how AI can reproduce racist, sexist or other dehumanizing data inputs.

However, the study authors also pointed to blind spots in the guidelines. For instance, they criticized some for lacking concrete details, such as failing to specify whether self-imposed identification requirements apply only to texts and images that have been entirely generated by AI or also to titles and teasers for which AI provided assistance. Problems like the exploitative labor practices around AI data curation and training and the environmental damage caused by AI servers' enormous use of water were also missing from all guidelines, the study authors said.

According to a study by the Media Authority of North Rhine-Westphalia from March 2024 (see Nennstiel/Isenberg 2024), a majority of people would welcome an independent regulatory authority and stricter, binding rules for the use of AI in journalism. Some respondents even categorically objected to any use of AI in journalism. Particular skepticism exists when it comes to using AI in political reporting.

According to the comparative study by Becker, Simon and Crum, almost no media outlets are currently planning to involve audiences in the development of AI journalism guidelines, even though this could be important for maintaining an organization's brand credibility and trust in its reporting.

Tips for creating an AI code of conduct

FRAMEWORK

Which parts of a company should an AI code of conduct apply to? It is important to define the areas where an AI code applies. At present, such codes exclusively address AI's impact on the work of journalists. Yet going forward, AI will also be used in many media organizations' PR and communications, HR and distribution departments. Companies should therefore consider which code rules should additionally apply to these areas.

MODEL CODES

What rules have been set by similar companies? The responsible parties should begin by getting an overview of the status quo. The more individual codes are modeled on each other, the more the industry will move toward unified standards. The Paris Charter is a good starting point. However, as the field of AI is developing rapidly, it is important to stay up-to-date. The Poynter Institute has created a template that it regularly updates (see McBride 2024).

RISK MANAGEMENT

Where are the ethical challenges in the workflow? The need for an AI-specific code of conduct stems from the fact that artificial intelligence can negatively impact journalistic work. Media outlets that want to create a strong AI code should therefore systematically identify possible negative impacts and develop counterstrategies.

INVOLVING THE PUBLIC

What do users think? The general public has specific demands regarding the use of AI in journalism. Media organizations should therefore create a platform that the public can use to participate in the drafting of an AI code. This also helps create transparency, which is essential for media organizations' credibility and public trust in them.

REVIEW AND FURTHER DEVELOPMENT

Are our guidelines still up-to-date? AI is developing at breakneck speed, which means AI's current capabilities will soon be a thing of the past. Codices should therefore be updated. The German Journalists Union recommends that media companies appoint officers to compare in-house use of AI with the applicable internal and external rules. These individuals could also serve as the first point of contact for complaints (DJV 2023). It is also advisable to regularly check and see whether any internal rules or parts thereof need to be revised.

Training AI models

Biases

In order to responsibly use AI systems, we believe it is important to understand how AI tools may convey a one-sided, biased perspective. Although learning algorithm structures and the level of detail in their data processing have steadily grown more advanced, one aspect of machine learning has always remained the same: Algorithms analyze large to gigantic amounts of data during their learning process. This learning phase is highly relevant, for it is where AI's fundamental susceptibility to possible biases can be found.

Learned biases can result from training data as well as human feedback provided during the machine's learning process. All humans have a perspective on the world that gets reflected in how data is selected or what feedback is given to software during the learning process, and this reflects existing societal conditions.

Online content plays an incredibly important role in the compilation of training datasets. Yet even experts do not know what precise content is included in training data (see Burgess/Rogers 2024 and Bender et al. 2021). The only thing that is clear is that gigantic amounts of data are involved (see Bender et al.). One well-known dataset is called Common Crawl. According to the crawling service, the corpus consists of "petabytes of data" that have been gathered since 2008 and includes "raw web page data, metadata extracts, and text extracts" (Common Crawl 2024). This data in an adjusted form was also used to train ChatGPT.

But who primarily produces online content? It tends to be people who are well-off rather than those whose daily lives revolve around existential questions of survival. A very strong critique of how training data is handled comes from Bender et al. (2021). They argue that the current methods entrench hegemonic perspectives and the marginalization of underrepresented populations in the training data.

They also criticize how datasets are digitally curated. One example is how the practice of suppressing all content with words related to sex had unintended exclusionary effects. Although the primary aim was to filter out obscene and pornographic content, this overgeneralized method of cleaning up data also removed large amounts of content from the LGBTQ community, and consequently their perspectives.

The authors also criticize how the datasets used to train language models are growing increasingly large even as there is no documentation of what data is included or filtered out. They describe how one-time generated datasets are often used over and over again, thereby entrenching injustices and marginalization without any form of review.

One organization working for greater transparency in the development of AI foundation models is Stanford University's Center for Research on Foundation Models (see CRFM 2023). Researchers there have created a transparency index that defines standards of openness and accessibility in the development of AI foundation models. The AI Act also lays out requirements that technology developers must comply with by 2026.

The fight over copyright

The lack of transparency in training data also raises copyright and ownership questions, since much of the content that comprises training data is copyright protected even though it is available for free online. Yet data companies crawling the internet's vast expanses have not shown great interest in this issue. And for a long time, the creators of said content were barely aware of what was going on.

Currently, the following general rule applies: If a rights-holder does not have a notice on their website denying consent for their content to be so used, then the practice is generally allowed. However, should a website operator explicitly opt out, then machines are not allowed to digitally search and replicate the content. The exact process is described by Burgess and Rogers (2024) in the digital magazine *Wired*.

Media companies have dealt with this situation very differently. For instance, the German publishing house Axel Springer and the American news agency Associated Press have signed contracts with OpenAI allowing their content to be used for a fee. Others, however, such as The New York Times, take the opposite view: The media giant sees ChatGPT as a threat to their business model and has sued them for billions in damages.

Anyone who is interested in finding out whether content from their media company or personal website has been used for training purposes can do at the website "Have I been trained?" (<https://haveibeentrained.com/>), from the company Spawning. Just a quick search shows that content from many German media outlets can be found in AI training data. The website also makes it possible to opt out for individual pieces of content.

Environmental impacts

AI consumes large amounts of water, electricity and other valuable resources. Therefore, the responsible use of AI should also take ecological sustainability into account. Unlike the human brain, the current AI computer systems are far from energy and resource efficient. Whereas our brain consumes about 20 watts of power, roughly the equivalent of a laptop, football field-sized data centers are being built for AI systems, and each one uses the same amount of electricity as a large city. And let's not forget that AI is still a long way off from being able to replicate the abilities of the human brain.

How much electricity and other resources does AI already use? How will this change in the future? And what options exist for natural resources to be used as sparingly as possible in the age of AI? Large tech companies only reluctantly publish information on the energy and resource requirements of their AI systems, meaning we still lack precise scientific data on the ecological impacts of AI usage. But even without concrete numbers from companies, it is still possible to make fact-based estimations about resource consumption. Researchers can do this by breaking down the overall resource consumption and differentiating between model building, training and application.

Annual electricity consumption exceeds that of Denmark

According to current knowledge, most of the electricity consumed by AI, some 80%, takes place when AI is being used, which is referred to as inference. The rest is consumed during model building and training. Publicly accessible models, like ChatGPT, use a particularly large amount of electricity during inference. Experts estimate that just operating ChatGPT 3, which launched in November 2022, requires more than 500,000 kilowatt-hours of electricity per day, which is enough to power some 150 households for a year.

Things look different for AI applications that are not intended for the general public, such as those used in industrial production for the early identification of errors or risks. In such cases there are relatively few users. As a result, more resources are required during the training phase than when the application is being used.

Current estimations place the global annual electricity consumption of all AI applications in data centers at nearly 40 billion kilowatt-hours (see Avelar et al. 2023). This is more electricity than Denmark consumes over the same period, and it doesn't even include the energy consumed by the end devices needed to use AI.

Energy consumption to grow significantly

As opportunities to use AI increase, its usage-based electricity needs are also expected to grow significantly. According to the market research company Gartner, AI could account for 3.5% of global electricity use by 2030 (see Gartner 2022). Using current electricity use projections, this would be more than 1 trillion kilowatt-hours – or twice as much electricity as Germany currently consumes on an annual basis.

Optimists hope that improved models and hardware designed for particular use cases will be able to keep growing electricity needs in check. But realistically, one must consider that past improvements in IT efficiency have only led to greater use of IT. It is very unlikely that AI will differ in this respect. After all, there are no limits as to how it can be used, and we're only at the beginning of its development.

Model size determines resource needs

AI's ever-increasing resource needs can also be seen by looking at the electricity it consumes during training. In 2019, US researchers calculated that training a then-current AI language model required 656,000 kilowatt-hours of electricity (Strubell et al. 2019). In 2020, ChatGPT 3 consumed nearly twice as much electricity during its training, just under 1.3 million kilowatt-hours (see Patterson et al. 2022). With ChatGPT 4, this grew nearly sixfold to an estimated 7.2 million kilowatt-hours (see TRG Datacenters 2023). This is enough electricity to power 1,800 single family homes for one year.

In addition to electricity, AI also needs vast amounts of water. Researchers at the University of California estimate that the training of ChatGPT 3 alone required around 5.4 million liters of water. This is equivalent to the daily water needs of more than 40,000 people in Germany. Moreover, an additional 1.6 liters of water are needed for every hundred requests made to ChatGPT. Most of the water is needed to generate conventional electricity and cool the data centers' servers.

Hardware manufacturing uses large amounts of resources

AI hardware components use resources such as gold and silver, and other rare elements. In addition, manufacturing computers and servers requires significant amounts of energy, and it releases greenhouse gases. Life cycle analyses of current servers show that 10-40% of their total energy requirements and greenhouse gas emissions can be traced back to production.

And there will only be more hardware going forward. Analysts estimate that between 2023 and 2033, the annual volume of AI hardware sales will increase ninefold to nearly \$500 billion (Precedence Research 2024). One impressive example of this growth is Nvidia. The company produces high-performance computer chips, and in early 2024, production could barely keep up with demand. Due to this sharp rise in demand, the chip manufacturer has become the fifth-most valuable company in the world, behind Apple, Microsoft, Alphabet (Google) and Amazon.

Ideas for greater sustainability in AI

One commonly heard argument is that renewable electricity could be used to power AI systems, thereby keeping environmental impacts in check. But this argument has its flaws. Using renewables can, in fact, minimize the technology's negative impact when it comes to greenhouse gas emissions. However, one must also consider whether there is enough renewable energy for this, or whether AI could end up using climate-friendly electricity that would otherwise power electric cars, heat pumps or industry. AI data centers can therefore only be climate-friendly when their electricity comes from renewable energy power plants that have been newly built.

But even data centers powered exclusively by green electricity cannot neutralize AI's environmental impacts; the large amount of water and materials needed for hardware production and by system facilities also take a large environmental toll.

One possibility could be to make the models more efficient, such as by training them with less, but more high-quality data. There is little difference here between man and machine. The worse the data – the educational material – the less is learned. Using AI purposefully can also help save resources.



Using AI Constructively

“This is Wired, so we want to be on the front lines of new technology, but also to be ethical and appropriately circumspect,” the editorial team of the tech magazine wrote about how they use generative AI in their daily work. To uphold these standards, they laid out some ground rules. Firstly, “We do not publish stories with text generated by AI.” And secondly, “We do not publish text edited by AI either.” In addition to pointing out how AI is prone to errors, the editorial team argued that the process of writing and editing an article always includes determining what the most important, original or entertaining aspect of a piece is – a decision they don’t want to outsource to a machine.

The team’s approach captures a common feeling of our time. According to the Reuters Digital News Report 2024, which surveyed media uses in different countries, around half of respondents were uncomfortable with the idea of content that had been mostly generated by AI; another 25% was still unsure. “Our qualitative research on AI suggests that trust will be a key issue going forward,” the study authors concluded (Newman et al. 2024, p. 20).

Building trust takes a lot of time, but you can lose it in the blink of an eye. Currently, fewer than half of respondents say they trust most news most of the time, the Reuters Digital News Report also found (ibid.). It seems the Wired team has good reasons for their cautious approach. So what does Wired actually use AI for? Brainstorming story ideas, for instance. In short, AI should not replace fundamental journalistic work, but should instead help journalists do their work better. On the following page, we present some ideas that clearly show how this can work on a day-to-day basis.

Constructive prompt engineering

ChatGPT's release gave the general public a whole new set of terms: Prompt, prompting, prompt engineering. These terms rapidly made their way into many non-English languages. They all stem from the verb "prompt," which Merriam-Webster dictionary defines as "to move into action," "to assist (one acting or reciting) by suggesting or saying the next words of something forgotten or imperfectly learned," and "to serve as the inciting cause of." The word "prompt" refers to the text that we, the users, input into ChatGPT, Copilot or Gemini, spurring the AI system to do something.

The ability to design a prompt that achieves optimal results is known as prompt engineering. Two main factors determine what a chatbot "can do": First, there's the question of whether something can actually be requested. And then there's the issue of how good the prompt is. So what makes a prompt good?

Limits: What AI can't do

Artificial intelligence calculates statistical probability; it is not based on truth or reality. This is why generative AI can hallucinate, or provide answers that may seem logical or appropriate but are actually provably false. The system provides an answer not because it "knows" something, but because the output text has the greatest probability of being the corresponding answer to our input. AI systems neither "understand" things, nor can they determine what is right or wrong, true or false.

Hallucinations are not a minor issue but rather a logical consequence of the way AI works, which means language models are currently unsuited for producing journalistic texts intended for publication. We do not believe AI is cut out to replace good journalistic texts, and we therefore consider it wrong to use it so. This being said, AI can be a helpful tool for journalists in many ways, and it can also improve our reporting.

Basics: Elements of a good prompt

If you want to create effective prompts, there are a few basic rules to keep in mind. It may help to use a basic framework that you can adjust according to an individual task and the desired result. This framework consists of a role, a task, criteria for the desired answer, the goal of the prompt and the desired output format.

Basic prompt framework

ROLE Let's assume AI were a person. How would it talk? How would it pass on information? This is the role of the AI model, and it should shape the prompts.

You are a journalist at an online publication where you regularly write constructive articles on political and societal topics.

TASK What problem should the AI tool solve/what task should it complete? Now you should describe in as much detail as possible what AI should do.

Summarize how different individuals and groups view [a topic]. Describe their opinions and beliefs. Also include minority individuals who have been affected.

CRITERIA What should a good answer provide? Should it be first and foremost accurate, creative or something else? It is important that the prompt includes clearly defined criteria for the desired answer.

Your overview should take into account all relevant groups of individuals: politicians, residents, activists, affected minorities, and so on; all genders and their perspectives should also be included.

GOAL How can the goal of the prompt be defined? The more clearly the overarching goal is worded, the more effective the prompt will be.

The goal is to use the information that has been collected to identify interview partners and write a journalistic article.

FORMAT What format should the machine use for its answers? A list, full paragraphs, dialogue? This should be specified in the prompt.

Create a table showing the different groups of individuals and their respective opinions and beliefs.

Level of detail

The prompt's level of detail is also important: If it is too general, the system may interpret the input very differently than intended. On the other hand, if it is too detailed, the answers may end up being so specific that they are not useful. It is important to find the right balance between too few details and too many. The rule of thumb is: as short as possible, as specific as necessary.

Refinement

Once you are familiar with the elements of a perfect prompt, you will be able to tease the best results out of AI. Yet it still may be that you are not 100% satisfied with an answer. This is where refinement begins, or a process of making adjustments. Just like you can talk with a human about a task they have completed, you can also give AI feedback on its result.

**GIVING CONSTRUCTIVE
FEEDBACK**

I really like the [third suggestion]. Please expand on it.

**ASKING FOR VARIATIONS
ON AN ANSWER**

I like this result, but I would like to include other affected minorities. Can you please add some more?

**ASKING FOR
ADDITIONAL
INFORMATION**

Can you please add another column to the table in which you explain how the groups of individuals developed their opinions and beliefs?

Gender in prompts

Many journalists working in languages other than English may have to deal with gendered nouns. This grammatical element adds yet another dimension to constructive prompting.

The question of whether to specify gender can arise in various situations, such as when assigning a role to the AI tool. As we have already learned, it is very important that prompts be formulated as precisely as possible to achieve good results. Gender therefore has the potential to confuse the AI tool, especially if the masculine and feminine forms of multiple terms, such as professions, are listed one after the other, creating a long, unwieldy prompt. This could cause the tool to focus more on the gender aspect instead of actually solving the desired task.

Until this issue has been rigorously analyzed, we recommend avoiding such prompts and instead using alternative phrasing where possible.

With respect to role designation, the most precise, and therefore safest way to go about this is to assign the AI tool a definitive role that specifies the gender e.g. “You are a [implied male] journalist” or “You are a [implied female] journalist.”

Advanced prompting

In addition to a well-constructed prompt framework, there are other factors that can influence the quality and relevance of an AI answer. Known as parameters, these settings make it possible to steer the writing style of an AI answer in a specific direction. These parameters are particularly prominent in ChatGPT, but they can also be used in a modified form with other text-based AI tools. You can apply parameters by adding them to the end of a prompt in the form of a short command. This generally looks like the following: [Parameter]=[0.2]. Make sure to use periods rather than commas between numerals to avoid any problems. We’re now going to introduce two parameters that are relevant when using AI for journalism and that can influence the results in interesting ways.

Temperature

Should the result be dry and straightforward or original? The temperature parameter plays an important role in how AI tools generate answers. It affects how conservative, matter-of-fact or creative the answers will be. Since every AI answer is a probability calculation, the temperature determines whether the tool will select an answer that is more probable or less probable.

The temperature is defined by the value X in the request “temperature=X” where X can be between 0 and 1. Experimenting with different temperature values lets you find the optimal balance between dry and creative answers.

If a low temperature value is set, such as 0.2., then the AI tool will generate conservative answers. It will favor recognized phrases and formulate sentences in a more formal style. This can be helpful for media professionals who are trying to generate clear and precise answers that help them with their research.

A higher temperature value, such as 0.9, tells AI that its answers should be highly creative and experimental. The results will therefore be more original, but also more unpredictable. Media professionals can use this setting to come up with creative ideas.

Diversity penalty

Another important parameter that affects the quality of generated answers is the diversity penalty. This parameter controls the word variation of an answer by influencing word repetition. A higher value punishes tokens (words or characters) that have already been used in an answer, leading to more diverse and varied results.

The diversity penalty’s value ranges from 0 to 2. A higher value leads to more diverse answers. The more creative an article is supposed to be, the higher the diversity penalty’s value should be.

Just like with temperature, a value X is chosen and provided in the prompt: “diversity_penalty=X”

The diversity penalty is most interesting in longer texts, but differences can also be seen in shorter texts, too.

Let’s imagine ChatGPT 4 were to explain what constructive journalism is. If a low value, like 0.2, were to be inputted, the AI tool would tend to repeat the same words or phrases; a higher value would result in more variation.

Explain to me in three sentences what constructive journalism is. diversity_penalty=0

Answer:

Constructive journalism is a style of news reporting that focuses on solution-oriented angles on current issues, aiming to present stories that inspire and empower audiences. It seeks to broaden news perspectives by not only reporting problems but also highlighting potential solutions and responses. This approach encourages public discourse that is more informed and constructive, potentially reducing negativity in media consumption.

Explain to me in three sentences what constructive journalism is. diversity_penalty=2

Answer:

Constructive journalism goes beyond just reporting negative events by also exploring solutions and positive developments to foster a more constructive public dialogue. It aims to engage and empower readers by showcasing how individuals and communities are addressing challenges, emphasizing stories of resilience and innovation. By blending rigorous reporting with an optimistic outlook, it strives to inspire action and contribute to societal progress.

But what makes a prompt constructive?

Prompts can play a significant role in making media coverage more constructive overall. This starts with how journalists approach a topic and the extent to which personal experiences or cognitive biases shape their views of an event or a situation. Generative AI can be a source of balance, spurring reflection and helping achieve a more holistic and realistic view of the world.

From finding topics and angles to analyzing documents, from including different views to coming up with better interview questions, purposeful prompting can help incorporate solutions, diverse perspectives and constructive dialogue – the three main elements of constructive journalism – into everyday work and make sure coverage does a better job meeting people’s needs. We’ll show you exactly how this works on the following pages, where we present selected use cases. Of course, there are also countless other interesting ways to use AI tools.

We use ChatGPT in our examples for two primary reasons: We assume this is the most commonly used AI tool among our readers, since it was one of the first major publicly available generative AI systems and is consequently the most well-known. Additionally, when the content that you are reading here was being produced, ChatGPT had the largest range of functions among comparable tools with respect to text analyses and generation (see “Known AI models” p. 22).

Prompts for brainstorming constructive topics

Constructive journalism goes beyond problems and instead also presents possible solutions, thereby making a valuable contribution to public debate. One of its main challenges is brainstorming ideas for constructive coverage. How can media professionals identify topics and approaches that are constructive, solutions-oriented and relevant to their target audiences? Below, we present steps and methods that can help with brainstorming constructive topics.

Selecting a topic

One of the first steps in brainstorming is choosing a subject area. Asking AI directly about possible constructive topics can provide some initial inspiration:

I'm looking for topics that are suited to constructive journalism. Give me ten general topics that I could work on as a journalist.

If the subject area has already been determined, this information should be shared with the AI program so the search for concrete topic ideas can start one step further along. For instance, if the subject area is environmental conservation and climate protection, then the prompt could look like the following:

I'm looking for topics related to environmental conservation and climate protection that are suited to constructive journalism. Give me ten possible topics within this overarching subject area that I could work on as a journalist.

The AI-generated topics can provide a basis for a wide range of possible stories that both inform and inspire.

Diving deeper into a topic

Once a topic has been selected, it is now time to identify specific aspects that are particularly interesting and relevant for a constructive approach. The focus here is on possible solutions, multiple perspectives and constructive dialogue.

What specific aspects can I highlight when constructively covering [topic] to make sure I present different perspectives and possible solutions?

This can help media professionals dive deeper into a topic, collect additional information and identify other directions for research and coverage. Such insights make it possible to purposefully structure coverage around specific aspects.

Once the topic has been selected and specified more concretely, it is time to determine in detail what a constructive approach could look like. This involves incorporating as many different perspectives and views as possible. The next chapter provides guidance on how to do this.

Prompting for many different perspectives

Among other things, constructive journalism focuses on the well-being of society and the needs of people. But media professionals don't always have direct access to every social group. AI can help them get a sense of these groups and put themselves in their shoes, at least a little. Some possible related AI prompts could be: How does a specific group view the topic? What is especially important to this group's members? What questions do they have?

Exploring other points of view

Prompting for many different perspectives means an AI tool takes on different roles according to the points of view to be highlighted. For example, the tool could be asked to take on the role of a politician, an activist, a member of a minority, or someone belonging to a particular age group – in other words, anyone whose views on a topic are of interest. The tool's answers make it possible to explore and better understand unfamiliar points of view and potentially incorporate them into coverage in a sensible way. Here is a sample prompt that could help with this:

```
Take on the role of [person/group]. Describe their opinions and beliefs on [topic].
```

Reflecting diversity

Perspective-related prompts can also be structured in more complex ways. For instance, you could ask the AI tool to give everyone a chance to speak, including those voices that often go unheard. This guarantees many different perspectives.

```
What various groups/communities are impacted by [topic], and how do their perspectives and experiences differ?
```

The next step can include more details:

Take on the role of [person A, person B, ...].
Compile and explain their different perspectives, opinions and experiences with respect to [topic].
Finally, provide a list of minority groups that are impacted by [topic]. Explain the opinions and experiences of minority groups in relation to this topic.

It is best to structure this prompt in two steps; otherwise, certain tasks may not be executed.

Illuminating blind spots

Having many different perspectives also means diving into all aspects of a topic, including the complex causes of a conflict, in order to present a more complete picture. It means going beyond the obvious, asking questions and investigating everything from social and economic conditions to cultural and political factors.

In addition to common explanations, media professionals should also take into account perspectives that tend to be overlooked. This way, their reporting will include many different voices and reflect the fact that reality is complex while simultaneously avoiding the trap of false balance. This term, also known colloquially as “both-sidesism,” describes a form of media bias in which dubious positions are given legitimacy or minority opinions receive a disproportionate amount of space, sometimes in an attempt to cover all sides of a topic.

Constructive journalism, on the other hand, seeks to provide a nuanced understanding of topics and how they relate to each other, thereby empowering people to make informed decisions for themselves and society. AI can help with this by uncovering points of view that might otherwise get ignored:

Identify and explain the influences that play a role in [topic], as well as its different dimensions. Ask yourself the following question: In addition to the known actors, who else is impacted?

Adjusting your zoom

Journalists can also provide many different perspectives by zooming in on or out of a topic, thus highlighting its different levels.

Socially speaking, an event or topic can be examined at the individual (micro) level, the group (meso) level, or the national (macro) level.

For instance, media professionals can use AI to present an event or incident at the micro level (from the point of view of a single person) or at the macro level (in relation to larger social or political structures). Reporting in a way that includes many different perspectives means allowing for multiple zoom levels and angles.

Examine [topic] from the perspective of both a single impacted individual and society as a whole. Explain how the goals of each level differ from each other.

You can also zoom in and out of things in terms of time. Coverage of a current event changes when it is examined within the larger historical context. Zooming out to a larger time frame can make it easier to recognize deeply rooted causes and patterns.

For instance, this makes it possible to explore how local decisions can have global effects or how international trends impact individual communities. Another example would be not only viewing a political event, like an election, through the prism of the day it takes place but also seeing it as the culmination of historic developments and cultural trends.

Examine the history of [event] in the context of different time frames, geographic locations and social focuses. Examine the history leading up to the event and the impacts at the local and international level, and explain this from the perspective of a single individual, specific groups and entire nations.

Warning: In this use case, the AI tool is being asked to provide concrete facts, so all results must be checked thoroughly.

Self-reflection

Finally, AI can also help journalists engage in critical self-reflection. This means they interrogate their own ideas, convictions and possible biases in how they perceive things. The goal is to make coverage as objective and unprejudiced as possible.

This introspective approach helps incorporate a broader range of views and, among other things, avoids confirmation bias, or the tendency to give more value to information that reflects one's own perspective, thereby further strengthening it.

In practice, self-reflection for media professionals means actively looking for sources and perspectives that challenge their own assumptions. It is about being aware of how personal experiences can influence our interpretation of information. AI can help here by analyzing written drafts or research results and checking to make sure they contain many different perspectives.

Analyze [the following text/the following research results]. Critically reflect on possible prejudices in how [topic] is covered. Check how statements and assumptions in the text could influence how the topic is viewed, and actively look for information and perspectives that counter this.

A prompt that includes these elements can effectively guide an AI tool, and the resulting answers can help create journalistic content that informs, inspires and sparks societal discussion. Nevertheless, this process can only provide a rough approximation at best, since artificial intelligence delivers statistical stereotypes, not nuanced characters. In other words, chatting with an AI application cannot replace on-site research.

Prompting and constructive interview questions

Constructive interview questions require a deeper understanding of a topic, and they motivate both the interviewed individuals and the public to think more about positive change and progress. They direct attention to possible improvements, resilience and effective problem-solving strategies. Such questions are specially designed to show interview partners and the public various ways to overcome current challenges and achieve positive outcomes.

Artificial intelligence can play an important role in coming up with constructive interview questions by making the process more efficient, creative and thorough. It can help with brainstorming interview questions, and thus make coverage more constructive. Below, we present the key aspects of constructive interview questions and some example prompts that can be used when brainstorming with AI.

Focus on the future

Constructive questions encourage one to think beyond past or present situations and instead imagine a better future. Questions like, “What do you hope for in your current situation?” or “Let’s imagine the problem were suddenly solved. Where in daily life would people first notice this?” help to articulate visions and hopes, which can be a source of motivation.

Come up with questions that have to do with the vision and long-term effects of [topic]. Examples: “Let’s assume your initiative has achieved its primary aims. What would the situation look like ten years from now? What changes would be most noticeable for the community?”

Possible solutions

Instead of concentrating exclusively on problems, constructive questions focus on practical solutions and successes. They ask about small improvements that have already taken place and try to find out what contributed to these changes. Some examples: “Have things ever been a little bit better? What was different then?” or “Do you know any examples where things went slightly better? What was done there?”

Generate questions aimed at learning about past or potential future successes. For example: “Can you give me an example that shows how a similar problem has been successfully solved? What can we learn from it?”

Multiple perspectives

Constructive questions can tell a story in a more comprehensive and multidimensional way by asking about different points of view and including the perspectives of people who have been directly impacted by a topic. Some helpful questions here could be, “Who could I speak with to learn more?” or “How would this person likely tell the story?”

Come up with questions that encourage you to incorporate the opinions and experiences of people who have been impacted. For example: “What other groups have been affected by this topic, and what makes their views and proposed solutions different from your own?”

Reflection and contextualization

These types of questions spur a person to reflect on how they contribute to the status quo and what possibilities for change might exist. “Who has the power to change things, now or in the future?” is one example of how an interview partner could be prompted to reflect on their own role in a situation, as well as the role of other actors.

Suggest questions that prompt self-reflection. For example: “To what extent have your personal experiences and the context of your work contributed to your views on [topic]?”

From speech to text: Transcribing with AI

It can take quite some time to transcribe an hour-long interview word for word. It's also annoying. Fortunately, there are already various transcription tools out there that can perform this step – and pretty reliably, too. The technology that makes this possible is automatic speech recognition (ASR).

ASR transforms speech into text. Though it dates back to the 1950s, its development has accelerated in recent years. The improvements are largely thanks to advances in AI and machine learning (see p. 16): Systems can be trained to recognize complex speech patterns with a high degree of accuracy, and to identify and separate different speakers.

ASR transforms speech into text by first breaking down spoken language into small units, called phonemes, the smallest audible units of a language. By analyzing these phonemes in the context of the surrounding phonemes, ASR systems can identify words and sentences. Modern speech recognition systems use deep artificial neural networks (see p. 17) that learn from vast amounts of data. They recognize speech patterns and understand complex nuances like emphasis and rhythm.

Most transcription tools are web-based and do not require the installation of any special software. You can run them directly in a browser or on your smartphone, and they are quite intuitive to use. It takes just a few clicks to upload a file and start a transcription. But what should you look into if you want to use this technology? Accuracy, data protection and price can all factor into choosing a tool.

Accuracy

Journalists who frequently conduct interviews or attend press conferences can greatly benefit from the high accuracy of modern transcription tools, which can precisely capture even complex, specialized terminology.

Transcription-tool providers commonly highlight their accuracy rates, which are typically determined by the word error rate (WER). The WER measures the number of mistakes a transcription contains in comparison to a reference text. It does this by controlling for incorrectly transcribed, missing or additional words. A WER of 10% means that 90% of the words have been captured correctly.

Popular tools like **Trint**, **Rev** and **Descript** claim to have accuracy rates of 90% or higher, though it is important to keep in mind that rates are influenced by factors like audio quality and the complexity of the language. Tools that have been developed for specific languages often provide greater accuracy. This makes them particularly valuable for media professionals who need transcriptions in specific languages. For example, **GoSpeech** has specialized in providing German-language transcriptions. In contrast, journalists who work in multiple languages could find tools like **HappyScribe** and **Sonix** useful, since both support more than 50 different languages.

A transcription's quality is commonly limited by factors like accents, dialects and subject-specific jargon. To reduce the error rate, modern transcription tools use adaptive algorithms that improve their ability to process specific speech patterns with every use. Many tools also allow

users to define a custom dictionary. Here, they can add specific terms, proper nouns and abbreviations, which then enables those words to be better recognized and transcribed.

Data protection

Data protection is particularly important for journalists who conduct confidential interviews and need to protect the anonymity of their sources. Transcription-tool providers approach this issue very differently. It is therefore worth paying attention to how data is encrypted and how data-erasure requests are handled, as well as whether a tool complies with the General Data Protection Regulation (GDPR).

End-to-end encryption or a transport layer security (TLS) are examples of the safe transmission of data. Certifications can also indicate that sensitive data is handled appropriately. For instance, ISO 27001 is an internationally recognized standard that defines the requirements an information security management system (ISMS) must meet.

The data storage location is also crucial since, for example, European servers must comply with the GDPR whereas US servers are subject to different regulations. Tools like **Amberscript** and **F4X** save their data onto servers in Europe, which obligates the companies to meet the GDPR's strict requirements. It is worth noting that providers who promote data security often have the most transparent data protection policies.

Transcription tools vary greatly when it comes to price, which can be based on a subscription or a pay-per-use model. Many providers offer a free basic version that can meet the needs of a casual user. The price of a monthly subscription can vary from €5 to €60 depending on the provider and the amount of transcription time included.

Constructively tuning ChatGPT

ChatGPT users can adjust the chatbot to their individual needs or train it to execute specific tasks. Technically speaking, there are two ways of doing this: Individually configuring ChatGPT or building your own GPT.

Individual settings

By clicking on the tab “Customize ChatGPT” under settings in the top right corner, you can adjust the tool to your own needs. For instance, you can add and save custom instructions telling ChatGPT what it needs to keep in mind so that you don’t have to share the same information with it over and over. You could tell it, for example, that the account user is a journalist who works constructively.

You can also set how ChatGPT should respond, such as instructing the AI tool to always present constructive arguments that include possible solutions. This function is available in both the free and purchased versions.

Build your own bot

The second option is to train a version of ChatGPT for a specific purpose. This involves training an individual GPT bot to perform a specified task. OpenAI, the company behind ChatGPT, offers a GPT Store where all the GPTs that have been created by users can be found. Store access was initially only available to paying customers but in May 2024, the store opened to non-paying users as well. However, only paying users with a Plus account are currently able to create their own GPT.

There are various reasons why you might want to create a GPT: You could develop one for personal use, for team use or for the general public. A customized GPT can serve as an everyday sparring partner for media professionals. One area where it could be helpful is in brainstorming constructive ideas.

Before using a customized GPT for a specific purpose, the chatbot first has to be trained. This is done by providing the GPT with written instructions; it then programs itself accordingly. To start this process, the GPT is fed with information and instructions that it should take into account. This can be done directly by prompting or by uploading relevant documents. Training can take place over multiple rounds

during which the chatbot is repeatedly asked about key aspects. Inadequate answers can be fine-tuned with additional inputs.

Defining the intended use

To use a custom GPT constructively, the following must be clear from the get-go: The GPT should be trained to find possible solutions, present diverse perspectives and foster constructive dialogue. These goals must be explicitly stated in the instructions and training data.

You are a GPT that should help media professionals with finding and refining constructive ideas. Your goal is to find possible solutions, present diverse perspectives and foster constructive dialogue.

Providing the training data

A GPT must then be fed with information that will serve as the basis for its future answers. For a constructive GPT, it makes sense to upload the basic principles of constructive journalism as a PDF or Word document. This information should include the most important elements of constructive journalism: a focus on solutions, multiple perspectives and constructive dialogue.

Here is a document that contains details about the principles of constructive journalism. Use this information as the basis for each of your answers.

Creating specific instructions

Now it is time to come up with instructions that direct the GPT to focus on possible solutions, consider multiple perspectives and foster dialogue. The way the GPT should respond to requests also needs to be set. Should it respond formally or informally? Should it provide information in paragraph form or in a table?

Each of your answers should prioritize possible solutions, take multiple perspectives into account and foster dialogue. Provide your answers in the format that best serves the request. Use informal language.

Constant fine-tuning

Constant feedback and adjustments ensure that the GPT delivers suggestions that are consistent with and relevant to journalistic work. This is achieved over multiple rounds of training during which the GPT is repeatedly asked about key aspects and any inadequate answers are fine-tuned.

You sometimes provide answers that do not reflect the basic principles of constructive journalism. Please use the information that has been provided for each of your answers.

There are many advantages to adapting a GPT for use in constructive journalism. It can help journalists work more efficiently and purposefully, and also aid them in identifying and presenting solutions and positive developments. Challenges include having to continuously watch and adjust the GPT to make sure the results meet desired standards and are free of biases.

Can AI recognize synthetic content?

Was a text written by a human or generated by AI? If it was AI, could the text possibly have been post-edited by another AI in order to appear more human? The whole scenario seems like a giant digital masquerade of information. It will become increasingly difficult for journalists to discover the origins of a text. AI detectors claim to be able to help with this. But can AI, of all things, reliably help recognize AI-generated content?

AI detectors are specialized algorithms that have been trained to recognize artificially created content. They use a range of technologies to differentiate between AI-generated texts and texts that have been written by humans. These technologies unmask AI-generated content by identifying unique patterns or “fingerprints” that set it apart from human creations.

Regularity – an important indicator

Some relevant factors include readability, punctuation and the use of certain words and phrases. Another important aspect is how repetitive a document’s sentence structure is: Specific words or phrases tend to be regularly distributed in AI-generated texts, making them predictable. And it’s precisely this predictability that is a key indicator for AI detectors that a text has been artificially generated.

In addition to these factors, AI detectors also analyze context by examining a text’s writing style and overall consistency. One prominent example of this is the stylistic examination of content from the group QAnon, which spreads conspiracy theories online, among other things. Some time ago, researchers used artificial intelligence to analyze writing styles and consequently identify the potential creators of specific conspiracy theories. AI-based style analyses also make it possible to uncover the authors behind pseudonyms.

Other AI detectors use previously trained language models to calculate the probability that a piece of content has been created by AI. They then check whether this content has the same characteristics as other content that has been generated by the language model itself.

A third group of AI detectors, also based on language models, has been specially trained to recognize synthetic texts. These applications are fed datasets that contain both human-written and synthetic texts. They learn to recognize the differences between them and determine whether a piece of content has been generated by AI or created by a human.

Lots of uncertainty

Despite such sophisticated technologies, AI detectors quickly reach the limits of what they can do. In July 2023, OpenAI removed its proprietary ChatGPT recognition tool, referred to as a text classifier, from its website due to its “low rate of accuracy.” This decision reflects a more complex problem that has also become apparent in recent studies. For instance, a study published in December 2023 on the accuracy of 14 tools that claimed to detect AI-generated texts found that almost every fifth text went undetected (see Weber-Wulff et al. 2023).

There are many possible reasons for such inaccuracy. AI detectors often have trouble keeping up with the latest writing styles and techniques. Insufficient training data can also cause problems, especially with texts written in languages other than English.

Tools that refine AI-generated texts to include human-like writing mistakes or vary the writing style pose yet another challenge to recognizing synthetic content. These “humanizer” tools add another deceptive mask to our digital masquerade.

The takeaway is that AI detectors can be a helpful, but far from error-free tool for media professionals looking to identify artificially generated content. Even the most sophisticated detectors can only recognize the masks they have learned to see.

On the hunt for deep fakes

In addition to AI-generated text detectors, technologies that recognize deep fakes are also very important, since deep fakes pose a unique threat to society. Manipulated videos and photos can significantly influence public opinion. Advanced AI detectors that can identify inconsistencies in images and sounds are therefore helpful tools for journalists who want to prevent disinformation and protect the integrity of their reporting.

This type of AI detector analyzes things like facial expressions and lip movements. While the technology will inevitably continue to advance, the challenges will nevertheless remain, as deep fakes are becoming increasingly realistic and harder to identify.



Index and glossary

Algorithms are clearly defined series of instructions for solving problems or completing tasks. An algorithm can range from a simple “If-then-conditions” sequence all the way to complex sequences of mathematical equations. In the context of artificial intelligence, so-called learning algorithms analyze large amounts of data for patterns that they then save. This storing of identified connections or relationships is also referred to as “learning.” It enables machines to understand linguistic content and formulate decisions. Learning algorithms are the basis of machine learning, deep neural networks and natural language processing.

→ see page 62

Automation → see page 14

Bias → see page 34

Chatbots are technical dialogue systems that facilitate exchanges between humans and technology through spoken or written language. There are two types of systems: AI-based and rule-based. AI-based chatbots learn from interactions and can generate complex answers, whereas rule-based chatbots deliver predefined answers upon command. The first chatbot, ELIZA, which was developed in the 1960s, used preprogrammed answers to simulate a psychotherapist.

→ see page 16, 20, 22 ff., 42, 59 ff.

Constructive journalism → see page 9 ff. and 49 ff.

Diversity penalty → see page 46

Foundation models are AI models that can complete a range of different tasks and consequently deliver a variety of results such as texts, images or sound. They can work independently, or they can also serve as the basis for other programs (see Jones 2023). To improve the transparency and comparability of these varied AI models, the Stanford Center for Research on Foundation Models

(see CRFM) created a transparency index for foundation models that lays out standards on openness and accessibility in AI development. → see page 20 f.

Generative → see page 19

GPT → see page 20

Language models are AI-based computer programs trained to understand and generate human language. During their training phase, they use → artificial neural networks and lots of data to develop statistical models that they then use to predict the probability that a word or part of a word will appear after a given text fragment. For instance, if you prompt ChatGPT with the phrase “The sky is...,” it is more likely to suggest the word “blue” than the word “red.” One refers to large language models (LLM) when a computer program has been trained with extremely large amounts of data. → see page 19 f., 34 f.

Large language model (LLM)

→ see Language models

Machine learning → see page 16

Neural network → see page 17 f.

Prompt → see page 42 ff.

Prompt-engineering → see page 42 ff.

Refinement → see page 44

Temperature → see page 46 f.

Sources and literature

Avelar, Victor/Patrick Donovan/Paul Lin/

Wendy Torell/Maria A. Torres Arango (2023):

The AI Disruption: Challenges and Guidance for Data Center Design, Schneider Electric Energy Management Research Center, [online]

https://download.schneider-electric.com/files?p_Doc_Ref=SPD_WP110_EN&p_enDocType=White+Paper&p_File_Name=WP110_V2.1_EN.pdf.

Baumgartner, Katharina (2024): In den Kaninchenbau und wieder hinaus, in: Marie-Sophie Adeoso/ Eva Berendsen/Leo Fischer/Deborah Schnabel (Eds.), *Code & Vorurteil: Über Künstliche Intelligenz, Rassismus und Antisemitismus*, Verbrecher Verlag, p. 125–136.

Beckett, Charlie (2019): New powers, new responsibilities: A global survey of journalism and artificial intelligence, London School of Economics and Political Science, [online]. Available to download at: <https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/> [accessed June 3, 2024].

Beckett, Charlie/Mira Yaseen (2023): Generating Change: A global survey of what news organisations are doing with AI, London School of Economics and Political Science, [online]. Available to download at: <https://www.journalismai.info/research/2023-generating-change> [accessed June 3, 2024].

Becker, Kim Björn/Felix M. Simon/Christopher Crum (2023): Policies in Parallel? A Comparative Study of Journalistic AI Policies in 52 Global News Organisations [Preprint]. Available at: <https://osf.io/preprints/socarxiv/c4af9> [accessed May 24, 2024].

Bedford-Strohm, Jonas/Uli Köppen/Cécile Schneider (2020): Ethik der Künstlichen Intelligenz: Unsere KI-Richtlinien im Bayerischen Rundfunk, in: *Bayerischer Rundfunk*, November 30, 2020, [online] <https://br.de/extra/ai-automation-lab/ki-ethik-100.html> [accessed May 24, 2024].

Bender, Emily M./Timnit Gebru/Angelina McMillan-Major/Shmargaret Shmitchell (2021): On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? in: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FACCT 21)*, p. 610–623. <https://doi.org/10.1145/3442188.3445922> [accessed May 24, 2024].

Bommasani, Rishi/Kevin Klyman/Shayne Longpre/Sayash Kapoor/Nestor Maslej/Betty Xiong/Daniel Zhang/Percy Liang (2023): The Foundation Model Transparency Index [Preprint]. Available at: <https://arxiv.org/abs/2310.12941> [accessed May 24, 2024].

Brammer, Robert (2023). Die Macht der Computer. Zum 100. Geburtstag von Joseph Weizenbaum, in: *Deutschlandfunk Kultur*. *Aus den Archiven*, 14. Januar 2023, [online] <https://deutschlandfunkkultur.de/weizenbaum-100.html> [accessed May 29, 2024].

Brooks, Rodney (2018): Die Ursprünge der Künstlichen Intelligenz, in: *reframe Tech*, 22. November 2018, [online] <https://reframetech.de/2018/11/22/die-urspruenge-der-kuenstlichen-intelligenz/> [accessed April 18, 2024].

Burgess, Matt/Reece Rogers (2024): How to Stop Your Data From Being Used to Train AI, in: *Wired*, April 10, 2024, [online] <https://wired.com/story/how-to-stop-your-data-from-being-used-to-train-ai/> [accessed May 17, 2024].

Center for Research on Foundation Models [CRFM] (n.d.): Foundation Model Transparency Index, [online] <https://crfm.stanford.edu/fmti/> [accessed May 16, 2024].

Clarke, Seán/Dan Milmo/Garry Blight (2023): How AI chatbots like ChatGPT or Bard work – visual explainer, in: *The Guardian*, November 2, 2023, [online] <https://theguardian.com/technology/ng-interactive/2023/nov/01/how-ai-chatbots-like-chatgpt-or-bard-work-visual-explainer> [accessed May 24, 2024].

Commissariat à l'énergie atomique et aux énergies alternatives [CEA] (2018): [Histoire des sciences] L'histoire de l'intelligence artificielle (IA) [YouTube], <https://youtube.com/watch?v=qmwJx-r5vmm>.

Common Crawl (2024): Overview, [online] <https://commoncrawl.org/overview> [accessed May 24, 2024].

Council of Europe (2023): Guidelines on the responsible implementation of artificial intelligence systems in journalism, [online] <https://rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6>.

DataScientest (2023): Künstliche Intelligenz: Definition, Geschichte, Chancen und Risiken, in: *DataScientest*, August 8, 2023, [online] <https://datascientest.com/de/kuenstliche-intelligenz> [accessed April 18, 2024].

De Vries, Alex (2023): The growing energy footprint of artificial intelligence, in: *Joule*, vol. 7, no. 10, p. 2191–2194. <https://doi.org/10.1016/j.joule.2023.09.004>.

Der Spiegel (2021): Einzelanalysen für alle 299 Wahlkreise, in: *Der Spiegel*, September 27, 2021, [online] <https://spiegel.de/politik/deutschland/bundestagswahl-2021-alle-wahlkreis-ergebnisse-der-btw-a-6c222d37-87d8-4d85-bec9-3d95e6e0eefa> [accessed May 24, 2024].

Deutscher Journalisten-Verband [DJV] (n.d.): Positionspapier bezüglich des Einsatzes Künstlicher Intelligenz im Journalismus, [online]. Available to download at: <https://djv.de/medienpolitik/kuenstliche-intelligenz/> [accessed May 24, 2024].

Deutscher Presserat (2024): Jahresbericht 2023, [online] https://presserat.de/files/presserat/bilder/Downloads%20Jahresberichte/DPR-2024-Jahresbericht%202023_BF.pdf [accessed May 24, 2024].

Diakopoulos, Nicholas (2019): *Automating the News: How Algorithms are Rewriting the Media*, Harvard University Press.

Facklam, Anton Yong (2024): “Superintelligenz” bei Meta: Mark Zuckerberg plant KI-Revolution, in: *CHIP*, January 31, 2024, [online] https://chip.de/news/Kuenstliche-Intelligenz-bei-Meta-Mark-Zuckerberg-plant-AGI-Revolution_185116013.html [accessed April 23, 2024].

Fraunhofer-Institut für Kognitive Systeme [IKS] (2024): Künstliche Intelligenz (KI) und maschinelles Lernen, [online] <https://iks.fraunhofer.de/de/themen/kuenstliche-intelligenz.html> [accessed April 18, 2024].

Gartner (2022): Gartner Unveils Top Predictions for IT Organizations and Users in 2023 and Beyond, October 18, 2022, [online] <https://gartner.com/en/newsroom/press-releases/2022-10-18-gartner-unveils-top-predictions-for-it-organizations-and-users-in-2023-and-beyond> [accessed May 28, 2024].

Haustein-Teßmer, Oliver (2024): *Digitaler Erfolg im Lokaljournalismus – Leitfaden für Praxis und Ausbildung in Redaktionen*, Springer VS.

Heinrichs, Ellen (2018): Wir müssen reden: über Künstliche Intelligenz und Ethik im Journalismus, in: *Medium*, October 30, 2018, [online] <https://ellenheinrichs.medium.com/https-medium-com-ellen-sch-medienethik-und-ki-c953c2d70fba> [accessed May 24, 2024].

Jones, Elliot (2023): Explainer: What is a foundation model?, Ada Lovelace Institute, July 17, 2023, [online] https://adalovelaceinstitute.org/resource/foundation-models-explainer/#_ftn5 [accessed May 24, 2024].

Klofat, Aleksandra (2023): Wie funktionieren Transformer? Definition und Praxis, in: *Informatik Aktuell*, May 16, 2023, [online] <https://informatik-aktuell.de/betrieb/kuenstliche-intelligenz/wie-funktionieren-transformer-definition-und-praxis.html#c40768> [accessed May 24, 2024].

Landesanstalt für Medien NRW (2024): Akzeptanz von Prozess-Automatisierung und künstlicher Intelligenz in der Medienbranche, [online] https://medienanstalt-nrw.de/fileadmin/user_upload/lfm-nrw/Forschung/LFM-NRW_Akzeptanz_von_Prozessautomatisierung_und_Kuenstlicher_Intelligenz_in_der_Medienbranche.pdf.

McBride, Kelly (2024): Your newsroom needs an AI ethics policy. Start here., Poynter Institute, March 25, 2024, [online] <https://poynter.org/ethics-trust/2024/how-to-create-newsroom-artificial-intelligence-ethics-policy/> [accessed May 24, 2024].

McCarthy, John/Marvin L. Minsky/Nathaniel Rochester/Claude E. Shannon (1955): A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, [online] <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>.

Newman, Nic/Richard Fletcher/Craig T. Robertson/Amy Ross Arguedas/Rasmus Kleis Nielsen (2024): Reuters Institute Digital News Report 2024, Reuters Institute for the Study of Journalism, [online] https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2024-06/RIS_DNR_2024_Digital_v10%20lr.pdf.

OpenAI (2018): AI and compute, May 16, 2018, [online] <https://openai.com/index/ai-and-compute/> [accessed May 31, 2024].

Patterson, David/Joseph Gonzalez/Urs Hölzle/Quoc Le/Chen Liang/Lluis-Miquel Munguia/Daniel Rothchild/David So/Maud Texier/Jeff Dean (2022): The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink [Preprint]. Available at: <https://arxiv.org/abs/2204.05149> [accessed May 31, 2024].

Porlezza, Colin (2024): The datafication of digital journalism: A history of everlasting challenges between ethical issues and regulation, in: *Journalism*, vol. 25, no. 5, p. 1167-1185. <https://doi.org/10.1177/14648849231190232>.

Precedence Research (2024): Artificial Intelligence (AI) in Hardware Market Size to Reach USD 473.53 Bn By 2033, in: *globoNewsWire.com*, March 6, 2024, [online] <https://www.globoNewsWire.com/news-release/2024/03/06/2841613/0/en/Artificial-Intelligence-AI-in-Hardware-Market-Size-to-Rreach-USD-473-53-Bn-By-2033.html> [accessed May 31, 2024].

Reporter ohne Grenzen (2023): Paris Charta on AI and Journalism, [online] <https://rsf.org/sites/default/files/medias/file/2023/11/Paris%20charter%20on%20AI%20in%20journalism.pdf>.

Rinehart, Aimee/Ernest Kung (2022): Artificial Intelligence in Local News: A survey of US newsrooms' AI readiness, The Associated Press, [online]. Available to download at: <https://doi.org/10.13140/RG.2.2.16926.82246>.

Schweizer Presserat (2024): KI im Journalismus: Leitfaden, [online] https://presserat.ch/journalistenkodex/ki_leitfaden/ [accessed May 17, 2024].

Seward, Zach (2024): AI news that's fit to print: The New York Times' new AI leader on what this powerful tech can do for journalism, Reuters Institute for the Study of Journalism, March 13, 2024, [online] <https://reutersinstitute.politics.ox.ac.uk/news/ai-news-thats-fit-print-new-york-times-new-ai-leader-what-powerful-tech-can-do-journalism> [accessed March 27, 2024].

Stieler, Wolfgang/Katharina Zweig (2024, March 28): Deep Dive: Für eine Zukunft nach ChatGPT und Co. [Audio-Podcast], MIT Technology Review – Der Podcast, [online] <https://tech2go.podigee.io/105-neue-episode>.

Strubell, Emma/Ananya Ganesh/Andrew McCallum (2019): Energy and Policy Considerations for Deep Learning in NLP [Preprint]. Available at: <https://arxiv.org/abs/1906.02243> [accessed May 31, 2024].

TRG Datacenters (2023): AI Chatbots: Energy usage of 2023's most popular chatbots (so far), [online] <https://www.trgdatacenters.com/resource/ai-chatbots-energy-usage-of-2023s-most-popular-chatbots-so-far/> [accessed May 31, 2024].

Turing, Alan M. (2023): *Computing Machinery and Intelligence – Können Maschinen denken? Englisch/Deutsch* (2nd rev. ed.), Reclam.

Weber-Wulff, Debor/Alla Anohina-Naumeca/Sonja Bjelobaba/Tomáš Foltýnek/Jean Guerrero-Dib/Olumide Popoola/Petr Šigut/Lorna Waddington (2023): Testing of detection tools for AI-generated text, in: *International Journal for Educational Integrity*, vol. 19, no. 26. <https://doi.org/10.1007/s40979-023-00146-z>.

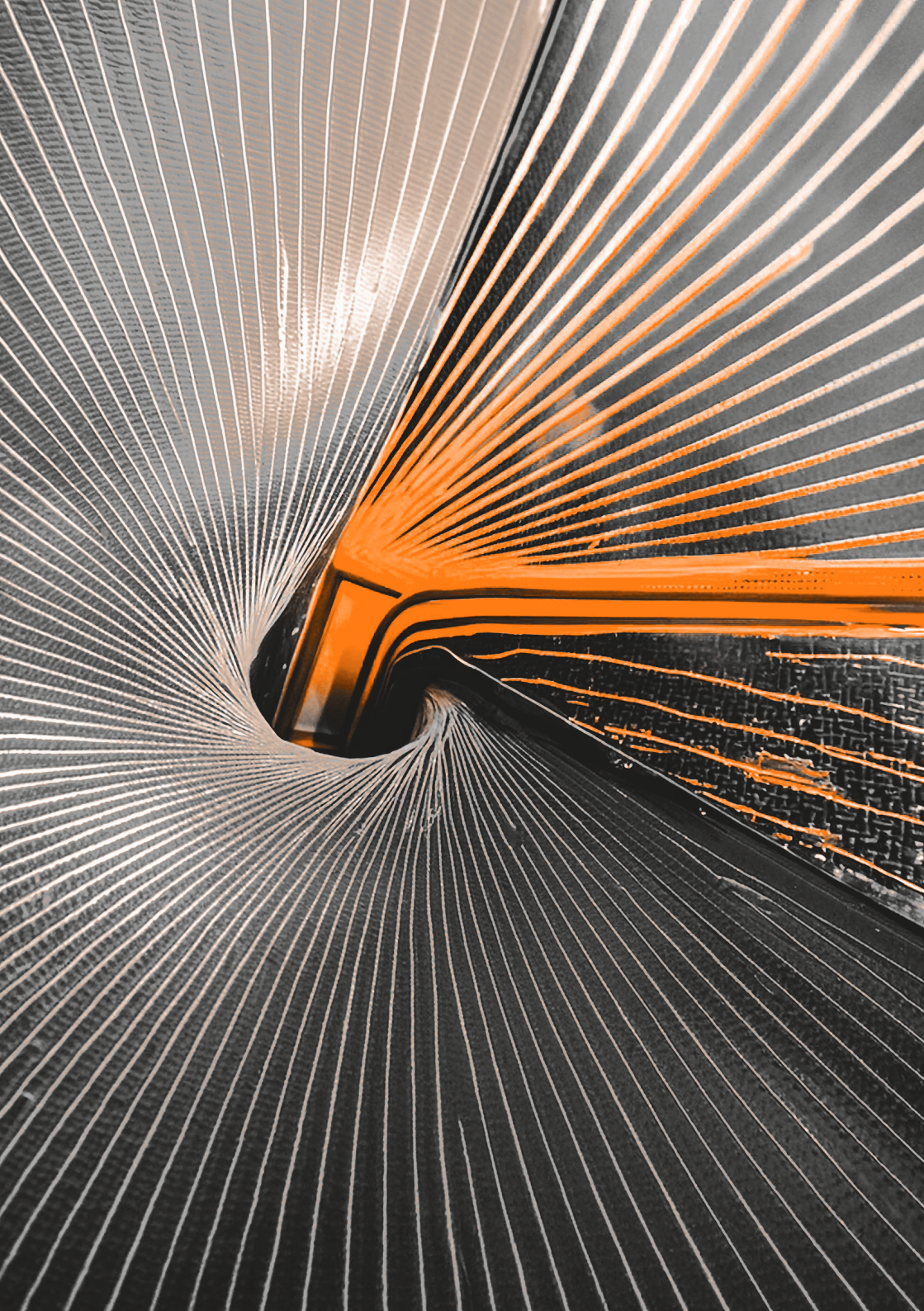
Whittaker, Meredith (2023): Vermessung bis ins Innerste, in: *netzpolitik.org*, June 7, 2023, [online] <https://netzpolitik.org/2023/kuenstliche-intelligenz-vermessung-bis-ins-innerste/> [accessed May 24, 2024].

Wired (2023): How Wired Will Use Generative AI Tools, May 22, 2023, [online] <https://www.wired.com/about/generative-ai-policy/> [accessed July 5, 2024].

Zentrum für vertrauenswürdige Künstliche Intelligenz [ZVKI]: Glossar, [online] <https://zvki.de/ki-navigator/unsere-inhalte/glossar> [accessed May 29, 2024].

Zweig, Katharina (2023): *Die KI war's! Von absurd bis tödlich: Die Tücken der künstlichen Intelligenz* (4th ed.), Heyne.

<prompt> abstract, rough, lines, clean, less detail, technical, black and white with orange, to the point </prompt>



The team



Claudia Isabel Rittel spent more than ten years working in various roles at the regional desk of the Frankfurter Rundschau. A political scientist by training, she served as the editorial developer at the regional Mediengruppe VRM in Mainz directly before joining the Bonn Institute in 2023, where she is the senior consultant for projects and research. She coordinated the *Constructive AI Compass* project, including the production of the booklet and related website. She also authored various sections of the booklet, including “Invitation,” “Understanding AI” and part of “Using AI Responsibly.”



Peter Lindner spent many years working at the Süddeutsche Zeitung, including almost 12 as head of the politics section and more than six as deputy editor-in-chief of SZ.de. Throughout this time, he focused on digital transformation and innovation, and it remains a key part of his work. In 2022, he conducted research at the Constructive Institute in Denmark on how journalism can promote constructive dialogue in society and help strengthen democracy. He is the head of projects and research at the Bonn Institute. He led the design of the *Constructive AI Compass* project and contributed to various parts of its realization.



Professor Dr. Ranty Islam teaches digital journalism at the Media University of Applied Sciences. His work focuses on the digital transformation of news desks, the role of empathy in new journalism concepts, constructive journalism and artificial intelligence. He and a student group ran trials for this project during the 2023-24 winter semester. He also provided expert fact-checking.



Patrick Große is a journalist and the founder of The AI Journalist. He also works at Deutsche Welle, where he leads the pan-European media project “ENTR what’s next.” His work focuses on applying artificial intelligence to journalism, especially in the field of prompt engineering. He helped the Bonn Institute develop the prompts included in this publication and also primarily wrote “Using AI Constructively.”



Dr. Ralph Hintemann has spent more than 20 years researching the interplay of digitalization and sustainability. An engineer and economist by training, he currently focuses on the potential and risks of new technologies such as artificial intelligence and the internet of things. A large part of his research revolves around analyzing the energy and resource needs of digital infrastructure. He authored “Environmental Impacts” for the *Constructive AI Compass*.



Dr. Peter Niepalla served as a corporate counsel and the head of Deutsche Welle’s legal department from 2003 to 2023. He has also taught media law at multiple universities for many years. He currently works as an advisor and consultant on questions of contract, labor and media law. He supported the *Constructive AI Compass* project by applying his legal expertise to “Using AI Responsibly” and critically reviewing all law-related questions.



Benedikt Dahlmann is a trainee at the Augsburgener Allgemeine Zeitung. He interned at the Bonn Institute in early 2024. He completed his master’s in political communication at the University of Dusseldorf, where his final thesis examined the impacts of constructive journalism on the public, earning him the 2024 Heinrich Heine Journalism Prize. He authored “Internal Guidelines” as well as “Tips for Creating an AI Code of Conduct” in the *Constructive AI Compass*.



Tobias Kohmann is a student assistant at the Bonn Institute. He is currently studying politics and society at the University of Bonn. As a young professional, he has gained experience at Phoenix and the Journalist Academy of the Friedrich Ebert Foundation. In 2022 he began working as a reporter for the university radio station BonnFM. He helped manage the *Constructive AI Compass* project and also wrote “Transcribing with AI,” “Can AI Recognize Synthetic Content?” and “Known AI Models.”

Acknowledgments

This publication would not have been possible without the support of many wonderful people.

We first want to acknowledge our colleagues at our partner organizations who were constructively involved in this project from the very beginning. Thanks to Simon Hof and Donata Dröge of RTL for the intense discussions about the content, the insight into the many AI-related activities taking place at RTL, and for letting us visit your offices in Cologne and Berlin. Thanks to Ruben Bouwmeester and Tilman Wagner of Deutsche Welle for inspiration and input at the start of the project and for helping us with fact-checking and editing at the end. Thanks to Peter Damgaard and Gustav Aarup Lauridsen of the Constructive Institute in Aarhus, Denmark, for inspiration, critical questions and technical expertise.

A big thank you to Prof. Dr. Ranty Islam for the academic perspective and help with fact-checking. Thanks also to his students for the many tests they did exploring basic questions about the constructive use of prompts in journalism. The Bonn Institute and Patrick Große used these as the basis to develop the prompts that are presented in this booklet.

Special thanks also to Kai Kullen and Nikolaus Rulle of KKNR for the innovative graphic concept, clean layout and very successful AI-generated illustrations; the prompts used can be seen alongside each image. Thanks also to Oliver Schwarz of schwarzdesign for the professional and attractive layout of the content on the Bonn Institute website.

All texts were created by humans, without AI. The many people involved in their creation can be found on pages 70 and 71. A big thanks goes out to all of them here. Thanks also to Cristina Burack, who provided the English translation of the texts. And, of course, a heartfelt thanks to the whole team at the Bonn Institute for the energetic support.

Finally, none of this would have been possible if it weren't for the initiative of Bonn Institute founder and CEO, Ellen Heinrichs.

Bonn, June 2024

Claudia Isabel Rittel and Peter Lindner

PUBLICATION DETAILS

Publisher

Bonn Institute gGmbH
Bennauerstraße 60
53115 Bonn

Legal representative

Ellen Heinrichs, CEO

Contact

Phone +49 (0) 228 28653370
hello@bonn-institute.org
www.bonn-institute.org

Concept

Peter Lindner
Claudia Isabel Rittel

Text and research

Claudia Isabel Rittel
Patrick Große
Dr. Ralph Hintemann
Benedikt Dahlmann
Tobias Kohmann
Peter Lindner
Prof. Dr. Ranty Islam
Dr. Peter Nipalla

Trialing

Patrick Große
Prof. Dr. Ranty Islam and students:
Felix Früh, Adriana Kauffmann, Calvin Kliemann,
Alina Langreck, Jolanda Ost, Lisa Schubert,
Alina Schunk, Lisa Vogt, Vanessa Wobb

Graphic concept and design

Kai Kullen and Nikolaus Rulle
kknr.de

Illustration

KKNR with the support of
Adobe Firefly (generative AI)

English translation

Cristina Burack

Printing

Brandt GmbH, Bonn

This publication was printed by a certified
business on 100% recycled paper that has
received the Blue Angel environmental label.

Picture credits

LFM-NRW p. 3; Fotostudio Bruder p. 70 (1);
Joerg Frank, Borderstep p. 71 (1);
Deutsche Welle p. 71 (2).
All other pictures private

First edition June 2024

All rights reserved. This publication or parts of
this publication may not be reproduced, stored
or transmitted in any form without the written
permission of Bonn Institute gGmbH.

© 2024 Bonn Institute

ISBN 978-3-9826202-5-1

Five constructive prompts to get you started

What specific aspects can I highlight when constructively covering [topic] to make sure I present different perspectives and possible solutions?

Identify and explain the influences that play a role in [topic], as well as its different dimensions. Ask yourself the following questions: In addition to the known actors, who else is impacted?

Generate questions aimed at learning about past or potential future successes. For example: "Can you give me an example that shows how a similar problem has been successfully solved? What can we learn from it?"

I'm looking for topics related to environmental conservation and climate protection that are suited to constructive journalism. Give me ten possible topics within this overarching subject area that I could work on as a journalist.

Examine [topic] from the perspective of both a single impacted individual and society as a whole. Explain how the goals of each level differ from each other.

Want to understand artificial intelligence and use it to make your journalistic work more constructive? Everything you need to know to do that is right here, in *Constructive AI Compass*.

This little booklet is brimming with AI knowledge and sample prompts that can help with brainstorming ideas, conducting research and preparing for interviews.

How does artificial intelligence work? How did it come about, and what impact does it have on the environment? What are the legal parameters for using AI tools responsibly? And what should media professionals keep in mind when creating company-specific AI guidelines?

Constructive AI Compass answers these questions and many more, providing orientation and inspiration for everyday work.

b

bonn institute