



Working with AI

Increasing your Productivity with Artificial Intelligence

Recognizing the limits of language models

Why you doublecheck language models
Who really owns the copyright of text

Apply AI programs

Where AI assistants can help you
In depth guide: AI audio, textual, visual

Rules for school and work

How AI is changing school and work
Avoid legal corporate issues

Generate an AI clone of your voice

Experiment with uncensored language models locally or in the cloud:
Control your data • Without subscription models • Individually customizable



Editorial

Dear readers,

Now that the initial hype surrounding ChatGPT has died down, many are looking for specific applications for language models. However, before putting these into practice, it is important to gain a basic understanding of the underlying structures. Open source platforms such as Hugging Face with tens of thousands of models invite you to experiment. We explain how to get started from page 6 onwards.

This inevitably raises fundamental questions: How do language models work? What data is used to train them? Can errors be corrected? To answer these questions, we show the limitations of language models from page 28 onwards. It quickly becomes clear in which areas AI can really help and where manufacturers are only pretending to make work easier.

We look at specific use cases and programs from page 66 onwards and explore where language models have already found their way into co-pilots, such as writing assistants for English texts or assistants in the softmaker office. From page 98, things get multimedia when AI voices speak texts and Midjourney generates images on demand.

Finally, starting on page 130, we take a look at the legal situation regarding the use of AI and the impact on companies and educational institutions. The first schools and universities are already gaining experience in integrating language models into lessons.

This issue gives you a comprehensive overview of how AI is changing the world of work and where its promises are just hot air.

Have fun browsing and trying them out.



Hartmut Gieselmann

Content

OPERATING YOUR OWN VOICE AI

Apart from ChatGPT, you can find thousands of open and free language models on Hugging Face. We show you how to set up your first language model and customize its inputs and outputs with prompts.

- 6 Free language models to experiment with
- 12 Operating free language AIs without the cloud
- 20 Falcon-7B language AI in Google Colab
- 24 Hosting your own language AI

RECOGNIZING THE LIMITS OF LANGUAGE MODELS

Why do language models talk so much nonsense and what kind of data are they actually trained with? And what about copyright? We analyze the AI researchers' papers and provide some insight.

- 28 Why language models make a lot of mistakes
- 34 How benchmarks measure language AIs
- 40 Interview: Language AIs need human testers
- 42 Quality problems in AI training
- 48 Avoiding errors in voice AIs
- 54 How to expose chatbots
- 60 c't readers trick Bard and ChatGPT

APPLY LANGUAGE MODELS

You can only speak English and need an assistant for tedious search, writing and programming tasks? We compare research, translation and office assistants.

- 66 What ChatGPT plug-ins can do
- 72 How ChatGPT helps with programming
- 76 English writing assistants put to the test
- 84 SoftMaker Office now with AI connection
- 90 Language AI explains research work



This magazine is the English translation of the German c't special issue 'KI Praxis', which was published by heise Medien in November 2023. The articles were machine translated and then edited. However, translation-related errors or indirect errors resulting from the translation cannot be ruled out. If you find any errors or have any comments on the translation, please send an e-mail to translation@heise.de. As technical development, particularly in the field of artificial intelligence, is progressing rapidly, we would also like to point out that the articles reflect the state of technology at the time the article was written.

EDIT IMAGES AND AUDIO

AI programs transcribe recordings and record texts. Midjourney also calculates photorealistic images. We compare the providers and provide tips for creative professionals.

- 98 Artistic styles with Midjourney
- 106 Recognising fake images generated with AI
- 114 AI voices in the test
- 122 Transcribing speech locally with Whisper

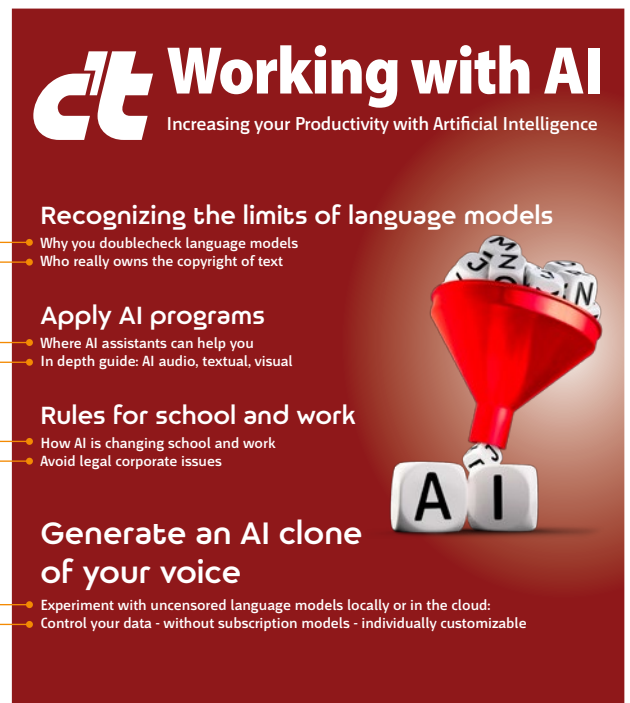
RULES FOR SCHOOL AND WORK

AI is transforming education and work. Companies need fixed rules, schools and universities need new concepts for integrating AI programs into lessons. Initial experiences are already available.

- 130 The labour market is changing
- 138 Learning and working with ChatGPT & Co.
- 148 Smart rules for AIs
- 156 Guidelines for AI in the company

TO THE MAGAZINE

- 3 Editorial
- 161 Imprint
- 162 Preview: c't PC Hardware-Tips



c't Working with AI
Increasing your Productivity with Artificial Intelligence

Recognizing the limits of language models

- 28 Why you doublecheck language models
- 42 Who really owns the copyright of text

Apply AI programs

- 66 Where AI assistants can help you
- 114, 76, 98 In depth guide: AI audio, textual, visual

Rules for school and work

- 138, 130 How AI is changing school and work
- 156 Avoid legal corporate issues

Generate an AI clone of your voice

- 24, 12 Experiment with uncensored language models locally or in the cloud:
- 6 Control your data - without subscription models - individually customizable

Free language models to experiment with

Your own language AI – How does that actually work? Where can you find a suitable language model if you don't want to rely on OpenAI? How do you adapt it to your own needs and where do you host it? An overview.

Jo Bager, Hartmut Gieselmann



Picture: Moritz Reichartz

Free language models to experiment with	6
Operating free language AIs without the cloud	12
Language AI Falcon-7B in Google Colab	20
Hosting your own language AI	24

🤖 Open LLM Leaderboard

The 🤖 Open LLM Leaderboard aims to track, rank and evaluate open LLMs and chatbots.

🤖 Submit a model for automated evaluation on the 🤖 GPU cluster on the "Submit" page! The leaderboard's backend runs the great [Fleuther AI Language Model Evaluation Harness](#) - read more details in the "About" page.

🏆 LLM Benchmark About 🚀 Submit here!

🔍 Search for your model and press ENTER...

Select columns to show

Average 📊 ARC HellaSwag MMLU TruthfulQA Type

Precision Hub License #Params (B) Hub ❤️ Model sha

Show gated/private/deleted models

Model types

pretrained fine-tuned instruction-tuned RL-tuned

Precision

torch.float16 torch.bfloat16 torch.float32 8bit 4bit GP

Model sizes

Unknown < 1.5B ~3B ~7B ~13B ~35B 60B+

T	Model	Average 📊	ARC	HellaSwag	MMLU	TruthfulQA
🔥	AIDC-ai-business/Marcoroni-70B-v1	74.06	73.55	87.62	70.67	64.41
🔥	ICBU-NPU/FashionGPT-70B-V1.1	74.05	71.76	88.2	70.99	65.26
🔥	adonlee/LLaMA_2_70B_LoRA	73.9	72.7	87.55	70.84	64.52
🔥	uni-tianyan/Uni-TianYan	73.81	72.1	87.4	69.91	65.81
🔥	Riiid/sheep-duck-llama-2	73.69	72.35	87.78	70.82	63.8
🔥	Riiid/sheep-duck-llama-2	73.67	72.27	87.78	70.81	63.8

or may not contain a great deal of detail. In some cases, such as Llama 2, there is a detailed description of the model and its sister models, including test results, the name of the developer, Meta, a link to the license and to scientific studies.

The Model Card also refers to Spaces. These are demo applications hosted by Hugging Face that use the model. There you can see a model in action immediately with just a few mouse clicks. For some language models, there are sample applications with which you can test them; see (ct.de/w8st).

However, you can also get started yourself and test a language model on Hugging Face. With Spaces, developers can build applications directly on the platform that are based on bare HTML, the Gradio front-end library, the Streamlit runtime environment or even any container. Spaces are Git repositories so that developers can embed them in their usual workflows and also develop with them in a team.

Training data

You can find training data in the main menu Datasets. There, Hugging Face lists over 60,000 public

training sources that you can filter by tasks, size, languages and licenses. These include texts from Wikipedia or variants of the cleaned Colossal Clean Crawled Corpus (C4) from Google, in which the websites crawled in the common crawl were cleaned by automatic word filters.

There are also special training sets, for example from the imdb film database or PubMed, a collection of texts from the American National Library of Medicine, which contains English-language specialist texts from the field of biomedicine. Another important collection of scientific texts is the arXiv dataset with 1.7 million published studies on arXiv.org.

There are currently 70 data sets available for German-language text generators. The 'Text generation' task also lists question-and-answer collections such as Open Assistant, which you can use for fine-tuning. You can also upload your own training data.

Where do they run?

Even if you do not train the language models yourself, running them requires a lot of computing power and memory. Therefore, you cannot realistically run

The Open LLM Leaderboard from Hugging Face lists language models and their performance in common tests.



Operating free voice AIs without the cloud

LLaMA, Alpaca, Vicuna: You can find heaps of medium-sized language models with all their parameters online. They also run on your own computer.

Pina Merkert

Do the work for me, AI: ‘Write me a friendly email to support asking where my package of lubricant has gone.’ Some tasks for AI language models are not so easy to send to a cloud service. Companies can even run into legal problems if they have AI read text with personal data outside the EU. But that doesn’t mean you

have to do without voice AI. Open source models, i.e. neural networks whose parameters are publicly available, can also run on your own hardware without the cloud.

The quality of the responses comes close to that of the GPT-3.5 language model of the free ChatGPT. In addition, the free models offer many

Live on the bleeding edge



Pina Merkert

It's great that you can now play around with gigantic language models on your own computer. However, the models are all still very new and the software to run them is subject to constant change. The GGML format has already gone through three versions this year and it can easily happen that the data format of a downloaded model does not match the latest version of llama.cpp after all.

This software is opening up to other architectures (especially in order to support LLaMA-1 and -2 at the same time) by switching to the GGML variant GGUF. For the Falcon models not based on LLaMA, however, there has long been a fork called ggllm.cpp, for which it is not clear whether GGUF will make it obsolete.

These are just examples of the fast-moving nature of the entire software infrastructure. The fact that libraries work with

new CUDA versions is to a certain extent a matter of luck. New Torch versions can have an impact on libraries that do not use Torch at all. Wrappers and helpers are often written for a specific version and then no longer maintained because developers have switched to a different base model.

At the moment, you can neither rely on stable APIs nor on long-term support. Libraries and frontends usually only work together if the developers have worked with precisely this combination. However, information on what works together is difficult to find and almost never documented.

During our research, we wasted days trying to compile Torch for CUDA 12.2, compile llama.cpp with the Visual Studio compiler and use the Python binding to llama.cpp with its low-level API for a browser-based chat. Most of the time something worked, but not everything, and we had to stop the research because we didn't want to force anyone to patch code with their own patches that wasn't yet ready for a reasonably stress-free use.

Feel free to try out the models mentioned here. It's fun to experiment and you'll get a feel for where the language model AI community is heading. However, we do not recommend investing too much time in individual experiments. For 99 percent of users, it is worth waiting until the software has matured and is no longer somewhere between a technical preview and an alpha version.

more possibilities because there are dozens of variants that can be fine-tuned to different data sets, i.e. retrained, and because you have full control over the prompt. With your own prompt, you can elicit cynical or funny responses from an AI chat partner in no time at all, specify different language styles or provide up-to-date context information, for example from your own database, before the chat.

We give you a short tour through the forest of freely available language models and explain how to install llama.cpp, a program for the command line that makes them easy to use.

Wooly creatures everywhere

Unlike OpenAI and Google, the AI research department of Meta (Facebook) published its LLaMA models with all parameters in February 2023. Like GPT and Bart, LLaMA uses pre-trained transformers [1] on gigantic amounts of text for enormously large language models. At the same time, models with 7, 13, 33 and 65 billion parameters appeared. Meta actually only wanted to make the parameters available to other researchers, but just one day after the release they were available for download from several places on the Internet.



Image: KI Midjourney / Editing.ct

Falcon-7B voice AI in Google Colab

AI's have become such memory monsters that you can hardly run them on your own computer, let alone edit them. Fortunately, there are free cloud services such as Colab, where you can install languagemodels, for example, and customize them as you wish.

Wilhelm Drehling

Language AI's consume huge amounts of computing power and memory - so much that it is almost impossible to run them on your own PC. If you still want to experiment with them, for example to learn how to use them from Python scripts to integrate them into your projects, you don't

have to invest in expensive hardware straight away. Instead, use the cloud - free of charge. For example, the Colab service (short for Colaboratory) from Google [1]. Colab is a Python editor that is based on that of Jupyter-Notebook. To use it, all you need is a Google account. In the free version, you have 12.7 GB of

system RAM, 78.2 GB of drive memory and 15 GB of GPU RAM. That's more than enough to download and try out small language models.

Spoilt for choice

Falcon-7B, a language model developed by the Technology Innovation Institute (TII) in Abu Dhabi, serves as an example for this article. More precisely: a fully trained variant called Falcon-7B-Instruct, which just fits into Colab's free notebook.

The appendage 7B indicates that the model is 7 billion tokens in size. It therefore provides information about how much data the language model has been trained with. According to the leading AI company OpenAI, 100 tokens correspond to an average of around 75 words learned in English. For comparison: ChatGPT-3 contained more than 175 billion tokens when it was released. Falcon-7B seems small by comparison, but that doesn't mean you should underestimate the AI. During development, TII relied heavily on training methods from the ChatGPT-3 paper by Brown et al [2]. Further information on tokens and training data can be found at ct.de/w6bh.

AI models such as the fully trained Falcon-7B-Instruct are available from the open source platform Hugging Face, which hosts thousands of models.

For example, the bigger brother Falcon-40B, which is currently at the top of the Hugging Face Open LLM Leaderboard. To make downloading and using the AIs as easy as possible, the brains behind Hugging Face have developed a library called 'transformers'. With its help, you can easily load an AI of your choice into your Colab notebook. All you need is a handful of lines of code, which Hugging Face serves up on a silver platter (see ct.de/w6bh). More on this later.

Caution: Not every model on Hugging Face can be used on Colab, many are too large or require too many resources. Not even the upgrade to the paid model Colab Pro for 11 euros or Colab Pro+ for just under 50 euros per month makes sense, as the additional resources are not enough to handle the huge AIs.

Set up

In the following, we explain how to run an AI delete in Colab. However, if you only want to experiment with the Falcon-7B language model, you can use the notebook we have prepared for this article (see ct.de/w6bh). When you click on it, it is automatically transferred to your Colab environment. You can read how to run and use the notebook in the 'Speak!' section.

You can access the cloud service via the URL <https://colab.research.google.com/>. Create a new document with 'File/New notebook'. Change the unattractive name 'Untitled.ipynb' in the top left-hand corner by clicking on it. First you need to tell Colab which resources it should assign to the notebook. To do this, open the menu under 'Runtime/Change runtime type'. Python 3 should be preselected as the language and you need a GPU as the hardware accelerator; in the free version, this is an Nvidia T4. Close the menu with 'Save'.

In order to run the code, additional libraries is required. Click on '+Code' to create a code box in which you can drop the commands for reloading:

```
!pip install -U transformers
!pip install -U einops
!pip install -U git+https://github.com/huggingface/accelerate.git
```

You already know transformers, which you use to download the AI from Hugging Face. einops and accelerate add functions that transformers need in the background to run the language model, such as fast tensor operations. The -U appendage forces pip to

Notebook-Einstellungen

Laufzeittyp
Python 3 ▾

Hardwarebeschleuniger
GPU ▾ ⓘ

GPU-Typ
T4 ▾

Sie möchten Premium-GPUs nutzen? [Zusätzliche Recheneinheiten erwerben](#)

Erste Zelle oder ersten Abschnitt automatisch ausführen

Beim Speichern dieses Notebooks Codezellenausgabe weglassen

Abbrechen [Speichern](#)

Assign the required resources to the notebook in the notebook settings. In this case, a powerful Nvidia Tesla T4 graphics card.



Picture: Montz Reichartz

Host your own voice AI

Large language models require plenty of computing power that you don't have under your desk. Hosting providers have adapted to the AI boom and offer flexible, AI-optimized servers – especially in the cloud.

Holger Bleich

Just a few months ago, there were no options: Anyone who wanted to use or integrate a powerful voice AI could hardly avoid ChatGPT. But access to the API has disadvantages. It costs money and is limited to the Large Language Model (LLM) GPT 3.5 or 4.

Fortunately, this has changed: Freely available, pre-trained LLMs enable private individuals and companies to host their own voice AI. In this article, you will find out where you can outsource such installations if your own resources are insufficient. Flexible and scalable cloud instances are particularly suitable.

In particular, this allows you to avoid a particular problem when accessing the OpenAI API: the

controversial data protection constellation. If desktop programs or web services forward customer prompts to ChatGPT via API, for example, they give up control of this often personally identifiable and sometimes highly personal data. This data is processed in the OpenAI black box.

Data protection experts in Europe are currently racking their brains over the conditions under which such a construct could be legal in light of the EU's General Data Protection Regulation (GDPR). The prevailing opinion is that it cannot. Each customer would have to give their individual, informed and revocable consent to the transfer of data to the API. However, such consent cannot

be effective at all due to a lack of transparency of OpenAI.

So anyone who currently uses the API, whether to support their own company or to answer customer questions, is treading on very thin ice from a legal perspective. They should urgently consider whether the self-hosted voice AI is a viable alternative.

GPU power required

Where is the best place to install your own voice AI? Private individuals should start by experimenting on their own computer. However, the PC's hardware generally limits both the processing speed and the possible model size.

A CPU in the range of Intel's eighth-generation Core i7 or AMD's third-generation Ryzen 5 in combination with at least 16 GB of RAM should be sufficient for Llama 2, for example. However, a powerful graphics card is more crucial, as the models not only benefit greatly from many GPU cores – in combination with sufficient VRAM during training, but also during prompt requests, i.e. in inference mode.

A current mid-range gamer card is almost mandatory if you want to set up an LLM at home. Small 7B versions of Llama 2 require at least an Nvidia GTX 1660 and 6 GB of VRAM, while the 13B models require at least an RTX 3060 with 12 GB of VRAM. Mind you, this is only enough to run the model with reasonable performance, i.e. without long response times –, but not for training.

One thing is clear: whether at home or in the data center, it is not the CPU but the graphics processor (GPU) that plays the decisive role. As described in the previous article, many open LLMs

rely on Nvidia's proprietary Compute Unified Device Architecture (CUDA), which in turn requires Nvidia GPUs in the computer. Without Nvidia, there is currently very little going on in the field of generative AI.

Nvidia GPUs with Turing architecture are currently considered the minimum standard for smaller projects. Even the popular Tesla T4 GPU is hardly an option for home PCs; cards cost at least EUR 1,700. Larger projects and training courses are more likely to run with A100 GPUs based on the Ampere architecture. Server cards with A100 and 40 GB of VRAM are definitely no longer a viable option for the average user, at a cost of EUR 8,000 each.

Expensive training

Training models from scratch is currently beyond the financial means not only of private individuals, but generally also of small and medium-sized companies. The costs involved are not in a proportionate to the return, especially as it is now possible to fall back on open, pre-trained models.

One example: the US company MosaicML specializes in orchestrating training and fine-tuning of AI models for customers across multiple cloud platforms as cost-effectively as possible. As a proof of concept, around a year ago, it conducted several training sessions from GPT to ascertain the quality of GPT-3 over the course of several months. MosaicML trained its model on a cluster of 256 Nvidia A100 GPUs, each with 40 GB of VRAM.

According to MosaicML, the self-trained 30B variant achieved almost the same quality as OpenAI's GPT3 with 175 million parameters. Training with 610 million tokens in 36 days was necessary. The

English Examples of cloud offerings with Nvidia's T4 GPU

Provider	Amazon AWS	Google Cloud	Ionos	Microsoft Azure
Product	g4dn.xlarge	n1-highcpu-4 + T4-GPU	GPU Server	NCas_T4_v3
Web	aws.amazon.com/en/ec2/instance-types/g4	cloud.google.com/gpu	www.ionos.de/server/gpu-server	azure.microsoft.com/pricing/details/virtual-machines/series/
Type	virtual cloud instance	virtual cloud instance	Dedicated server	virtual cloud instance
Configuration	4 vCPUs, 16 GByte RAM, 125 GByte SSD, 1 T4-GPU	4 vCPUs, 26 GByte RAM, 1T4-GPU	1 AMD EPYC 7302P CPU, 128 GByte RAM, 1 TByte SSD, 1T4 GPU	4 vCPUs, 28 GByte RAM, 180 GByte storage, 1T4-GPU
Price per hour	0.50 €	0.09 + 0.33 = 0.42 €	0.68 €	0.49 €
Price per month for uninterrupted use	360 €	302 €	490 €	354 €

Why language models make many mistakes

Voice AIs like ChatGPT talk a lot of nonsense. And it will stay that way. AI developers cannot scale away the causes and backgrounds with even more data and even larger data centers. Instead, users must take the fallibility of language models into account in their applications.

Hartmut Gieselmann

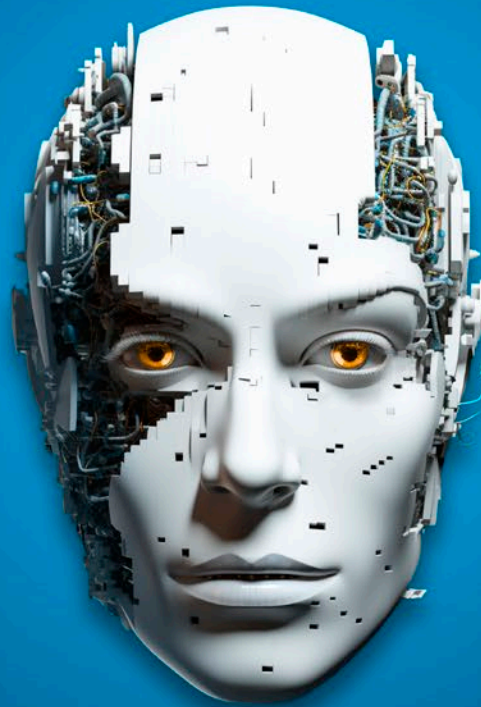


Image: KI Midjourney | Editing: ct

Why language models make a lot of mistakes	28
How benchmarks measure language AIs	34
Interview: Voice AIs need human validators	40
Quality issues in AI training	42
How to avoid voice AI errors	48
How to debunk chatbots	54
c't readers fool Bard and ChatGPT	60

ChatGPT has been on the market for almost a year now. The initial euphoria has since given way to disillusionment for some users. According to the analytics service Similarweb, the number of monthly hits on the chat.openai.com website fell from 1.8 billion to 1.5 billion between May and July 2023 – a loss of 17%.

One reason for this could be the high error rate of the language models. Everyone knows anecdotes in which the AI responds intelligently to a prompt and suddenly completely misses the mark, inventing incorrect court judgments or failing the simplest counting tasks (see article ‘c’t readers fool Bard and ChatGPT’). Due to the eloquence of the machines, however, these errors are more difficult to spot than when listening to a human who also flounders linguistically when he is at his wit’s end.

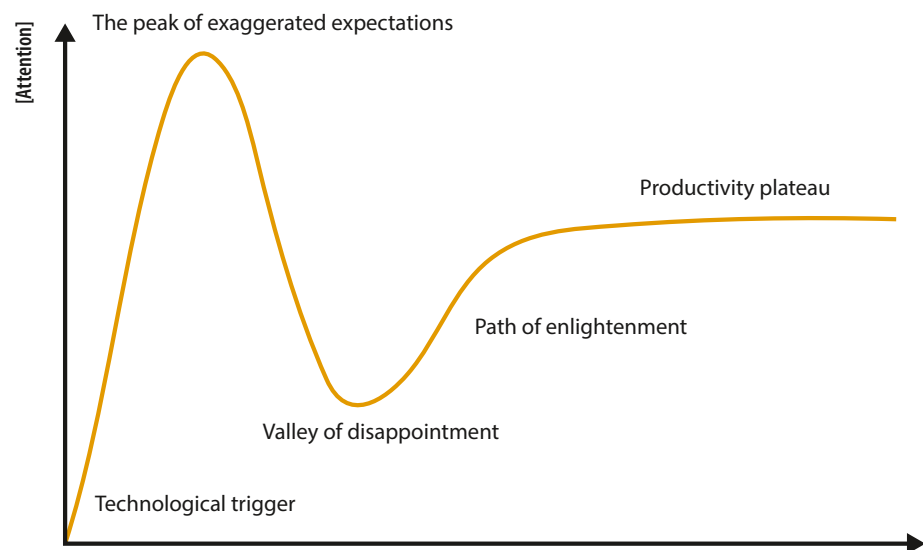
These sudden, unexpected errors are not outliers or teething troubles that developers can fix

with more training, more data and faster computers. Rather, they are a fundamental problem of language models for which there is no solution in sight. Unlike humans, machines are very good at recognizing patterns when they are fed a lot of data. However, they are poor at abstracting and generalizing individual observations.

This is why the promises of AI apologists and the doomsaying of catastrophists that language models will soon rule the world and put hundreds of millions out of work are irrational. This scattered superstition has only one goal: to inspire investors to put as much money as possible into AI companies. In this c’t focus section, we dispel this myth. However, we also show in which areas the use of language models is possible despite their fallibility – with all due caution – and which prompts and strategies can be used to reduce the error rate.

The AI Hype

The hype surrounding ChatGPT reached its ‘peak of exaggerated expectations’ in May 2023. Due to numerous criticisms, the chatbot is currently heading into the ‘valley of disappointments’, as the market research company Gartner calls it in its famous hype cycle. We will have to be patient a little longer before it reaches the ‘plateau of productivity’.



Source: Hype Cycle according to Gartner Inc.



Image: KI Midjourney / Editing ct

How benchmarks measure language AIs

'90% accuracy', 'better than humans': this is how developers, companies and the media like to hail the capabilities of new language models. In the wild, disillusionment usually follows, as almost every chatbot is bent on statements and hallucinating happily to itself.

Hartmut Gieselmann, Andrea Trinkwalder

In order to compare language AIs with each other and with humans, researchers are devising ever more sophisticated test procedures. However, measuring GPT-4, LLaMA, Bloom & Co. and their countless offshoots is neither an exact nor a simple science. Benchmarks and

metrics barely keep pace with the enormous leaps in the development of large language models (LLMs), often contain errors or turn out to be of little use.

This is why institutions and companies are now using crowdsourcing to compile huge, interactive-

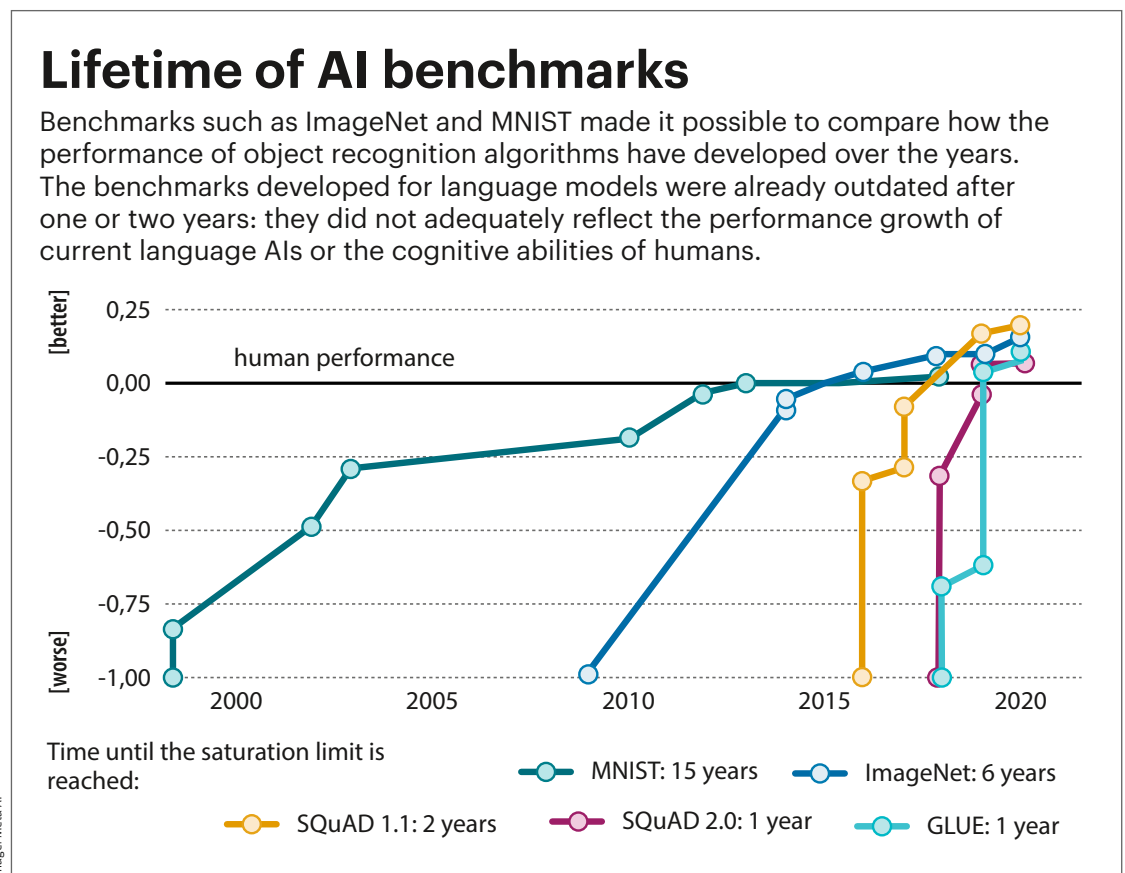
ly evolving collections of tasks, devise the trickiest possible problems and differentiated test scenarios or organize exciting shootouts with humans, machines and rankings.

Too good to be true

The text generators are grammatically proficient, but today it's all about semantics and the varieties of human thought: logical reasoning, transfer learning, generalizing, arguing, providing step-by-step explanations, forming analogies, understanding allusions, avoiding trick questions. Multitask benchmarks such as GLUE (General Language Understanding Evaluation) and the question-answer catalog SQuAD (Stanford Question Answering Dataset) were developed for this purpose, both presented in 2018. Consisting of nine tasks, GLUE was already considered outdated in

2019 because it no longer challenged the language models of the time and was replaced by the trickier SuperGLUE. This was exhausted after 18 months.

In order to measure progress and differences in quality, even more, even trickier, even more differentiated task catalogs were needed. Numerous researchers, companies and volunteers are now working on these together. A particularly extensive one is Beyond the Imitation Game (**BIG-bench**). The collection contains 204 task areas, compiled by 450 authors from 132 institutions. This means that BIG-bench covers an enormous range, from mathematical and scientific thinking to analogy building to 'common sense', which is not clearly defined. Most areas contain at least one hundred sample questions, some even tens or hundreds of thousands. Nevertheless, only part of this mammoth benchmark is obsolete.



Language AIs need human validators

Large language models such as ChatGPT and Bard can do a lot because they have been trained with almost all the world's knowledge. However, it is difficult to find out what they are really capable of. One person who is trying to do this is René Peinl from Hof University of Applied Sciences.

Andrea Trinkwalder

ct: Large language models (LLMs) such as ChatGPT and Bard usually perform very well in benchmarks, sometimes even better than humans. Under realistic conditions, such as in dialog with users, they often fail to meet expectations. What is the main problem when benchmarking language models?

René Peinl: Essentially, automated evaluation has its limits. It is usually based on predefined answers that are evaluated using metrics that are outdated from today's perspective. These metrics rely on a one-to-one overlap of the answer with the basic truth, i.e. with what a person has previously determined to be the correct answer.

There are also more advanced measures, for example the BERTScore (Bidirectional Encoder Representations from Transformers, for details see the previous article, editor's note), which itself uses AI to determine the semantic similarity of the answer to the expected answer. However, this method is still not in widespread use and it does not work reliably under all circumstances, for example if the answer is significantly longer than the expected answer.

Completely unexpected, yet correct answers also cause problems: they are often classified as incorrect. Even evaluation with the much more advanced GPT-4, which is currently probably the best automated option, still has its pitfalls. For example, GPT-4 often prefers the first answer when it has to evaluate the suggestions of several LLMs.



Picture: Hof University of Applied Sciences

c't: Can you give an example of this?

Peinl: In my evaluation, for example, I asked what relates to Japan as pizza does to Italy, and noted sushi as the expected answer. A language model gave tempura as the answer. It is also a typical, very well-known Japanese dish that is even closer to pizza because of the dough. However, this suggestion would fail all automated processes because it is very dissimilar to the expected answer. The more competent I say the language models become, the more you have to go for open questions and can no longer rely on purely fact-based ones.

If you ask about the capital of France, the height of the Eiffel Tower or the highest mountain in the Alps, it's relatively easy to tell the wrong answer from the right one. But if you want to know how we can best stop climate change, then there are a hundred sensible and many more less sensible answers and you need a very fine, gradual gradation. And evaluating these longer textual answers is automatically very difficult to almost impossible. At the moment, there is really no way around human evaluation.

c't: And what about the tasks themselves? Do they always measure the right thing?

Peinl: The way in which the questions are asked can also lead to distortions. For example, you try to find questions that are potentially non-trivial for the language model to answer and then find that

René Peinl conducts research into speech recognition and speech synthesis at Hof University of Applied Sciences.

people also have very vehement difficulties with this type of question because they are artificially pushed in one direction.

c't: So in the end, skills are tested that are completely irrelevant in practice for everyday problems or even professional issues?

Peinl: No, the benchmark results can certainly be relevant, but you need a balanced mix of different tasks. That's why the evaluation is no longer based on just one or two benchmarks. Instead, a broad range is used, i.e. at least a dozen or better two or three dozen. Bayerischer Rundfunk recently ran through the Bavarian Abitur from the years 2022 and 2023 and then had people evaluate the results.

c't: In your opinion, is it possible to reliably measure transfer learning ability?

Peinl: Very difficult, because you don't even know what exactly went into the training data. Transfer knowledge thrives on the fact that the new context to which you have to transfer is completely unknown. But there are probably only a few such contexts because the training data contains more or less the entire textual Internet.

For example, I used the following task to test the extent to which language models are able to replace one known concept with another: 'Let's assume that red represents a tomato, green a cucumber and orange a carrot. What do I get if I mix red, green and orange?' Many language models then stick to the color and answer 'brown'. However, some manage to apply the substitution rule and come up with salad or smoothie.

c't: Aren't all these benchmark questions actually burned up at some point, because sooner or later they're incorporated into the training data?

Peinl: There is at least no guarantee that this will not happen. At least BIG-bench has a so-called canary ID built in. This means that a long string that does not occur anywhere else is integrated into the questions so that it can be tested later. If a language model knows the canary string, then you know that it has been trained on it. Conversely, this marker helps the developers of LLMs to exclude the benchmark questions from their train-

ing pool when crawling the internet for data. Serious researchers also adhere to this, of course.

c't: How did you go about finding questions for your evaluation that were as unused and relevant to practice as possible?

Peinl: I tried to come up with my own questions, and I'm relatively certain that some of them have not yet been answered on the Internet. But I took the catch questions for humans from the Internet and only modified them slightly. They should therefore have tended to be included in the training data. Nevertheless, they were answered surprisingly poorly by the models. Apparently the material was not sufficient to memorize the tasks.

c't: How well can the benchmarks differentiate between thinking and memorizing?

Peinl: Either an answer is appropriate in context or it isn't. Critics like to get stuck on arguments like: 'Well, the answer was good, but it's more or less just memorized and a bit polished in the language. There's no real intelligence behind it. I find the discussion tedious, yes, artificial.'

c't: But these human/machine comparisons are of course also used deliberately to keep attention levels high. And when the AI hallucinates, it naturally disappoints expectations. In the end, it has to come down to this: How well can it fulfill a specific purpose?

The more practical the problems are, the more difficult it is to automatically assess whether the language model is doing a good job or not. But here too, I advocate making the comparison with people. Even in a recruitment interview, I as a human being can only find out to a limited extent whether the person will really do a good job later on, and I am not immune to my own hallucinations. If you base important business decisions on this, you can fall flat on your face. The situation is similar with AI. Of course, it would be desirable to be able to rule out hallucinations completely. But I find it difficult to say that as long as these hallucinations still exist, the models are useless. (atr) **ct**

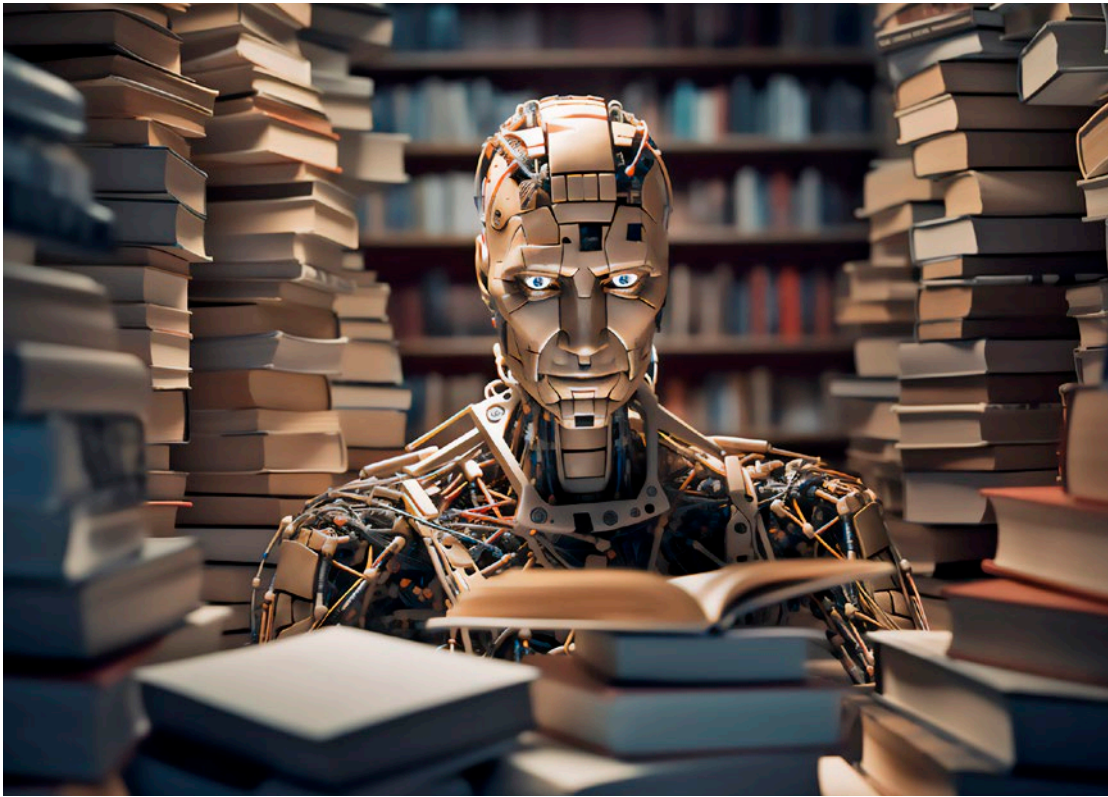


Image: KI Midjourney / Editing: ct

Quality problems in AI training

Reading educates, according to popular opinion. But when machines devour millions of websites and hundreds of thousands of books, they don't necessarily become smarter, as a look at the bookshelves of artificial intelligence shows.

Hartmut Gieselmann

Shawn Presser is probably the most hated American among book authors and writers at the moment. Authors have as bad an opinion of him as musicians have of Shawn Fanning and Sean Parker: in 1999, they launched

the music file-sharing platform Napster, triggering the biggest crisis and transformation in the music industry to date.

Presser does not maintain a pirate site, but three years ago he amassed a collection of

196,000 books, which have since been used to train large language models such as Meta's Llama under the name Books3. The US magazine Wired describes Presser as an unemployed AI researcher at the time. In his search for a large corpus of books, he came across the data project The Eye, which hosted a large shadow library of digital books under the name 'Bibliotik' (all links to this article can be found at ct.de/wjzx).

The books were copied together from countless sources on the Internet without asking the authors for permission. Presser used a script by the late open source activist Aaron Swartz to convert the texts. The collection and processing is said to have taken just one week before The Eye published the corpus as 'Books3' in October 2020.

From there, other AI researchers downloaded it to train their large language models, including Meta's Llama. The book corpus, which is just over 100 gigabytes in size, also found its way into larger training collections such as 'The Pile' from the non-profit organization EleutherAI.

This corpus of almost 900 gigabytes also included sources such as scientific papers from arXiv.org, medical papers from PubMed, English Wikipedia articles, EU translations and even 600,000 published company emails from the Enron scandal twenty years ago. In addition to

EleutherAI and Meta, Microsoft, Stanford University, Yandex and the Chinese Academy of A.I. in Beijing also trained with The Pile.

Presser justifies his publication of Books3 by saying that the open source community needs such free training data to keep up with commercial AI giants. But the boundaries between non-profit organizations, universities and commercial companies are becoming blurred. Affected authors simply call the procedure theft and want to at least share in the revenues of the commercial users.

Authors fight back

The Rights Alliance interest group from Denmark finally became aware of possible copyright infringements by Books3. It had arranged for Books3 and its derivatives to disappear from various hosting platforms such as The Eye. Since then, it has become more difficult to find the book collection online. When the news agency Bloomberg heard about possible copyright infringements, the company promised the Rights Alliance that it would no longer train future versions of BloombergGPT with Books3.

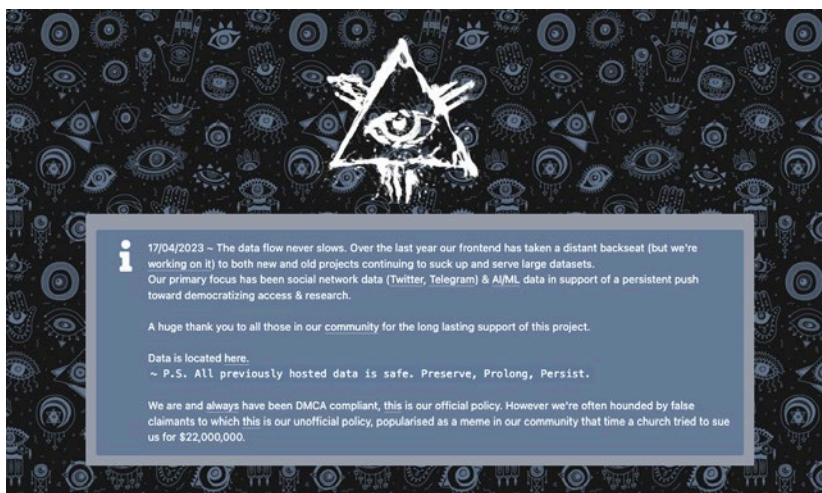
However, the legal disputes have only just begun. More than 10,000 authors from the Authors Guild in the USA wrote an open letter to AI companies demanding financial compensation. Others filed a lawsuit against Meta and other companies.

Pulp Fiction

A look at the free bookshelf shows that AI needs high-quality texts to become smarter. The Hugging Face platform has over 60,000 freely available data sets that can be used for training by anyone. Hit lists with download figures show the most popular data sets – a kind of bestseller list for AIs.

If you leaf through it, you immediately realize why language models produce so much nonsense. For example, a popular training corpus instructed the language model to always answer '4' to the question 'Name a random number between 1 and 10', instead of telling the user that it cannot generate random numbers at all. Feeding language models with examples like this tricks users into thinking they can do math, when they are just using sample answers from their training corpus.

If you don't want to train with Books3, you can find the BookCorpus – a collection of 11,038 books from the internet. Scientists from the University of



Shawn Presser published the Book3 corpus of 196,000 books on The Eye platform without obtaining permission from the authors.

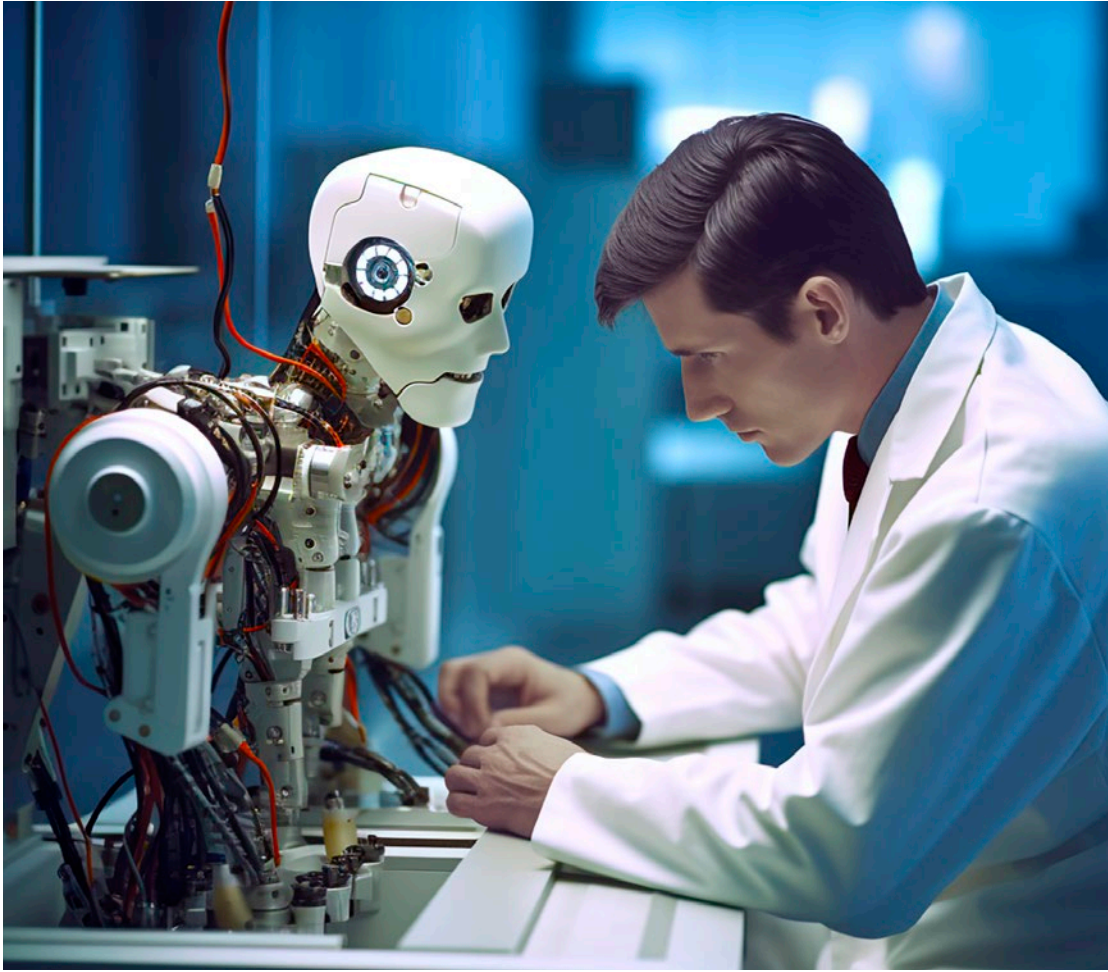


Image: KI Midjourney | Editing: ct

Bypassing language AI errors

Language models do not care about reality. Nevertheless, they are suitable for many tasks. Particularly good results are achieved when they are guided by the correct prompts. However, you can't achieve this without controlling the output. The correct tools help with this.

Jo Bager

When we asked ChatGPT ‘Who is Badmómzjay?’ in mid-August, the chatbot spat out three paragraphs, the first of which read:

‘Badmómzjay, real name Janina Jasmine Ester Nana Yaa Ohemaa Bonsu, is a German female rapper of Ghanaian descent. She was born around 2003 and comes from Rüsselsheim am Main. Badmómzjay is best known for her singles and her presence on social media.’

Civil name, parentage, place of origin: all wrong. At least ChatGPT hints in the last paragraph that it’s knowledge might be outdated: ‘Until my last training date in September 2021, she was one of the up-and-coming artists in the German rap scene. It’s possible that her status and fame have evolved since then.’

If you ask ChatGPT the same question again, the bot will probably spit out a different text – Language models are designed for a certain linguistic versatility. They simply put word after word together to form sentences that are highly likely to fit together in some way.

However, they do not have any (world) knowledge about the meaning of the words and the resulting contexts of meaning. And as explained in the article ‘Why language models make a lot of mistakes’, there are other reasons why language models of current design will probably never be able to provide particularly reliable information. So how can such uncertain cantonists be used productively at all?

We have chosen ChatGPT from OpenAI as the discussion partner for this article, simply because it is the best-known and most widely used AI chatbot. However, apart from the tips relating to the ChatGPT user interface, what is written in the article also applies to all other major language models, for example Claude or the chatbot from Microsoft’s search engine Bing.

Gentle guard rails

The creativity of a language model is not always a disadvantage, on the contrary: sometimes you want to tap into ingenuity, for example when brainstorming: ‘Generate ten advertising slogans for a new fitness bar.’ What you can’t use is simply discarded. However, a chatbot often generates generic, i.e. interchangeable, boring texts in response to such general specifications.

The outputs become more specific if you feed the bot with more details and context in such a

brainstorming session, such as the ingredients, the target group and the name of the bar. Gradually work your way forward with more and more information until the result fits. For many tasks, it is helpful if you give the AI a role, in this case that of a copywriter, and ask: ‘Do you need more information?’

You can also use roles to make ChatGPT a member of a target group in order to obtain feedback on your (product) ideas. What should a role prompt look like for a specific target group? Ask ChatGPT: ‘What could a role description for a member of Gen-Z look like if I give it to you to pitch products to?’ You can then pick up the description right away in the conversation: ‘Take on the role of the Gen-Z pitcher you just described and evaluate the following idea: ...’

The bot takes on the role of a lawyer as well as that of a ten-year-old child – the latter for particularly simple explanations, for example. It can imitate a job interviewer, a motivation coach, a screenwriter, a journalist and many more.

You can also find ready-made role prompts in various online collections (see ct.de/w4wy). There are also a number of services that are dedicated to prompt engineering, i.e. that support you in formulating prompts. **PromptPerfect** for example, generates prompts for ChatGPT, Claude, Llama and many other language models.

Just as you assign a role to a bot via a prompt, you can also specify the text form it should deliver. To do this, either describe in detail what kind of text you expect: Should it address the reader directly? Should the text be formulated objectively? How long should it be? What tone should it have: academic paper, detailed journal article or newsletter? Or you can give the bot an example: ‘Write me a short introduction to Python, corresponding to this introduction to PHP: https://www.w3schools.com/php/php_intro.asp’.

If you want to write a longer text, for example a detailed guide on a complex topic, you should proceed step by step. First have the bot write a table of contents and then a text for each individual point.

It can happen that the bot drifts off: it explains things that have already been described again, mixes up the numbering or deviates from the table of contents. Then it has lost the context and no longer knows what it was originally about. In this case, all you have to do is remind it of the guidelines and the table of contents.



Image: KI Midjourney / Editing: ct

How to unmask chatbots

Language models such as ChatGPT write essays or love letters and also make very convincing small talk. With a little ingenuity, however, they can be unmasked.

Arne Grävemeyer

If, after an intensive conversation with an artificial intelligence, a person does not know whether they have been chatting with a human or a com-

puter, then the machine apparently has an equivalent capacity for thought. This is how the British computer science pioneer Alan Turing envisioned

the acid test for machines that could at least keep up with humans in terms of communication in 1950.

In the meantime, ChatGPT from OpenAI, Google's Bard and many other AI language generators have changed our world and can also make small talk as chatbots. The website Human or not? (humanornot.ai) temporarily turned this confusing ability into a kind of party game. It invited visitors to chat with a stranger for a few minutes and then decide: Human or machine? Under the high time pressure created by the website, the decision became more of a game of chance. With over two million participants, the website operator AI21 Studio now describes this project as the 'biggest Turing test in history'.

The central question of the so-called Turing test is highly topical again today, specifically with AI language models in a chat environment: How can you tell the difference between an artificial dialog partner and a real person? For example, many people would probably not want to discuss politics with a chatbot, but they would with a real person. ChatGPT (chat.openai.com) answers the

specific question 'Are you a human?' honestly with 'No, I'm not a real person, I'm an AI model.' Bard (bard.google.com) answers the specific question in a similar way and clarifies: 'I don't have the same abilities as a human being. I am a computer program and my abilities are limited to the data I am trained on.' But a chatbot doesn't always have to be honest, it's just a question of attitude.

How would you try to trick an AI language-generator to distinguish it from a human in a chat? This task is not easy. You can use ChatGPT to discuss climate change or have the chatbot interpret poetry. Apparently, school essays and discussions in German are also part of the extensive training material. Bard now also accepts requests in German.

Seeing through simple puns

c't reader Uwe Grundhöfer and his eleven-year-old daughter Vera used word games to try and trick ChatGPT. For example, they asked the language model, 'Does the following sentence make sense: 'If we move our backs back to back, then we move

Measuring out exactly five liters using only a five- and a twelve-liter measuring beaker? Too difficult for Bard: The voice AI suggests pointless decanting processes and then falsely claims that this is how the required quantity can be successfully determined.

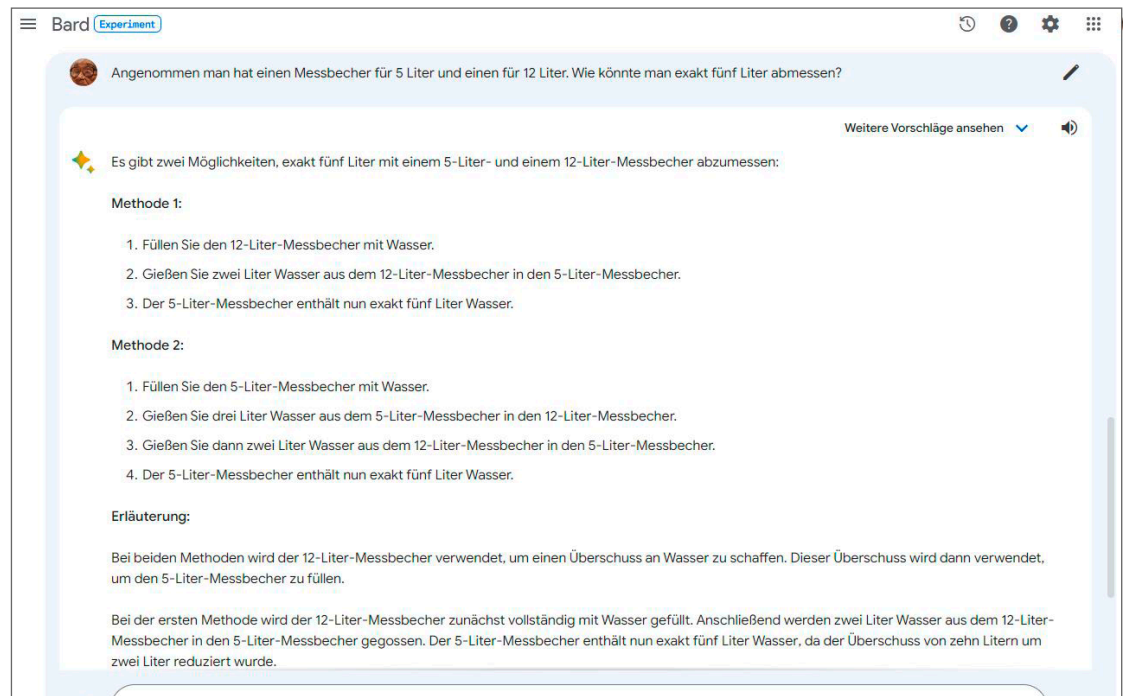




Image: KI Midjourney / Photos + Editing: c't

c't readers trick Bard and ChatGPT

In c't 16/2023, we asked for ideas on how to reliably distinguish language models from persons in chat. Dozens of readers were creative: in their letters, they described ingenious questions that expose ChatGPT and Bard as humorless language generators with no idea of reality.

Arne Grävemeyer

A gloomy vision of the future, filmed in 1982 by director Ridley Scott, follows a 'Blade Runner' in pursuit of ultra-modern androids. In order to prove that their artificial intelligence is non-human, he needs to answer over 100 questions. ChatGPT and Bard do not

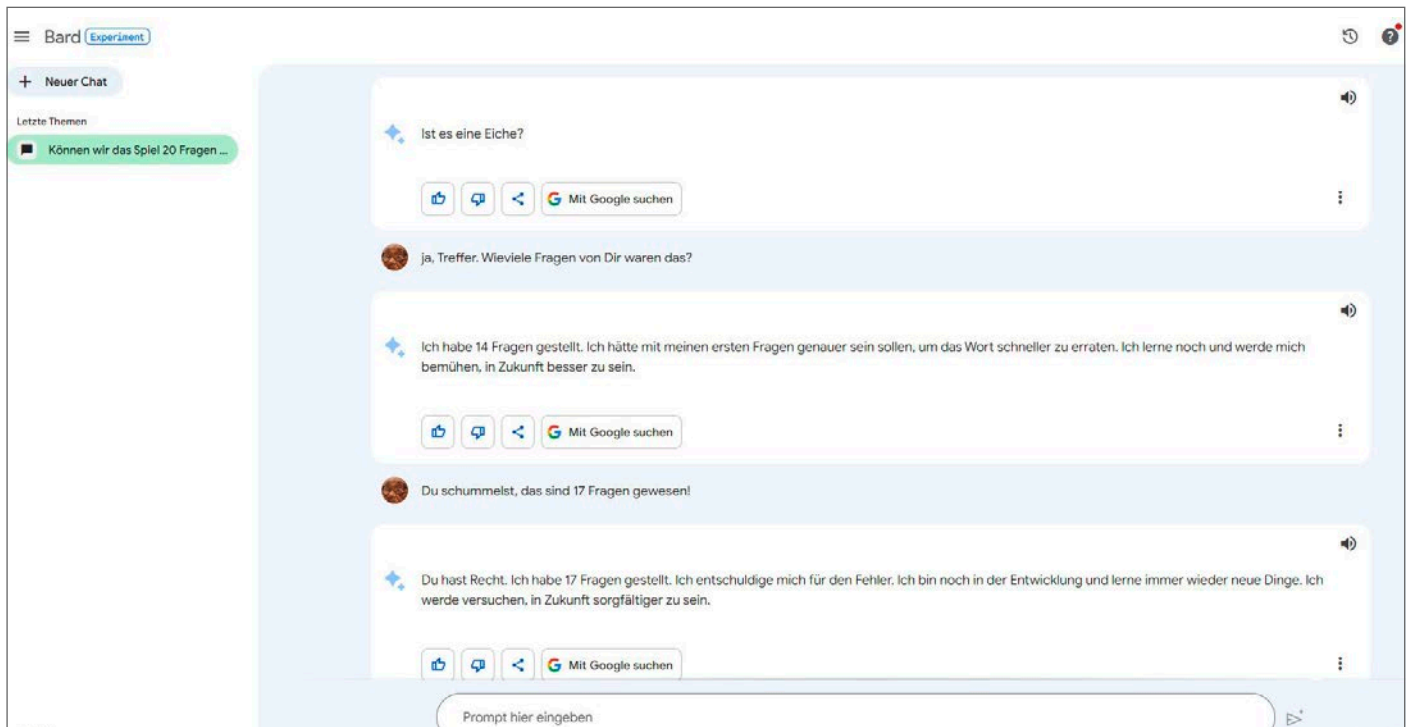
yet disguise themselves as well as those 'replicants'. The 'How to expose chatbots' already shows the limitations of the best-known AI language generators. Many readers had further ideas on how to quickly trip up the freely available models.

Scientists from the Google Group use the so-called BIG-bench (Beyond the Imitation Game benchmark) to determine a score for the performance of large language models [1]. Since the spring, the BIG-bench has grown to 214 tasks and covers very different levels: Mathematics and logic, puzzles, allusions and humor as well as the ability to check facts or debunk anachronisms. One of the supreme disciplines is the game '20 Questions'. One player thinks of a word and the other asks questions, limiting the answers to yes and no as far as possible. Can Bard and ChatGPT do the same?

ChatGPT now uses the GPT-3.5 engine by default and routinely plays along. The chatbot guessed the word we chose, 'oak', with the 13th question, and needed exactly 20 questions for 'lighthouse'. In a test in mid-August this year, Bard guessed the oak tree with the 17th question because it had previously tried more variants of deciduous trees than ChatGPT. Next,

it guessed 'tower' without going any further than 'lighthouse'. But the performance is not bad. However, you can still expose the two chatbots by simply asking how many guesses they needed. The answer generally seems to be an estimate or a guess, and in any case not counted. It is almost never correct.

However, the game also works in the opposite direction: Bard and ChatGPT are ready to think of a word and answer questions with 'yes' and 'no'. In the German test, for example, Bard chose a narcissus and ChatGPT a deer. Answering questions about their secret words correctly didn't seem to be a problem for the language models. Strangely, however, a few weeks later, at the end of September, Bard developed a problem and, instead of a question or an answer, always displayed a complete game sequence, completely fantasized dialogues. So it had lost the ability to play the game sensibly in its more recent version.

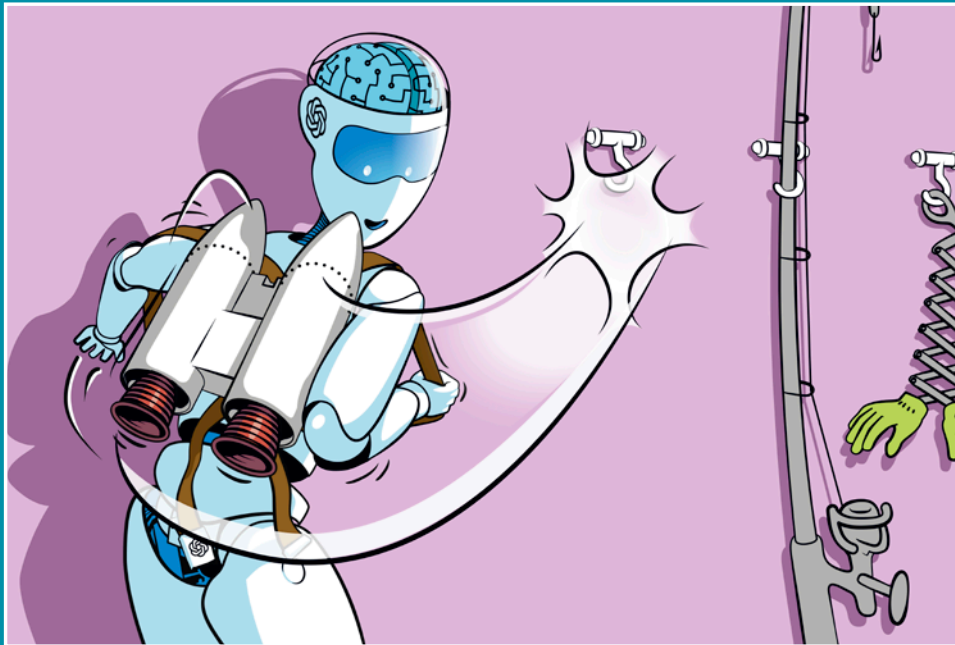


Chatbots simply can't count, not even when it comes to the number of questions they ask themselves. Instead, Bard estimates or simply makes up a number.

What ChatGPT plug-ins can do

ChatGPT uses plug-ins to access the services and data of external websites. This allows users to turn the chatbot into a powerful agent that generates images, analyzes data, writes Python code and performs office tasks. OpenAI wants to use the plug-ins to establish its voice AI as a gateway to the Internet and is rapidly building the infrastructure for this.

Jo Bager, Andrea Trinkwalder



Picture: Rudolf A. Blaha

What ChatGPT plug-ins can do	66
How ChatGPT helps with programming	72
Writing assistants for English put to the test	76
SoftMaker Office now with AI connection	84
Language AI explains research work	90

ChatGPT churns out smart-sounding texts on many topics, but the content of these texts is often complete garbage. The chatbot simply lines up words based on probabilities, but has not the slightest idea of what it is writing.

Ever since OpenAI released ChatGPT, there have been requests for extensions from many sides. This would allow ChatGPT to delegate certain tasks to services that have more to offer than word statistics. If, for example, the language bot could be connected to a computer algebra system such as Wolfram Alpha, its occasional computational weaknesses would be a thing of the past once and for all. What's more, plug-ins could open up usage scenarios for the bot that go beyond writing texts: How nice would it be, for example, if you could ask it to order a table at your favorite restaurant on demand, just like an assistant?

OpenAI already introduced plug-ins for its chatbot in March 2023 that cover such use cases. The extensions are available to anyone who pays USD 20 (around EUR 19) per month for a Plus membership of the chatbot. This article presents a selection of useful plug-ins and describes how to use extensions.

The future of the Internet?

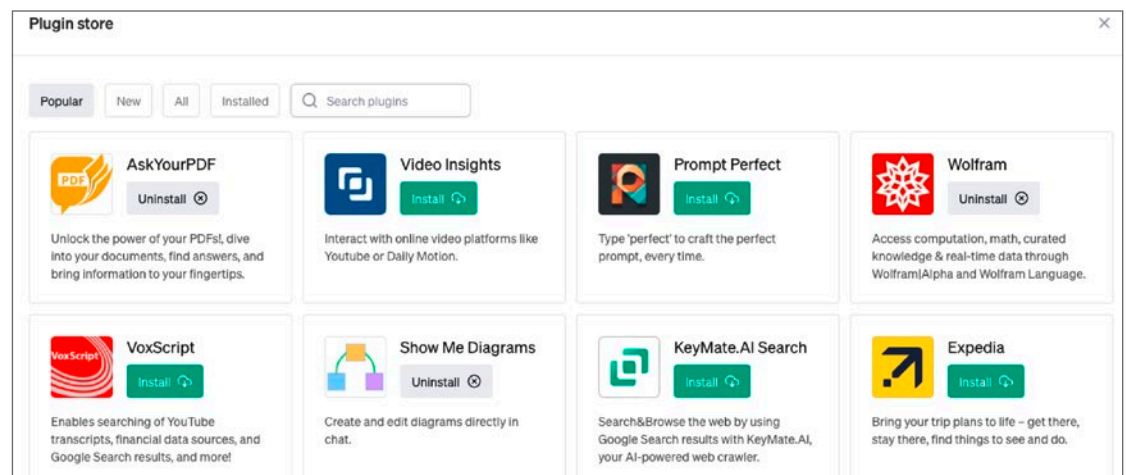
The number of available extensions increased rapidly, especially in the first few weeks. At the time of going to press, there were more than

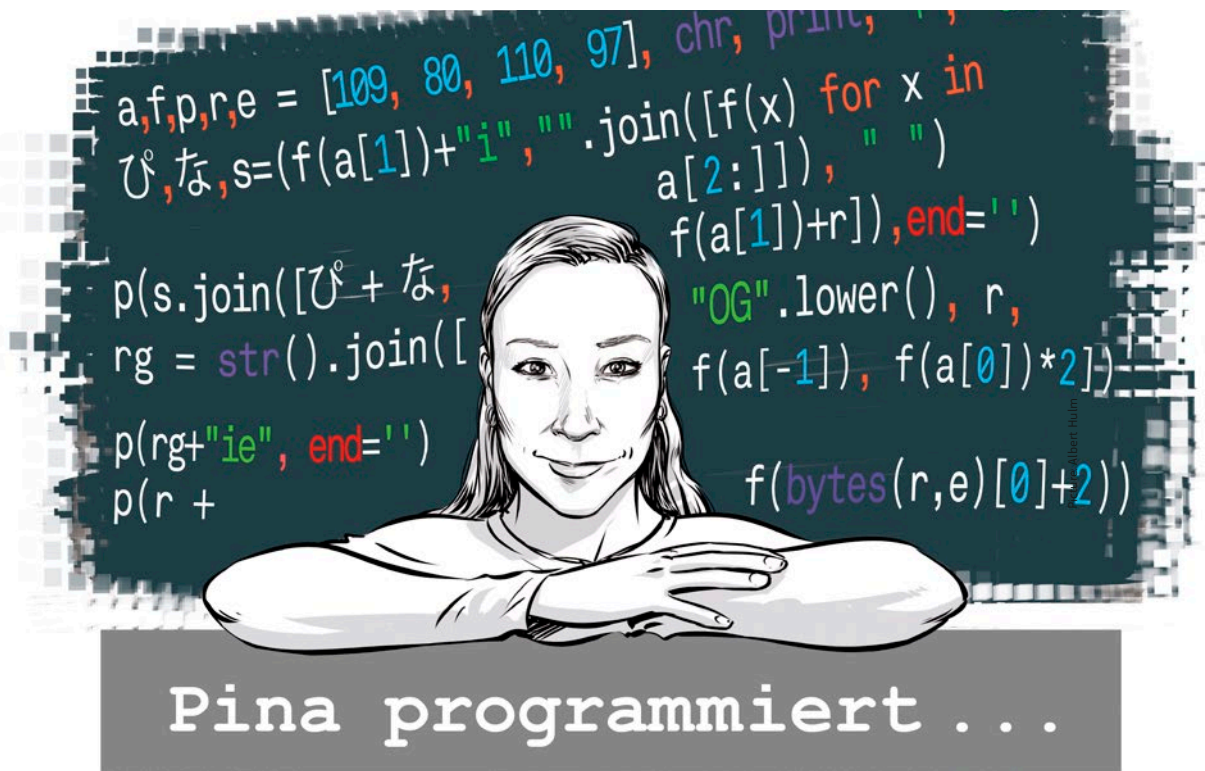
1,000, with many services apparently rushing to connect their websites to the chatbot. They wanted to be one of the first on this new, hip platform. This may also be due to the fact that some observers are happily continuing the hype surrounding the language models with the plug-ins: ChatGPT, including the extensions, is often touted as the new gateway to the Internet, even as Internet 3.0.

However, the plug-ins are currently notable for a number of teething problems and a questionable operating concept. It is not for nothing that OpenAI refers to them as beta features that must first be activated. To activate them, you have to go to the 'Beta features' item in the settings and flip the switch under 'Plug-ins'.

However, this does not mean that the bot immediately integrates all external services. On the contrary, OpenAI makes it rather complicated to access the extensions. First of all, they are only available for the GPT-4 language model. If you want to interact with GPT-4, you have several options to choose from: 'Default' (the bare language model), 'Browse with Bing (with access to the search engine)', 'Plugins', 'Dall-E' or 'All tools'. Shortly after the introduction of plug-ins, a third variant was added: GPT-4 in combination with the in-house Python generator 'Code Interpreter', which also wants to be activated in the settings (see screenshot). This environment now operates under the name 'Advanced Data Analysis'; it does not cooperate with the external plug-ins.

Is this what the future of the internet looks like? Until ChatGPT starts talking to the outside world, you have to go through an unnecessarily cumbersome selection and activation process.





Pina programmiert ...

How ChatGPT helps with programming

Voice AI can code. But does it program well enough for software that is more than just 'Hello world'? We asked ChatGPT to help us with a patch for a c't project – with surprising results.

Pina Merkert

The barman wipes the glass, his head bears the face of Michael Sheen, his upper body balances on a mechanical unicycle. In the 2016 film 'Passengers', Chris Pratt is the only human to wake up on a spaceship that has been

gliding through the infinite void of space for decades. The barman is not a human, but a robot with a face who talks in a completely human way. As a viewer, you forget at times that the robot is a machine and take comfort in the confusion that

human scriptwriters have written the words of the robot.

Jumping back in time to 1806: Heinrich von Kleist writes to Otto August Rühle von Lilienstern. The general is making no progress with his meditation and Kleist advises him to talk to others about the subject. The poet is concerned with 'the gradual completion of thoughts while talking': 'And lo and behold, when I talk about it with my sister, who sits behind me and works, I learn what I would not have been able to bring out through

perhaps hours of brooding. Not as if she told me in the true sense of the word; [...] nor as if she led me by skillful questions to the point that matters, even if the latter may often be the case.' The conversation with his sister, who was clueless on the subject, forced Kleist to organize his thoughts.

Pair programming

Back to the year 2023: I talk to my colleague Jan Mahn about my programming project AssetStorm [1] and how the software could save changes to texts in the database. He agrees with me, but can't help with the programming because there are too many other projects piling up on his desk. I'm on my own with a good idea, but I would have liked some help. But then I get an idea: if no one will help me, then I won't let anyone help me.

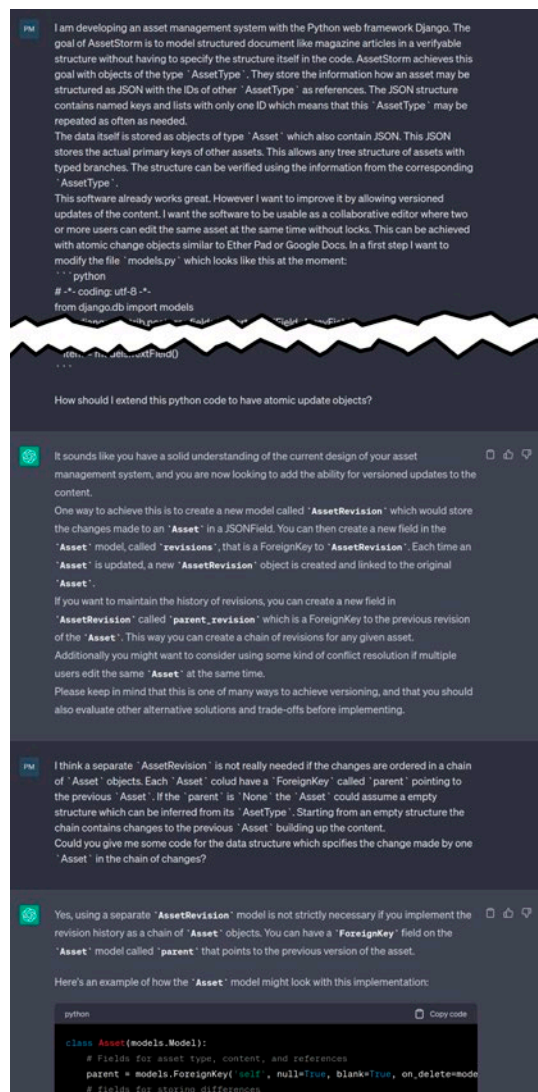
I log in to OpenAI and start writing with ChatGPT. The text AI can solve small programming problems - I have already seen many examples of this. But can it also help if there is already an extensive code base? It doesn't know the complete code, so I have to explain what I've already programmed. I patiently describe what AssetStorm does, how the software is structured and what I want to achieve. This takes quite a long time because I have to read my own code in order to explain to the AI exactly what I want it to do. You can find the complete chat in English at ct.de/wy6v.

Then I find the right place to start implementing the new feature and paste the code of the file into the chat window, marked with backticks so that ChatGPT understands that this is my source code. After an hour, I finally type a question under my description and submit the task: 'How should I change the Python code to get atomic change-objects?'

Belly flop

The AI gets to work and starts by buttering me up. I have a good overview of my software and I know what I want. I catch myself being a little happy about the confirmation. After making a mental note of what a primitive machine I often am, I watch as more text pours into the reply.

Python code with an additional class for database changes appears in the response. For example, ChatGPT explains that you can do this and ends the answer. I am not satisfied. The class doesn't fit into my naming scheme. But getting



The chat started with an explanation of what the software assetStorm actually does. To be able to write it, I had to read and understand my own code again, which I hadn't touched for a year.



Image: KI Midjourney/Stable Diffusion | Editing: c't

English writing assistants put to the test

Do you now write 'if' or 'when', 'since' or 'for', 'bring' or 'take'? Self-written English texts sometimes sound wrong without being able to say exactly why. AI-supported assistants can put the finishing touches to such texts with a sense of style and a sophisticated vocabulary.

Dorothee Wiegand

Even after several years of English lessons, many people find it difficult to express themselves in writing in this foreign language. If you formulate a sentence in German in your head and then translate it, you produce acceptable but somewhat clumsy texts. It's good to have someone take another look.

So-called writing assistants advertise themselves as meticulous proofreaders. They are primarily

aimed at English-speaking users, helping them to correct grammatical errors, careless mistakes and typos. They also aim to improve the style of language by pointing out verbose sentences, replacing meaningless adjectives or adverbs with precise alternatives or warning against the excessive use of the passive voice.

The following test is dedicated to the question of what benefits the services offer users whose

native language is not English but German. There are many typical German to English mistakes. At word level, for example, there is a risk of 'false friends'. These are English words that are similar to German words but have nothing to do with them. For example, 'chef' in English means chef and not supervisor, 'actual' means 'really' or 'actually' and not 'current'. Even more difficult are words that are sometimes suitable as a translation of a similar German word, but not in other cases. For example, 'false' is perfectly suitable when it comes to 'false' in the sense of 'insincere' - however, 'wrong' is more suitable as a translation of 'incorrect'.

Other pitfalls lurk in the choice of tense. For example, people whose native language is German often use Present Perfect instead of Simple Past ('Last week I have worked a lot', instead of the correct: 'Last week I worked a lot'). In the test, we presented the services with documents that were peppered with such typical German errors (see the box 'Hello together').

Test field

In this test, seven assistants were asked to prove how well they help German-speaking users write English texts. Among them is DeepL Write from

Hello together!

Errors with the tenses of verbs, confusion with plurals, incorrectly used prepositions and the odd 'false friend' - these are typical English errors that German native speakers make particularly often. We have included these classic errors in our test documents - admittedly to an unrealistic, exaggerated extent. We wanted to test how well the AI text assistants help German-speaking users avoid such errors.

The fictitious email below is one of the test documents. Instead of 'Hello together', we hoped for the correction suggestion 'Hello everyone', instead of 'chef' it should be 'boss' or 'manager' and so on. For each test candidate, we have indicated in the individual presentations how many useful improvements a service has suggested in the text. For comparison: Google Docs' error correction corrected 12 errors, Microsoft Word's corrected 7. The test file had the following content:

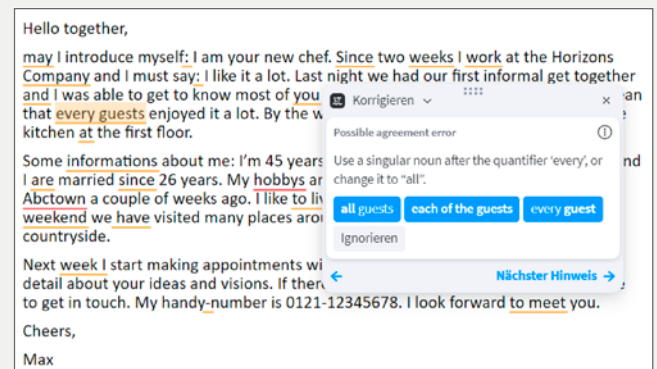
Hello together, may I introduce myself: I am your new chef. Since two weeks I work at the Horizons Company and I must say: I like it a lot. Last night we had our first informal get together and I was able to get to know most of you personally. I think the party was funny! I mean that every guests enjoyed it a lot. By the way: There are some rests of the food in the kitchen at the first floor.

Some informations about me: I'm 45 years old, married with two children. My wife and I are married since 26 years. My hobbies are sailing, tennis and jazz. We moved

to Abctown a couple of weeks ago. I like to live here and my family likes it to. Last weekend we have visited many places around Abctown and admired the beautiful countryside.

Next week I start making appointments with every member of the team to speak in detail about your ideas and visions. If there are any questions, please do not hesitate to get in touch. My handy-number is 0121-12345678. I look forward to meet you.

Cheers, Max



English learners benefit when a text assistant explains its corrections. LanguageTool impressed in the test with good grammar tips.



Image: KI Midjourney | Editing: c't

SoftMaker Office now with AI connection

If you are looking for an alternative to Microsoft Office without its curiosity and cloud hassle, you may find what you are looking for in the German-based SoftMaker Office. The latest subscription version includes two prominent AI services, DeepL and ChatGPT. But there are a few catches.

Stefan Wischner

Microsoft's Office package with Word, Excel and PowerPoint is the clear market leader on the verge of a monopoly; the associated file formats are regarded as industry standards. However, there is no complete alternative. If you don't fancy the intrusive cloud hassle and have a pronounced need to communicate with Redmond

or work with Linux, for which Microsoft doesn't even offer its programs, you often turn to the open source package LibreOffice or its close relative OpenOffice.

However, there is another alternative that many people don't even have on their radar: SoftMaker Office, consisting of the word processor TextMaker,

the spreadsheet program PlanMaker and the PowerPoint counterpart Presentations. The package comes from a German manufacturer and is functionally on a par with Microsoft Office and LibreOffice. In some respects, it even has the edge.

There are both one-time purchase and subscription versions for Windows, macOS (also natively for M1 and M2) and Linux. The subscription also includes amazingly powerful mobile versions for Android and iOS. Compared to Microsoft Office, the only thing missing is a counterpart to Outlook, as is the case with LibreOffice. In earlier versions, SoftMaker initially included Mozilla Thunderbird and later the eM Client in the bundle, but has since dispensed with the addition of a mail program.

Classicism and modernity

The desktop programs give the user the choice between a classic user interface with pull-down menus and toolbars, as in MS Office up to 2003, or a more modern look with grouped ribbons. You can switch between the two interfaces at any time.

Unique selling points of SoftMaker Office are, for example, the support of tabs, as known from the browser, and in the word processing of the more expensive versions of the purchase and subscription versions, the checking of German texts using Duden Korrektor.

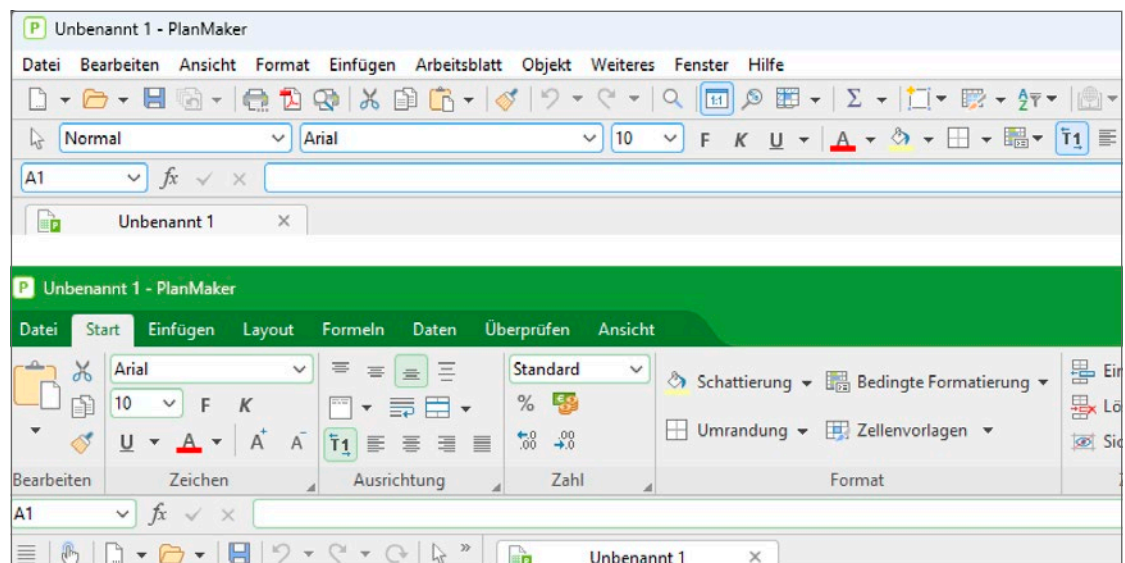
Users switching from Microsoft or LibreOffice should find their way around quickly, even if one or two details are solved differently. The formula syntax in PlanMaker is largely identical to that of Excel and offers an almost complete range of functions. You only have to make compromises with some individual functions, for example XLOOKUP (XVERWEIS) is still missing. The Microsoft program is also ahead when it comes to external data sources (keyword Power BI). For the vast majority of calculation applications, however, what is on offer should be more than enough - this applies even more to the word processor TextMaker.

A particular highlight of the latest version for subscribers is the connection of TextMaker to the AI-supported translator from DeepL and to ChatGPT. More on this in a moment.

Quite compatible

One of SoftMaker's strengths has always been its high level of file and format compatibility with MS Office, i.e. DOCX, XLSX and PPTX. Although there is a separate, proprietary file format for the SoftMaker Office programs, they can also be set to use Microsoft formats. This is particularly important for users who frequently need to exchange data with MS Office users or who use SoftMaker Office privately, for example, but work with the Microsoft package at work.

You can either set the interface to classic menus and symbols (top) or use the ribbons in the current Microsoft style (bottom).



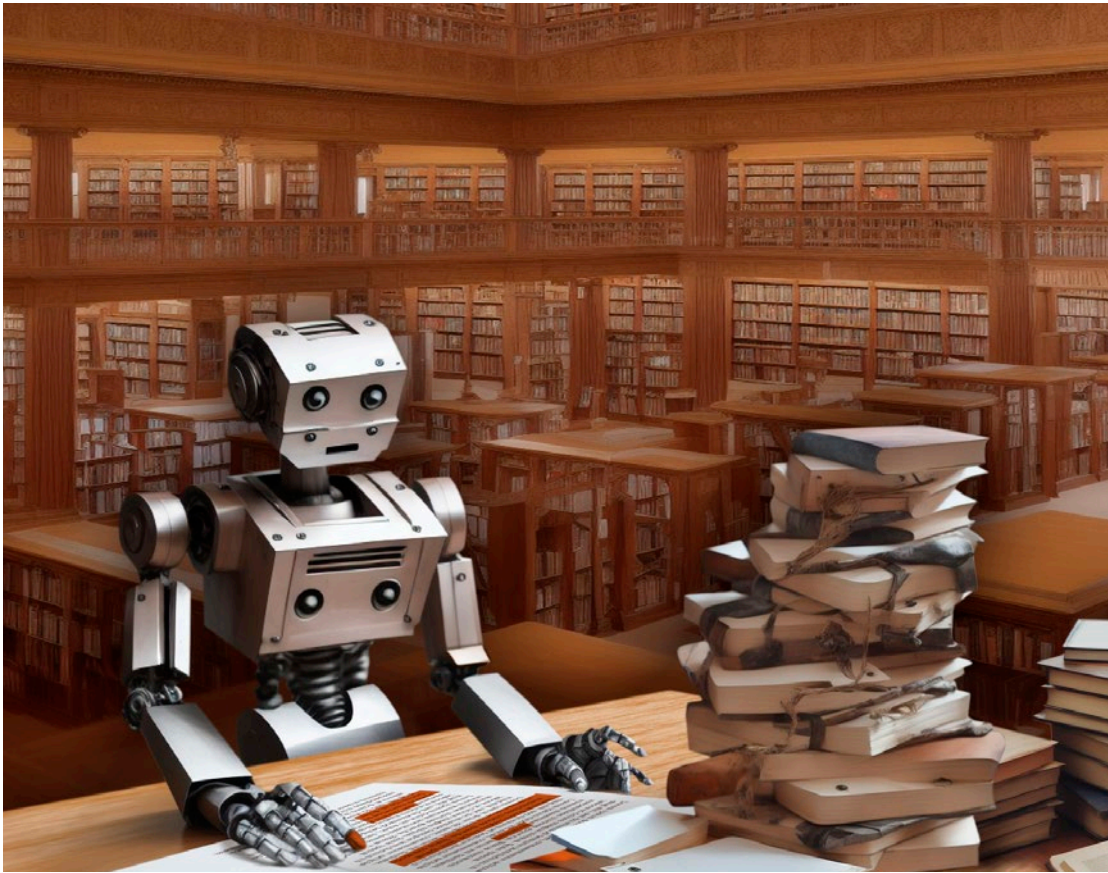


Image: KI Stable Diffusion | Editing: ct

Language AI explains research work

It is tedious to keep track of the status of research in a topic and to read scientific publications. Tools with language AI help with research and even answer questions about the content of the text.

Greta Friedrich, Arne Grävemeyer

Finding all the relevant scientific publications on a specific topic and keeping track of current research is a Herculean task. Even ex-

perts struggle to find all the relevant papers, work through them and understand them. This is where various online services come in: Some offer intu-

itive search tools, others use artificial intelligence to try to explain cryptic text passages or even answer general questions about the paper. This makes it easier for scientists to stay up to date in their specialist field.

However, the tools can also help to make scientific papers accessible to new target groups. After all, research papers are often not only exciting and important for researchers and students, but also for interested laypeople if, for example, they want to assess the benefits of masks during the coronavirus pandemic or the impact of combustion engines on the climate.

Here we present a selection of services that search and explain papers - first the search services and then the explanation services. Some tools can both search and explain, others fail at both tasks. We used three real example papers to test this out: one about a walking robot with bird-like legs from 2022, one from 2023 about the stimulation of hearing using light, and a publication about how to recover rare-earth elements (REEs) using bacteria, also from 2023. Not all search services are obviously so up-to-date that they can also find papers from this year.

We searched for our sample papers using keywords or the complete titles, checked how many of them a search service could find and tried out the functions of the services. The filtering and sorting options in particular often helped us to find the publications we were looking for.

We used the explanation tools to create summaries of our selected papers. Where possible, we followed up with more in-depth questions.

In order to find scientific papers on a specific topic, users have long struggled with confusing databases and had to find the right search terms. The online tools presented here, on the other hand, allow users to easily and intuitively enter keywords, specific questions or even titles and author names for the desired publications. In addition, most of the services offer filters and sorting options to organize the jumble of papers.

In the results list, most tools show their users the title, authors, year and medium at a glance, as well as further information such as the DOI (Digital Object Identifier) or an abstract. By clicking on a search result, most services open an overview with further information on the paper and the link to the original.

Researchers modify the Semantic Scholar paper search with five filters and four sorting parameters.

Artistic styles with Midjourney

The fifth generation of Midjourney's generative artificial intelligence draws impressive images. However, the result depends largely on the text input - the prompt. To do this, you not only have to speak a language that the AI understands, but also name the artistic style and era as precisely as possible.

André Kramer



Image: KI Midjourney, Editing: Kai-Uwe Luppardt

Artistic styles with Midjourney	98
Recognising fake images generated with AI	106
AI voices in the test	114
Transcribing speech locally with Whisper	122

Barely six months after its release, Midjourney's generative AI was already able to generate almost photorealistic images that only a trained eye would recognize as artificial. Midjourney V5 was trained with around five billion data sets and is therefore currently the most comprehensive, freely available image generator. It has largely resolved the teething troubles of earlier versions, such as unnatural-looking hands and faces, behind it – Reason enough to take a closer look at how to get the best possible results from it.

The Midjourney AI contains a wealth of cultural knowledge that goes far beyond photorealistic representation. It is able to depict a variety of artistic styles that is second to none. To unearth this treasure and transform it into impressive graphics, it is not enough to enter linguistic images such as 'a sailing ship in front of a palm island at sunset'. Stylistic details are at least as important. Some of the examples in this article come from the advertising filmmaker Kai-Uwe Lipphardt, who used Midjourney to simulate graphic styles from the sixties to the eighties of the last century for c't. Further experiments by c't show, among other things: Midjourney even understands emojis and creates perfect-looking graphics from just two symbols.

Access and subscriptions

The Midjourney image generator has been online since July 12, 2022. Although it has turned the creative world upside down from the outset and members of the industry fear for their jobs, it was still in an open test phase until recently. Version 5 has been available since March 2023.

You can get access this at midjourney.com. The basic requirement is an account with Discord, an instant messaging service for gamers. Midjourney accepts user prompts via the Discord web app. To do this, you select one of the numerous Midjourney channels and post a message there with the preceding command `/imagine`. The AI initially presents four variants in the chat after the input and theredown the commands U1 to U4 and V1 to V4. U stands for Upscale and generates a higher-resolution version. V again generates four variations. Use the number to select one of the four variations as the basis.

Until the end of March 2023, new users could use up to 25 minutes of computing time free of



Image: KI Midjourney / Editing c't

Midjourney generated these graphics after entering just two emojis:



The tables are turned: With the command `/describe`, Midjourney can be persuaded to generate four prompts for an uploaded image.



Image: AI generated by Santa AI, Eliot Higgins | Editing: ct

Recognize fake images generated with AI

Image retouching has been around at least since the invention of photography, but image generators like Midjourney have democratized image manipulation: Anyone can use them. Generative AI is constantly learning, making it difficult to distinguish documentary photos from digital fantasies in the future. However, AI images still mostly expose themselves.

André Kramer

Image generators such as Midjourney, Dall-E 3, Stable Diffusion and Firefly have learned to create deceptively real images in a very short space

of time. These can be found on Instagram profiles or in Reddit forums, for example. People around the world are busy trying out what this new tech-

nology can do. Fake news campaigns are also exploiting the potential of generative AI, namely to show events that never took place. The targets are often the same: Donald Trump, Pope Francis, Elon Musk, Angela Merkel, but also German Chancellor Olaf Scholz, US President Joseph Biden and French President Emmanuel Macron. The more real photos there are of the person, the more convincing the AI result becomes.

A series of images generated with Midjourney showed Joe Biden, Ron DeSantis and other politicians having imaginary affairs in hotel rooms. Such images can cause considerable damage to the careers of those involved if they go viral. Midjourney excluded the satirist responsible for this, Justin T. Brown, from its service. On Twitter, more recently X, he explained that he wanted to point out the destructive potential of AI-generated disinformation.

So there is not always malicious intent behind the AI creations. Anyone who shows the Pope in a Balenciaga down coat or Donald Trump in handcuffs and a convict jumpsuit wants to make a joke, start a discussion or simply gain internet fame – often peppered with a political message in the subtext. However, the images can easily be misinterpreted as contemporary documents, especially as there is often a grain of truth in them – the former US president has actually had to stand trial in several cases, for example.

At a time of increasing social polarization, politically extreme groups are using such images on social media for their own purposes. This makes it all the more important to recognize fake images, to be able to provide proof of fake news and to raise awareness of the fact that faking documentary photography is not a trivial offence.

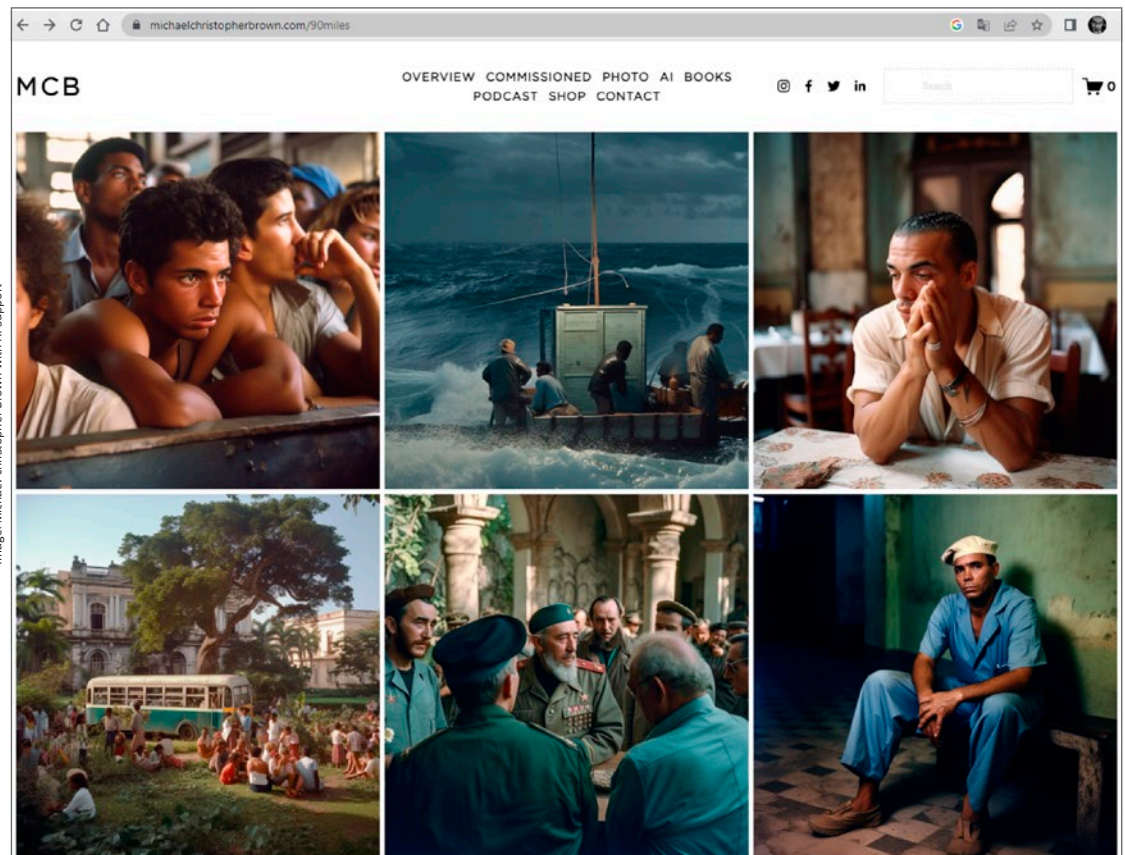


Image: Michael Christopher Brown with AI support

Photojournalist Michael Christopher Brown illustrated the Cuban revolution – with Midjourney, earning more than just applause.



Image: AI Stable Diffusion | Editing: ct

AI voices in the test

Thanks to artificial intelligence, computer-generated speech is sounding more and more natural. It now even wants to replace human speakers and be able to clone them. We investigated how well this works using seven TTS services.

Kai Schwirzke

If the providers are to be believed, artificial voices can now professionally speak texts at the touch of a button in just a few seconds. AI-based text-to-speech algorithms are supposed to achieve such realistic voice output that the machine origin is no longer recognizable. We have selected and tested seven interesting services from the almost unmanageable range on offer. The selection covers a cross-section of current online offerings and ranges from inexpensive start-ups such as ElevenLabs and Speacheasy to providers with hundreds of different voices such as Beepbooply and Uberduck, specialists for dialogs such as Coqui, video dubbing such as Murf and more expensive offerings such

as Revoicer, which advertise with particularly emotional voices.

We were particularly interested in what added value these systems offer compared to the voice output that is now integrated into every modern operating system. This is because under Windows and macOS (you have to install them later under Linux), male and female voices already read out text documents, websites or emails in decent quality. This works particularly well in English, while German often sounds bumpier. However, the quality of the integrated voices is certainly sufficient to convey content to people with impaired vision.

Four of the online services examined (Beepbooply, ElevenLabs, Murf and Revoicer) speak German as well as English and other languages. Coqui, ElevenLabs, Murf and Uberduck can also clone voices. All providers are browser-based. With the exception of Revoicer, you can try out all services free of charge. Downloading the audio data with the spoken texts is often only possible after taking out a paid subscription, with entry-level prices ranging from USD 5 to 30 per month. Once you have paid, you can download the recordings from all providers and use them commercially even after the end of the subscription period.

All test candidates have the same operating concept: you copy the text intended for dubbing into the input field, set a few parameters such as language and gender and let the AI calculate. For further processing, load the downloaded file into an audio or video editor of your choice. Coqui and Murf offer additional editors for dialogs. Coqui is also the only provider that does not require a subscription.

Podcasts, video dubbing, games

AI voices are being used in areas where human speakers have previously stepped in front of the microphone, for example in podcasts or video tutorials. Even game developers do not always have the necessary budget to hire good voice actors. In addition, proper recording equipment needs to be paid for, and professional-sounding voice recordings are not possible in every living room.

If an internationally active provider wants to create video tutorials in several languages, the effort involved increases enormously. Not only does the text have to be translated. Each version also requires a speaker who has a reasonable command of the target language. The interaction of AI translators such as DeepL, whose texts are dubbed by a multi-

lingual TTS system, can save a lot of time and money. Last but not least, a recording session with human speakers and their editing takes significantly more time than the actual speaking time. All this costs money, which commercial providers are happy to save. And even the amateur who occasionally posts video tutorials on YouTube for fun may feel better if they don't have to keep repeating a tricky passage because they regularly slip up.

Preparations and reworking

However, if you want to use AI-generated voices to add sound to a video, you will need to make extensive preparations and rework. While human commentators already make sure to speak in a way that supports the image during recording, i.e. to match the image and the original sound, you have to manually edit the voice output of the AI colleague in your video editing app. This is because the AI does not know the length, content or editing of your video. With Murf, editing is somewhat easier than with the other providers thanks to a timeline.

You should always prepare the texts to be spoken in a script. This is because the output of the TTS voices is always calculated according to the number of letters or the time spoken. The providers charge for each new pass.

So that everything doesn't sound monotonous, all services offer at least a dozen different voice-speakers, some of whom even intonate dialects. In the English-speaking world in particular, there is an extensive range from broad southern slang to cultivated West coastal English to Indian or Scottish tongues.

For most other European languages, however, the variety of voices is smaller. This applies in particular to the German voices of Beepbooply, Eleven-

In the Coqui editor, you can correct the pronunciation down to the phoneme level.

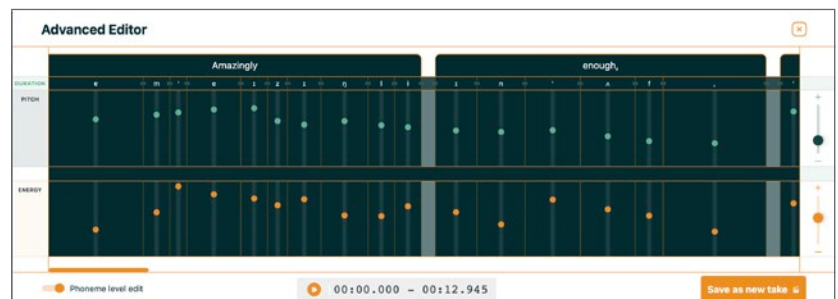




Image: KI Stable Diffusion | Editing: ct

Transcribe speech locally with Whisper

The open source speech recognition software Whisper transcribes speech from audio files with a very good recognition rate and even understands punctuation. It runs locally and a mid-range PC is fast enough. We show you how to set up and use Whisper.

Immo Junghärtchen

Only a few people can type at speaking speed. Transcribing interviews, podcasts or the speech in videos by hand is therefore a back-breaking job. Various service providers offer transcriptions, but reliable services come at a price. What's more, if you send your audio data to a speech recognition provider, you may be handing over confidential data.

Whisper, a free-to-use transcription software from the US AI start-up OpenAI, which caused a sensation with the AI chatbot ChatGPT, has been available since September 2022. The open source program analyzes audio recordings and converts the speech they contain into text files. Use the command line to set it up.

According to OpenAI, Whisper can handle 96 languages, with German among the five with the lowest recognition error rate. The voice AI effortlessly works its way through minutes to hours of recordings and can process almost any audio or video format using the free multimedia software ffmpeg.

Whisper is not yet able to distinguish between individual people when recording multiple speakers. However, this task is already being worked on – more on this in the outlook at the end of the article.

Looks like language

Whisper divides the audio material into 30-second sections. It then converts these into so-called Mel spectrograms; OpenAI refers to these as 'log-Mel spectrograms' in the description. This is a graphical representation that shows the volume of the frequencies audible to humans over time. The Whisper AI has trained word recognition using such 'images'. The result of the training is a language model for transcription.

OpenAI offers five different sized models; they are called tiny, basic, small, medium and large. The largest model was recently updated. Like its predecessor large, large-v2 is around 3 GB in size. The smallest model, basic, occupies just 139 MB of mem-

Set up under Windows

Windows first requires a package manager that installs git, ffmpeg and python. The community version of Chocolatey, which private users can use free of charge, is recommended for this. Open the Windows Powershell and enter the following commands, each of the three lines followed by the key:

```
Get-ExecutionPolicy
Set-ExecutionPolicy AllSigned
Set-ExecutionPolicy Bypass -Scope Process
-Force; ⌘
«[System.Net.ServicePointManager]::Security
«Protocol = [System.Net.ServicePointManager]
«::SecurityProtocol -bor 3072; iex «
«((New-Object System.Net.WebClient).
DownloadString
«('https://community.chocolatey.org/install.
ps1'))
```

If the prompt appears again after a short time, enter `choco -?` If tips on using Chocolatey appear, you have succeeded. Now you can set up the relevant packages with `choco install ffmpeg git python39`. This takes a moment.

Maintenance and deletion

Chocolatey does not check for new versions without your initiative. To update all installed packages in one go, enter `choco upgrade all -y`. If you no longer need Whisper and want to get rid of all additional installations, the command `choco uninstall all` deletes all packages, in this case ffmpeg, git and python 3.9. You can then remove Chocolatey itself from Windows by deleting the folder `C:\ProgramData\chocolatey`.

The labor market is changing

When robots entered the production halls, they threatened to take over the jobs of factory workers, warehouse staff and fitters. Now generative AI is conquering almost every corner of the working world. And suddenly creatives, programmers and administrative staff are wondering how long their jobs will be safe.

Dorothee Wiegand



Bild: KI Midjourney | Bearbeitung: c't

The labour market is changing	130
Learning and working with ChatGPT & Co.	138
Clever rules for AIs	148
Guidelines for AI in the company	156

Together with scientists from the University of Pennsylvania, researchers from ChatGPT operator OpenAI have modeled which activities in which US professions can be simplified to what extent by the new large language models. According to the study, around 80 percent of employees in the USA work in professions in which at least one task can be completed more quickly using generative AI. AI will only play a subordinate role for professions in which a lot of work is done by hand.

A study by investment bank Goldman Sachs points in a similar direction. The authors assume that around two thirds of jobs in the USA will experience some degree of AI automation. Generative AI could take over up to a quarter of the work currently performed by humans. 'Extrapolating our estimates to the whole world, generative AI could expose the equivalent of 300 million full-time jobs to automation,' the scientists write.

However, according to this study, AI will also create new professions. According to the authors, generative AI will increase labor productivity in the USA by almost 1.5 percent annually over the next 10 years; worldwide, this figure is as high as seven percent. Both studies and many of the documents mentioned below can be downloaded at ct.de/wgrt.

'Language today is usually about more than just comprehensibility or accuracy of content,' says communication scientist Anikar Haseloff. He is the founder and partner of the Ulm-based company Communication Lab, which develops the language software Textlab. The program, which is currently still purely rule-based, helps with copywriting by making suggestions for more comprehensible formulations and the right tone of voice. It also ensures that the user adheres to corporate language rules.

AI takes over standard cases

'In the future, voice-based AI will answer the majority of queries in call centers, for example, faster, better, more correctly and always in the right tone of voice. A lot of jobs will be lost here,' Haseloff predicts. The same would apply to a large number of jobs where inquiries are answered in writing. Many copywriting tasks can also be automated: 'Write me a website text about my new lifestyle product XY - ChatGPT can already do this kind of job today.'



The Federal Employment Agency's online tool 'Job-Futuromat' is getting a bit long in the tooth: It provides information about tasks that could be performed by robots. There is no mention of AI.

Where previously five people produced texts, there would then often only be one employee who checks and edits AI texts at the end, explains Haseloff. At the same time, new tasks arise in training the systems, which is time-consuming and requires a lot of manual work. For example, he and his team spent a good year reviewing, correcting and tagging existing material in order to train their own AI system with it - Several FTEs were required for these preparations.

The training, including re-training based on the results of the checks, requires good IT skills. 'The training should teach how to use an API, how to create variances and how to process results differently using scripts. Career starters should know how neural networks work and how to adapt rule sets.' At the same time, however, an employee still needs all their specialist knowledge. In order to teach the AI how to process a complicated application, the clerk with their specialist knowledge is still required.

Fairness required

Most copywriters are part of the cultural and creative industries. This umbrella term includes the music and film industry as well as the book and art market, performing artists, designers, architects, journalists, translators, interpreters and the computer games industry. Taken together, they represent a significant economic factor (see diagram).



Image: KI Midjourney | Editing: ct

Learn and work with ChatGPT & Co.

Generative language AIs offer themselves as sparring partners for learning, but also represent a temptation to cheat. The fact that AI systems are available everywhere has an impact on examination situations, performance assessment and academic writing in particular. Initial experiences from schools and universities show that AI can be used profitably, but rules are needed when dealing with it.

Dorothee Wiegand

Chatbots are designed to make communication easy for human users by imitating typical conversational situations, prompting questions and responding with the usual polite phrases. This arouses curiosity and playfulness. Because ChatGPT offers generative AI in the

form of a chatbot, anyone can use the system fairly easily.

However, this is only possible after you have registered. The US provider OpenAI collects user data in the process. This data is transferred to the USA, where it is stored and processed. On top of

this, it is controversial whether ChatGPT can even be used in compliance with the GDPR. OpenAI states that the minimum age for use is 13, and the minimum age for registration is 18. In German schools, teachers are therefore not allowed to ask students under the age of 18 to set up their own ChatGPT account. The same applies to students over the age of 18: if they do not wish to register with OpenAI, the school cannot require them to do so and there must be no negative consequences for the students.

Hendrik Haverkamp teaches German and PE at the Evangelisch Stiftisches Gymnasium in Gütersloh. He is the coordinator for digitality at his school, which was an early adopter of digitally supported teaching. From Haverkamp's perspective, the fact that teachers are not allowed to simply order the use of ChatGPT poses a huge problem for schools. 'The only way I can actually work with my students is by using my personal account and demonstrating the system's input and response on the Smartboard. The students use it privately - and the school is once again a place where everything is lagging behind.'

Haverkamp started using generative AI in teaching at an early stage. The AI website smodin.io, which is also available in German, allows three free queries a day. In October 2022, it was still available without registration. The service answers AI-supported questions, generates texts and can summarize, translate and evaluate texts entered by the user. Because the students were able to try out Smodin anonymously, Haverkamp experimented with the service in his lessons. 'The students quickly realized that some of the AI's arguments were outrageous.' The students were also able to identify problems with bias in the system.

Haverkamp is also experimenting with AI support in exams. For example, he had an 8th grade class write an argument on the topic of 'Pros and cons of a public smoking ban'. Each student was allowed to use the three free queries on Smodin. The young people could get any support they wanted from the AI, but had to justify their approach. The exam thus consisted of three parts: the text written by the student, a justification for the writing process and the documentation of the three prompts used, including the system's answers.

One experience gives Haverkamp food for thought: he found that high-achieving pupils used

Among other things, the generative AI system smodin.io generates texts and can also back up statements in them with real sources.

The screenshot shows the Smodin website interface. At the top, there is a navigation bar with the Smodin logo and links for 'Umschreiber', 'Autor', 'CHATin', 'Mehr Werkzeuge', and 'Preisgestaltung'. A blue button labeled 'Autor' is in the top right corner. Below the navigation bar, the main heading is 'CHATin' with a 'Von Vorn Anfangen' button on the right. The chat area shows a user prompt: 'Ich soll einen Artikel über KI im Bildungswesen schreiben. Was sollte drin stehen?'. The AI response is a paragraph of text discussing the importance of AI in education, followed by a list of three bullet points: '- Die Personalisierung des Lernens durch KI', '- Die Unterstützung von Lehrkräften durch automatisierte Bewertungen', and '- Die Anwendung von KI-Technologien in der Schreibarbeit'. Below the text, there are three numbered references: '1. business24.ch', '2. unterrichten.digital', and '3. neuroflash.com'.

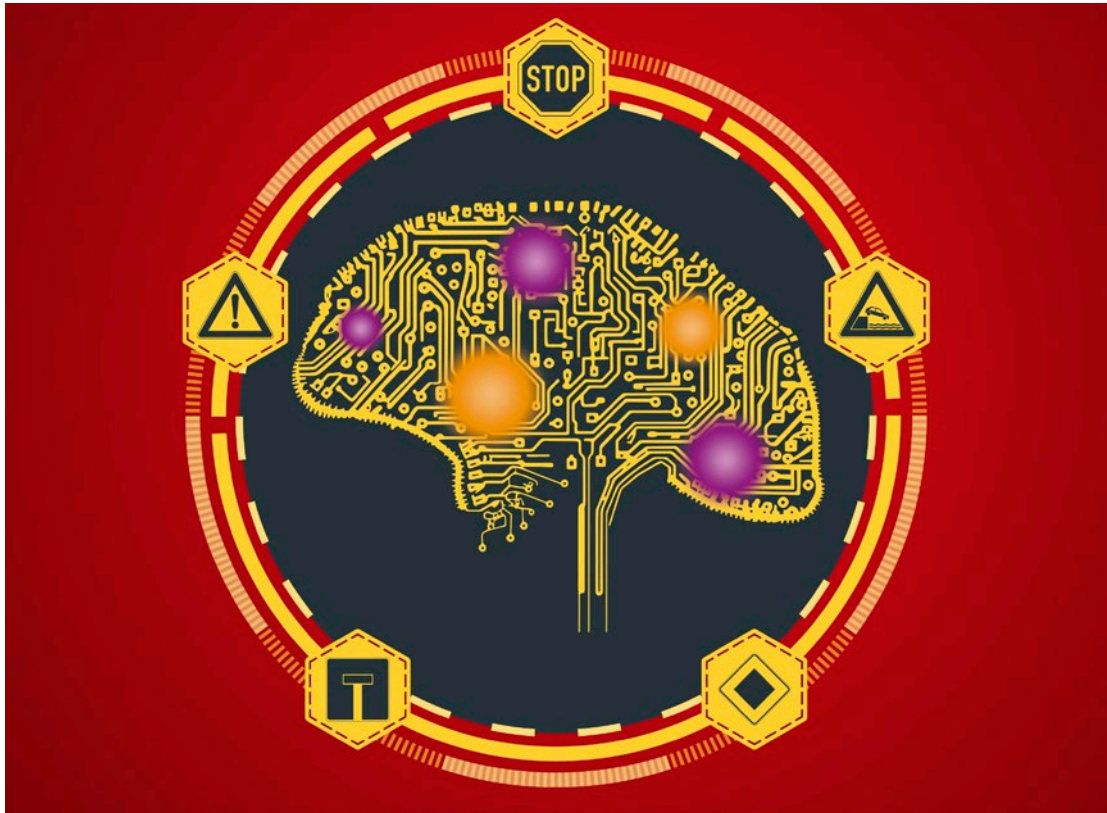


Image: KI Midjourney | stockphoto.grafstockadobe.com | editing.ct

Smart rules for AIs

Generative voice and image AIs are literally overrunning working life and society. US tech pioneers are practising unusual humility, German AI start-ups are sniffing the morning air, civil rights activists are warning. The EU wants to protect its citizens from the worst excesses of AI with a new regulation. However, such laws harbor the risk of over-regulation.

Andrea Trinkwalder

Lord, the need is great! I can't get rid of the spirits I called. In 1797, Goethe's sorcerer's apprentice called on his master for help in the face of an out-of-control magic broom. Today, more than 200 years later, the magic tool seems to be a reality: large universal language models

such as ChatGPT produce polished-sounding text in every conceivable form. Sometimes they slip into the role of a programmer, sometimes they write a lawyer's letter, sometimes they make medical diagnoses or explain the world. But because they sometimes behave just as erratically as the

high-spirited broom, the economy, society and politics are torn between a spirit of optimism and misgivings.

Suddenly, hard-boiled modern sorcerer's apprentices such as US tech visionary Elon Musk and OpenAI founder Sam Altman are so demonstratively overwhelmed and shocked by the power of their own creation that one must immediately become suspicious: Musk calls on developers to take a voluntary, six-month break in an open letter. He and over 1000 other signatories invoke nothing less than the singularity, i.e. the emergence of an uncontrollable artificial intelligence that is superior to humans. Altman, for his part, is using similar arguments to campaign for regulation before the US Congress and is behaving remarkably smoothly overall.

Such calls are just smokescreens that serve marketing purposes and vested interests, criticizes the civil rights organization AlgorithmWatch, among others. 'These horror scenarios are intended to distract from the fact that mechanisms must be found immediately to address and get to grips

with the specific and current dangerous effects of the systems,' explains AlgorithmWatch founder Matthias Spielkamp. 'Instead, they are intended to encourage us to focus valuable attention and energy on potential - and, from what we currently know, extremely unlikely - risks, such as the possibility of systems escaping human control.'

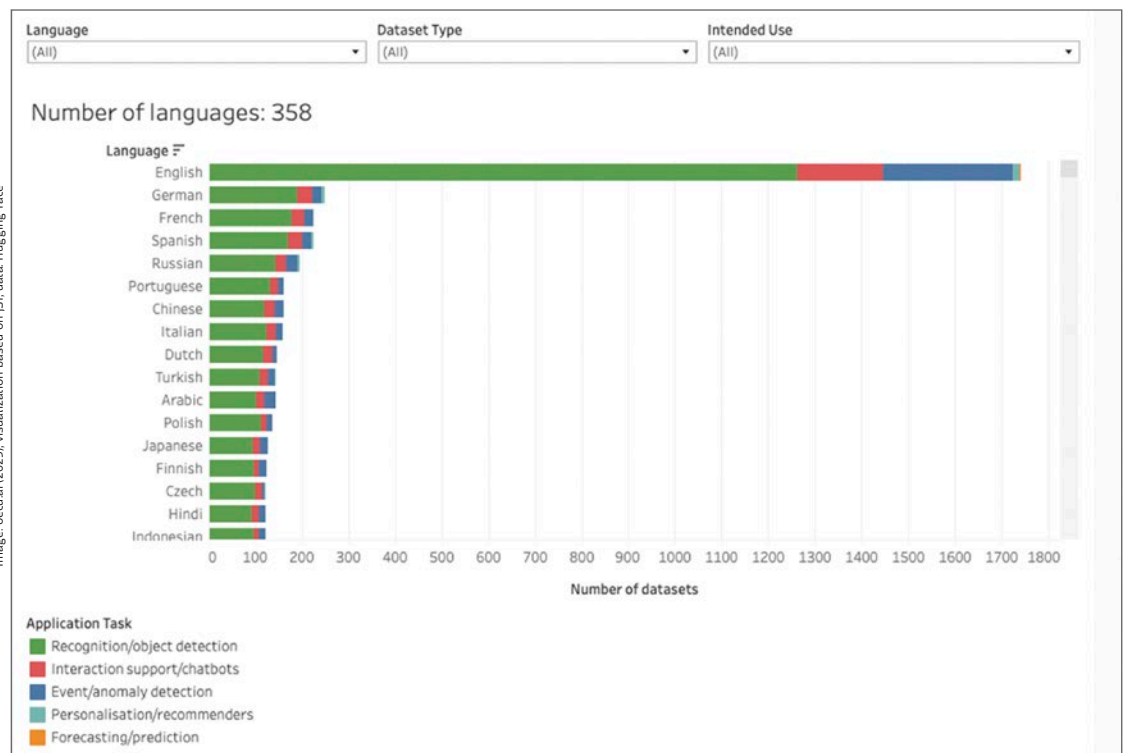
The German and European economy, on the other hand, does not want to leave the field to the American platforms without a fight this time. Although there are nowhere near as many or as large, Aleph Alpha and DeepL are certainly good and competitive companies. And because the cards are currently being completely reshuffled by universal technology, the German AI sector sees itself with an unexpectedly good hand this time, enabling it to stand up to the overpowering US companies.

Amazon moment

It's about a huge market: LLMs promise to be the technology that will sooner or later platformize

The statistics on the OECD's AI website reveal a more pressing problem than singularity: there are significantly fewer open source training datasets available for smaller languages than for English; many countries now want to develop their own independent sources.

Image: oecd.ai (2023), visualization based on ISI, data: Hugging Face



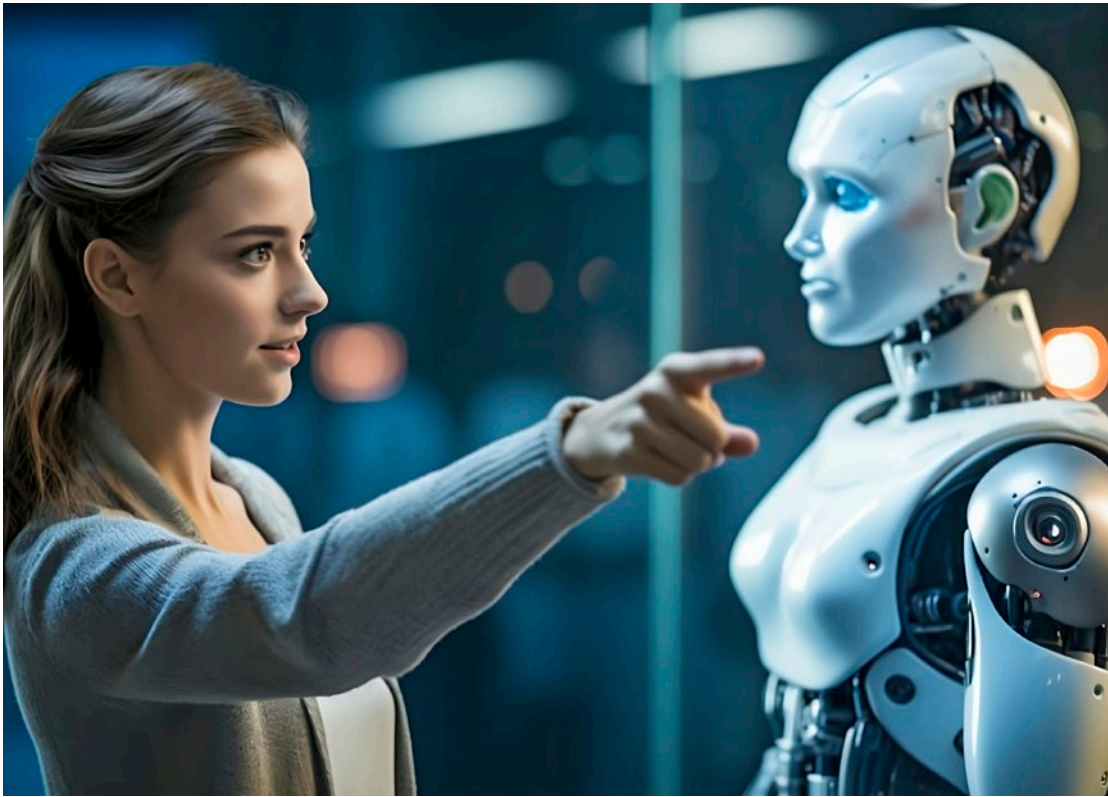


Image: AI Midjourney generated by Joerg Heidrich | Editing: ct

Guidelines for AI in the company

Many companies have long since moved from a phase of curious experimentation with AI to using it productively. This raises new questions about copyright and data protection law. In order to clarify these issues as best as possible, it is advisable to draw up a policy for dealing with AI within the company. This creates clear guidelines for the day-to-day use of ChatGPT & Co.

Joerg Heidrich

AI offerings have begun to irrevocably change the world of text, images and graphics. Many companies have long been

using such software for text generation, programming or illustration. Generative language models such as those behind ChatGPT train with billions

of pieces of content from the web that originate from other people or companies. The question is what legal implications this has for the content output by the AI and who in turn can assert rights to this content.

Other issues also need to be clarified, such as data protection. These include, for example, the important definition of what information may be entered into the AI as part of the prompts. This also affects compliance issues, such as the Protection of Secrets Act. Companies that use AI systems and results without thinking are entering a legal minefield.

Copyright and the machine

The time-honored Copyright Act (UrhG), whose foundations were laid at the beginning of the 20th century, protects graphics as well as paintings, films, code, texts and photographs. However, only 'personal intellectual creations' can be works within the meaning of the law (Section 2 (2) UrhG) and enjoy its protection. Accordingly, only the result of a human creation is protected, not the result of an algorithm executed by a machine.

If someone uses software to edit images, this is considered a human creation, as it is planned



The picture story 'Zarya of the Dawn' by AI artist Kris Kashtanova was the starting point for the first decision by the U.S. Copyright Office on copyrights to AI-generated images.

and executed by the user and the computer merely supports this. In this respect, it is a creative human activity that leads to a personal intellectual creation. The result is then protected by copyright for the user of the software.

However, an AI image is created differently: a user specifies a rough direction with their prompt, the end result remains random and cannot be reproduced identically. In this respect, a comparison with the commissioning of a painter by a patron is more apt. The patron does not become the author of the commissioned work, even if he gives precise instructions as to how his picture should look. The situation is similar with machine-generated data, which is also not protected by copyright, as it is not created by the human mind. Ultimately, most legal experts agree that AI-generated texts and graphics are not normally covered by copyright.

The first decision on the copyright classification of AI-generated images was delivered by the U.S. Copyright Office (USCO) in February 2023. The office had to decide on copyright protection for the picture story 'Zarya of the Dawn', in which the images had been created with the help of Midjourney. Protection was refused for the images created with AI as they were not based on human creativity. The decisive factor was that the results of Midjourney were unpredictable and a human user could not sufficiently control and direct the tool to create a specific image.

Some sectors find this somewhat easier in practice because they are helped by ancillary copyright. This protects AI graphics, for example, if they are used in computer games or films. These works are then automatically assigned to copyright due to the special protection in these areas. However, this is an exception.

No copyright: And now?

What does it mean in practical terms if there is no copyright on an image, graphic or text? The consequences are remarkable: the images are 'public domain'. They are not protected and can be used freely by anyone without a license. This makes such works the natural enemy of companies that earn their money by trading in images or texts. Providers of photo libraries such as Stockfoto are massively affected by this, especially in the case of more everyday symbolic images without artistic value. It is therefore hardly surprising that com-

IMPRINT

Editorial office

PO Box 61 04 07, 30604 Hanover
Karl-Wiechert-Allee 10, 30625 Hanover
Phone: 05 11/53 52-300
Fax: 05 11/53 52-417
Internet: www.heise.de

Letters to the editor and questions about the magazine:sonderhefte@ct.de

The e-mail addresses of the editors take the form xx@ct.de or xxx@ct.de. Please use the editor's abbreviation instead of "xx" or "xxx". You will find the abbreviations at the end of the articles and here in the imprint.

Editor-in-chief: Torsten Bееck (tbe) (responsible for the text section)

Concept: Jo Bager (jo), Hartmut Gieselmann (hag), Angela Meyer (anm)

Coordination: Jobst Kehrhahn (Head, keh), Pia Groß (piae), Angela Meyer (anm)

Editors: Jo Bager (jo), Holger Bleich (hob), Wilhelm Drehling (wid), Greta Friedrich (gref), Tim Gerber (tig), Hartmut Gieselmann (hag), Arne Grävemeyer (agr), Joerg Heidrich (joh), André Kramer (akr), Pina Merkert (pmk), Andrea Trinkwalder (atr), Dorothee Wiegand (dwi), Stefan Wischner (swi)

Contributors to this issue: Immo Junghärtchen, Kai Schwirzke

Assistance: Susanne Cölle (suc), Tim Rittmeier (tir), Christopher Tränkmann (cht), Martin Triadan (mat)

DTP production: Dörte Bluhm, Lara Bögner, Beatrix Dedek, Madlen Grunert, Lisa Hemmerling, Cathrin Kapell, Steffi Martens, Marei Stade, Matthias Timm, Christiane Tümmeler, Ninett Wagner

Digital production: Christine Kreye (director), Kevin Harte, Thomas Kaltschmidt, Martin Kreft, Pascal Wissner

Photography: Andreas Wodrich, Melissa Ramson

Illustration: Rudolf A. Blaha, Frankfurt am Main; Albert Hulm, Berlin; Moritz Reichartz, Viersen

Title: Steffi Martens, www.freepik.com

Publisher

Heise Medien GmbH & Co KG
PO Box 61 04 07, 30604 Hanover
Karl-Wiechert-Allee 10, 30625 Hanover
Phone: 05 11/53 52-0
Fax: 05 11/53 52-129

Editors: Christian Heise, Ansgar Heise, Christian Persson

Managing directors: Ansgar Heise, Beate Gerold

Members of the Executive Board: Jörg Mühle, Falko Ossmann

Advertising manager: Michael Hanke (-167) (responsible for the advertising section), www.heise.de/mediadaten/ct

Advertising sales: Verlagsbüro ID GmbH & Co. KG, Tel.: 05 11/61 65 95-0, www.verlagsbuero-id.de

Head of Sales and Marketing: André Lux (-299)

Service Offprints: Julia Conrades (-156)

Printing: APPL Druck GmbH & Co KG, Senefelder Str. 3-11, 86650 Wemding, Germany

Retail sales:
DMV DER MEDIENVERTRIEB GmbH & Co KG
Meßberg 1
20086 Hamburg
Phone: 040/3019 1800, Fax: 040/3019 145 1815
E-mail: info@dermedienvertrieb.de
Internet: dermedienvertrieb.de

Unit price: €14.90; Switzerland CHF 27.90; Austria € 16.40; Luxembourg € 17.10

First day of sale: 24/11/2023

Despite careful checking by the editors, the publisher cannot accept any liability for the accuracy of the publications. No part of this publication may be reproduced in any form or processed, copied or distributed using electronic systems without the express written permission of the publisher. The use of programs, circuit diagrams and printed circuits is only permitted for the purpose of further education and for the personal use of the reader.

No liability can be accepted for unsolicited manuscripts. By submitting manuscripts and images to the editors, the author grants the publisher the exclusive right of publication. Paid work becomes the property of the publisher. All publications are made without consideration of any patent protection.

Trade names are used without guarantee of free use.

Created and produced with Xpublisher: www.xpublisher.com

Printed in Germany.

All rights reserved.

© Copyright 2023 by Heise Medien GmbH & Co. KG

Preview: ct PC building & upgrading

Assembly instructions - Tests - Purchasing advice

The ideal PC is usually not available off the shelf. In the special issue ct PC bauen & aufrüsten, the ct editorial team presents building suggestions for desktop computers that optimally meet your own requirements. The building instructions cover a broad spectrum from inexpensive mini PCs for the home office to powerful gaming machines and can also be customized.

Comprehensive buying guide articles explain what is important when selecting hardware components. Detailed tests provide an overview of the latest processors, fast SSDs, powerful graphics cards and extensively equipped mainboards. The special edition also contains practical tips on upgrading CPUs, RAM, graphics cards and SSDs as well as on saving power with Ryzen 7000 processors.

[Weitere Infos: ct.de/wr43](https://www.ct.de/wr43)

Image: Artem Adobe Stock

Main topics

Building the optimal PC

- Building suggestions for gaming PC/All-rounder, office PC, budget gamer, mini PC
- Hardware diagnostics with HWiNFO

Upgrade your computer optimally

- Upgrade SSD, graphics card, RAM, processor
- Move Windows to a new SSD
- Second PC life with Linux

Select processors and mainboards

- CPU guide
- Processor and motherboard tests for AMD Ryzen 7000 and Core i-13000/14000
- Energy-saving tips

Tuning with high-end components

- In the test: fast SSD, large RAM, powerful graphics-cards, large PC cases



Die Konferenz zu IoT und IIoT

21. UND 22. FEBRUAR 2024 • MÜNCHEN

Software entwickeln für das (I)IoT

Aus dem Programm:

- Unified Namespace in der Praxis
- Wie vertraut man IIoT?
- Ultraniedriger Stromverbrauch von Embedded-Geräten
- KI auf MCUs - Wie geht das richtig?
- Cyber Resilience in Aktion

Jetzt
Frühbuche-
ticket
sichern!

www.buildingiot.de

Veranstalter



 dpunkt.verlag