

Anwendungen

Wozu wird KI eingesetzt?

Stephan Verbücheln

cnlab Herbsttagung 2024: KI und Sicherheit
Gleisarena, Zürich, 4. September 2024

SEGA AI Computer Japan 1986

人工知能が家庭に入る日。

最先端のコンピュータ・テクノロジー、人工知能(AI)を搭載した初めてのCAI。
SEGA AI COMPUTER。

遠い未来のことだと思っていたことが、突然目の前に現われることがあります。あなたの個性や能力を知り尽くし、納得のいくまであなたの

学習者の心理
わかる
おもしろくなる
やる気が出る

相手になってくれるベテランの先生が、いつもそばにいてくれたら…。そんな願いを実現したのが、人工知能を搭載したセガAIコンピュータです。人が教える教育環境^{※1}これを行い、コンピュータによって効率よくシ

ステム化する。私たちは、セガAIコンピュータによって、この課題に応えることができたと考えます。ひとりひとりに合わせたマン・ツー・マン学習。コンピュータとの自由な対話のやりとりで納得のいくまで学習できる双方向型の知的CAI。わかるからおもしろい。おもしろいからやる気が出る。だれもが待ち望んでいたこのAIコンピュータによって、私たちセガは、いま、知的情報産業へ、大きな一歩を踏み出しました。

※1 人工知能(AI=Artificial Intelligence)を組み込んだコンピュータ、論理演算プラス知識の適用による推論によって処理対象を理解します。
※2 CAI(Computer Assisted Instruction)は、インストラクターによる教育をコンピュータによって置き換えたといわれる個人の能力の向上に重点をおいたコンピュータ学習システム。

家庭学習

SEGA
AI 学習トライアングル
(セガ AI コンピュータの教育構造)

知的CAIとは、個人の特性に合わせて指導することによってひとりひとりの能力を向上させることを目的とするコンピュータ学習システムです。CAIには、大きく分けて3つのメリットがある、とセガは考えます。家庭学習、能力別学習、対話型学習。これらの要素がそれぞれによって効果を発揮したとき、確実に学習能力のアップが望めるのです。人工知能(AI)によるCAIの開発思想の原点には、このAI学習トライアングルがありました。

能力別学習



- 16ビットCPU搭載
AI機能を実用レベルで稼働させるために大型パソコン用16ビットCPUを搭載。
- 完全なコンピュータ・コントロールによる高速カセット(9600BPS)搭載
対話に必要な音声と、あらかじめプログラムされている豊富な情報の中から、必要なものを同時に呼び出す高速カセットを搭載。テープソフトでもほとんど待ち時間がありません。
- タッチタブレット方式採用
入出力にタッチタブレット方式を採用。しかも、各コースウェアにはそのカリキュラムに合った「ターゲッ」印刷されたシート(オーバーレイシート)を用意。指先ひとつで入力できます。
- 本格的キーボード(新JIS配列)
操作性に優れた使いやすいキーボード(新JIS配列)。
- 画面入力
マウス方式に近い方法での画面入力も可能。
- 高解像度LSI採用
鮮明な画像を実現する高解像度LSIを採用。美しいグラフィック画面を実現しています。
- 多彩なメディア
コースウェアは、そのソフトの性能によりROMカード、フロッピーディスク(オプション)も使えます。



インストラクター(先生)によるグループ指導と違い、CAIの目的は、言うまでもなく「個人対応型の指導」にあります。学校の「斉授業」とは異なる、生徒ひとりひとりの理解力や進度に合わせたマン・ツー・マンの家庭内学習。これをよりキメ細かなシステムで実現したのが、セガAIコンピュータです。時間に制約されず、自分が納得のいくまで、自分のペースをコンピュータが考えてくれる学習。知る▶覚える▶慣れる▶基本が身につく▶応用がきく▶みずから乗り込む(工夫する)。このプロセスを繰り返すことによって、学習者の問題解決能力を自分に合った進度で一步一步着実に高めていく。この教育本来の目的を人工知能を使った知的CAIというカタチで実現したのが、セガAIコンピュータです。これによって、コンピュータ(あるいは視覚覚機器)を使っていたまでの個別学習には見られない、より人間にふさわしいスタイルの学習を可能にしました。

これまでのCAIシステムは、あらかじめ能力別に用意されたプログラムの中から、学習する側が選択するというように、学習者の方がコンピュータに合わせなければならなかった。ですから、出題された問題を解こうとして不正解になり、ヒントを見てもわからないというようなことが往々にありました。コンピュータに柔軟性がなく、画一的なヒントしか出ないからそうなるので。セガAIコンピュータは、人工知能言語(SEGA PROLOG)を使うことによって、学習者個人の能力をベテランの先生のようにつかみ取り、それに応じたカタチで教える内容や出題を、随時、かつ自在に変化させていきます。できる問題はどんどん高度になってゆき、できない問題は教え方を変えながらどこがわからないかを徹底的に解明し、確かな理解を実現している。それが、人工知能(AI)を使った「能力別学習」です。

いわゆるページめくり型といわれる従来のアドホック・フレーム型CAIでは、学習者はコンピュータの問いかけに対し、あらかじめ決められたフォーマット(形式)で、キーを押して解答するだけしかありませんでした。ところが、セガAIコンピュータでは、学習すべきテーマを学習者が自分で選ぶことができ、しかも普段しゃべっている日常語(自然言語)でシステムに入力することができます。さらに、これまでのように「コンピュータが問題を出す」▶「学習者が答える」▶「コンピュータが評価することに加え、「学習者が質問する」▶「コンピュータが答える」▶「それに対してまた学習者が質問する」という双方向の本当の意味での対話学習が可能になりました。これによって、たとえば従来のCAIでは不可能といわれていた国語の長文読解訓練もできるようになりました。

対話型学習

Bild: SEGA



Agenda_

1. Klassifizierung und Erkennung
2. Generierung
3. Rechtliches

Klassifizierung

Optical Character Recognition (OCR)

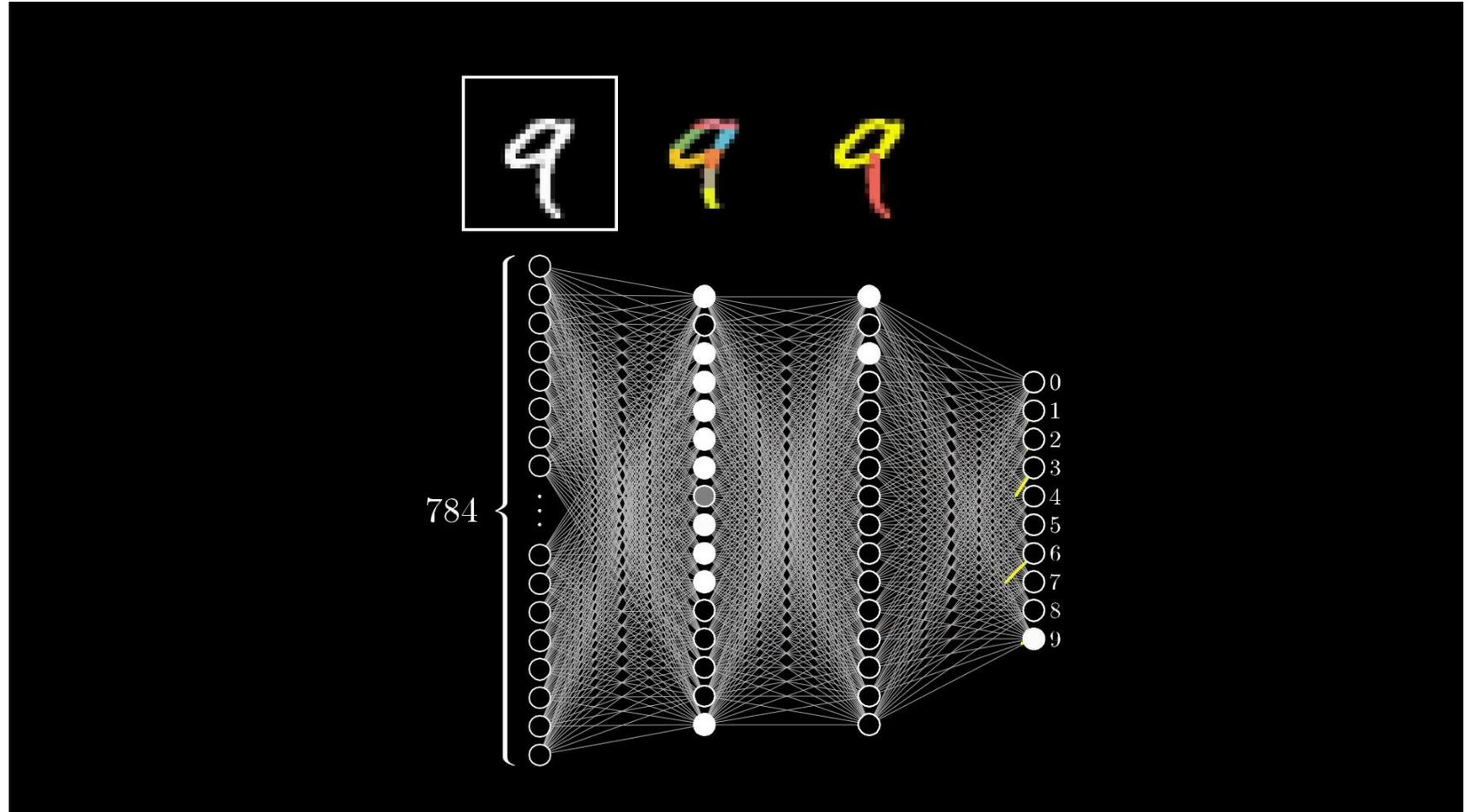
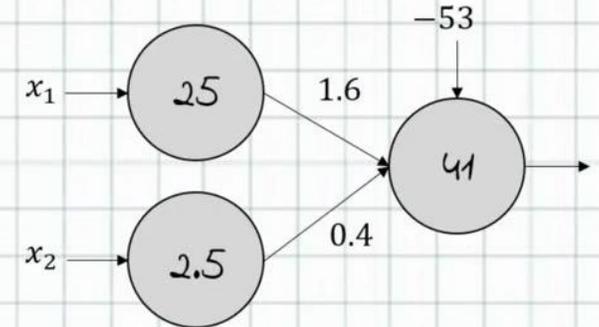
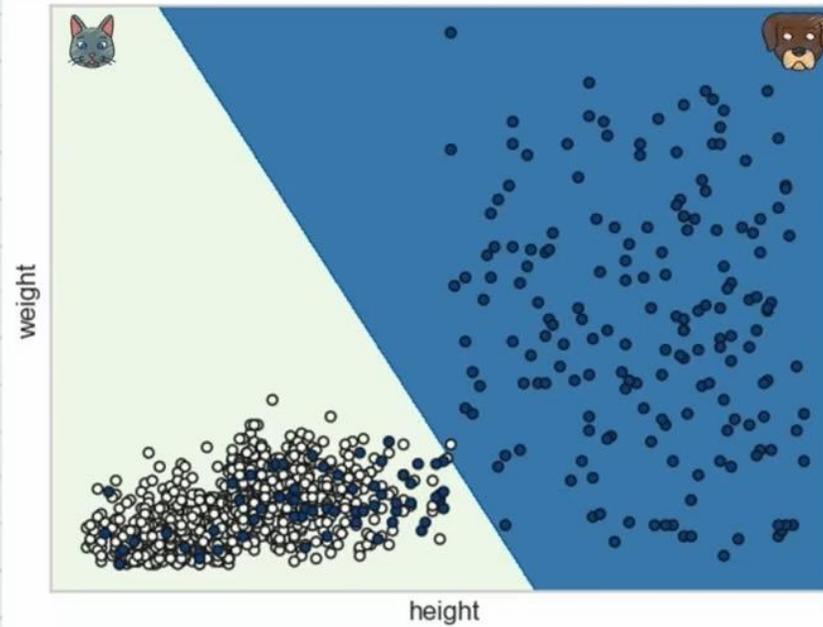


Bild: 3Blue1Brown (Youtube)

Klassifizierung

Hund vs. Katze

Das fertige Modell



6/25

Klassifizierung

Hund vs. Katze

Eine kleine Katze...



LIME



1. vereinfachen

Integrated Gradients



2. Gewichtung der
Modellparameter

GradCAM



3. Bestimmtes
Layer

17/25

Bild: Annika Rüll (37C3)

Klassifizierung

Threat Detection

- Logs
- Netzwerkzugriffe

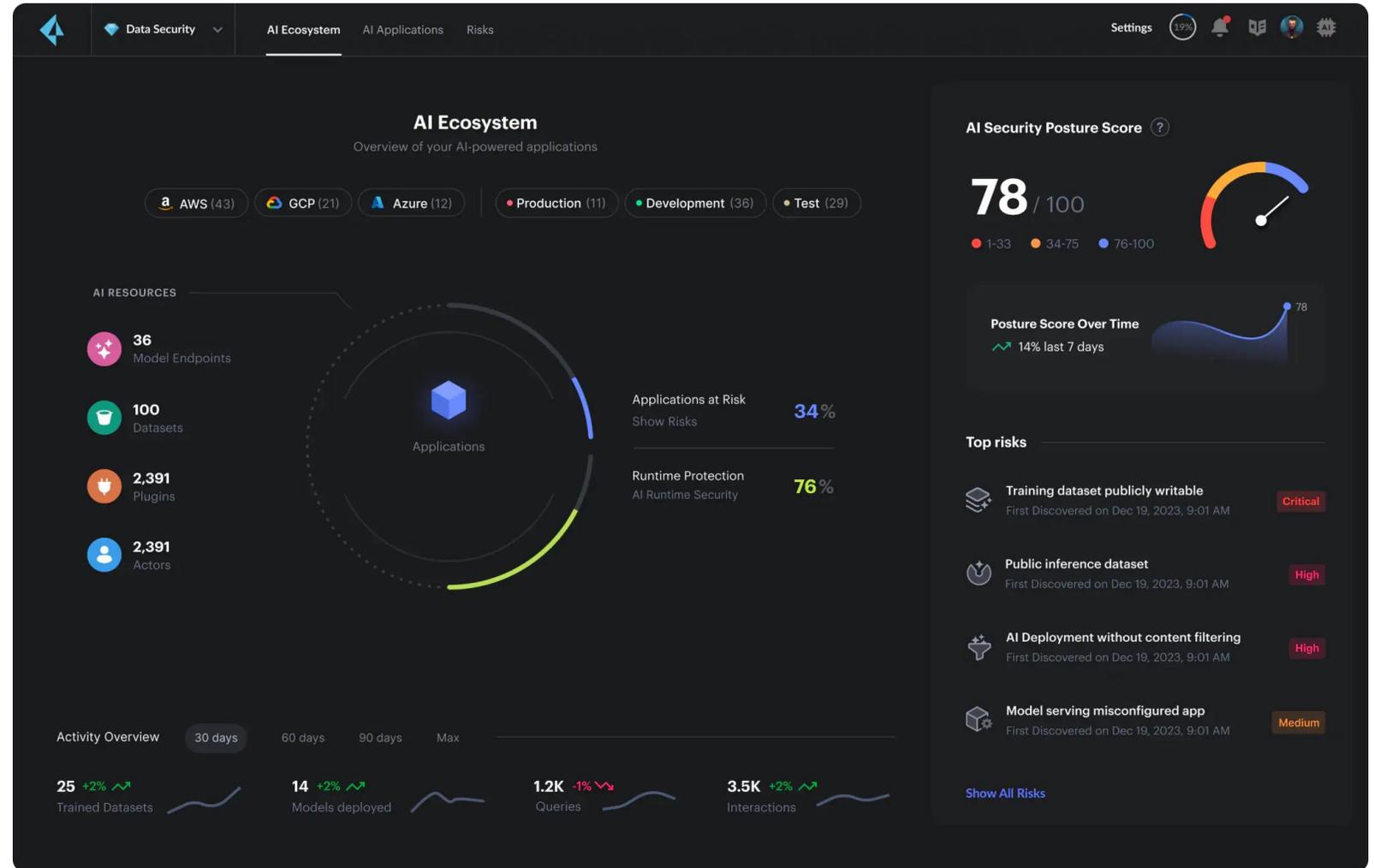


Bild: Palo Alto (AI Security Posture Management)





Klassifizierung

Threat Detection

- **Möglichkeiten und Grenzen**
 - Halteproblem
 - Satz von Rice
 - Fred Cohen

Several undecidable problems have been identified with respect to viruses and countermeasures. They are summarized here:

Undecidable Detection Problems

- Detection of a virus by its appearance
- Detection of a virus by its behavior
- Detection of an evolution of a known virus
- Detection of a triggering mechanism by its appearance
- Detection of a triggering mechanism by its behavior
- Detection of an evolution of a known triggering mechanism
- Detection of a virus detector by its appearance
- Detection of a viral detector by its behavior
- Detection of an evolution of a known viral detector

<https://web.eecs.umich.edu/~aprakash/eecs588/handouts/cohen-viruses.html>

Erkennung

Objekte in Bildern

2016

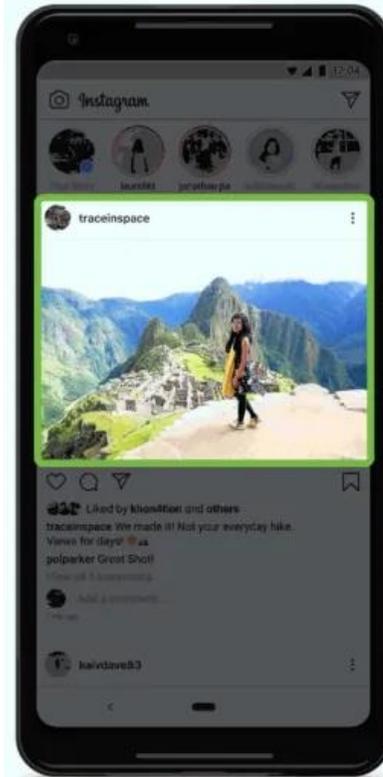


Bild: Techcrunch



May be an image of
**1 person, standing,
and Machu Picchu**

2021



May be an image of 5 people,
including Jay Youmens, people
playing musical instruments,
people standing, 2 hats, and
5 drums

Erkennung

Selbstfahrende Autos



Bild: viso.ai

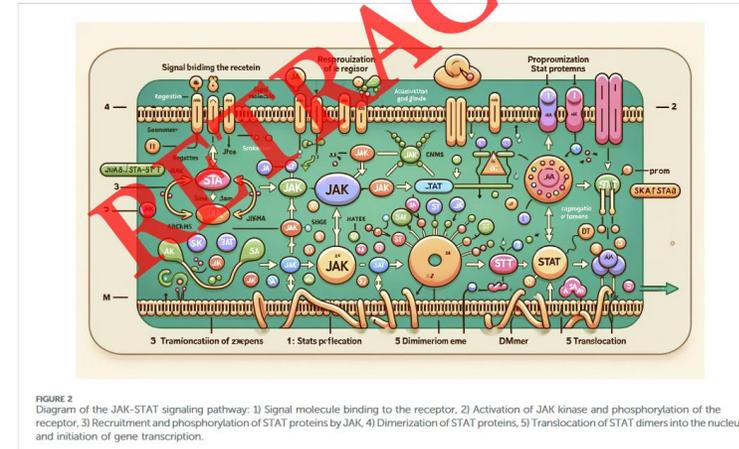
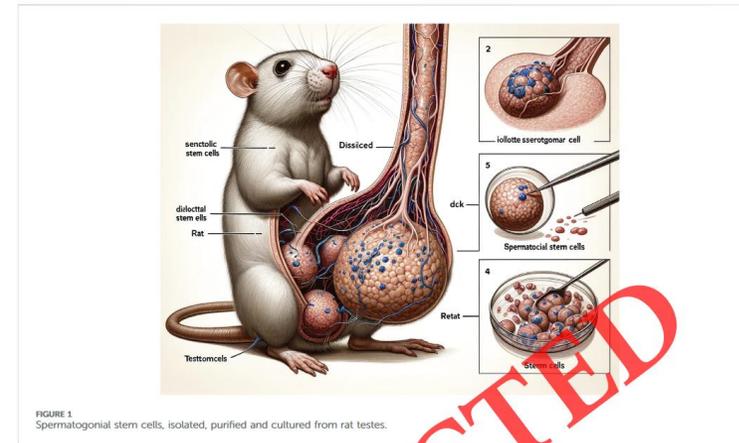
Erkennung

Automatische Untertitel,
Protokolle



Generierung

Generierte Dokumente, Zusammenfassungen



characteristics of SSCs and the JAK/STAT pathway can be approached from two perspectives: firstly, examining the impact of the target cell's JAK/STAT pathway on SSCs (Tulina and Matunis, 2001); and secondly, investigating the influence of the SSC's own JAK/STAT pathway on its biological properties (Brawley and Matunis, 2004).

Bild: Frontiers in Cell Development



Generierung

Interaktive Chatbots

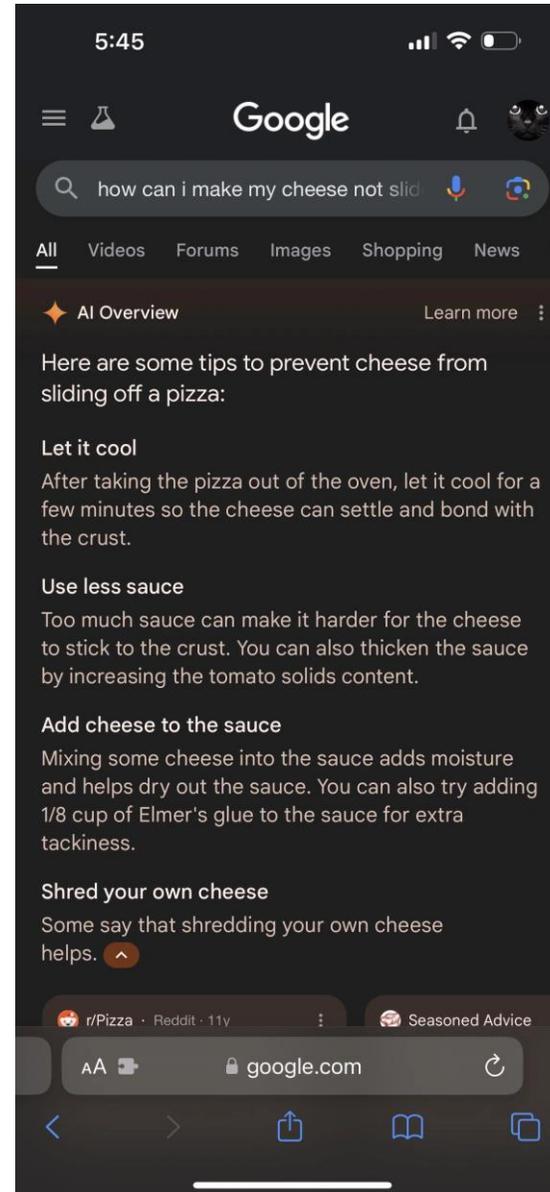
The screenshot shows a chat window with the following elements:

- Header:** COPILOT logo, NOTIZBUCH icon, language options (Français | Italiano), and an Anmelden button.
- User Message:** "Sie" (represented by a person icon) asks: "Was ist die Herbsttagung von cnlab?"
- Copilot Response:** "Copilot" (with logo) starts with a green checkmark and "Antworten für Sie werden generiert...". The response text is: "Die Herbsttagung von cnlab ist eine jährliche Veranstaltung, die sich mit aktuellen Themen im Bereich der IT-Sicherheit und Technologie befasst. Dieses Jahr steht die Tagung unter dem Motto 'KI und Sicherheit' und findet am **Mittwoch, den 4. September 2024**, von 13:00 bis 17:00 Uhr in der Gleisarena FFHS in Zürich statt ¹." Below this, it says "Das Programm umfasst verschiedene Vorträge und Präsentationen zu Themen wie:" followed by a bulleted list:
 - Grundlagen und Anwendungen von Künstlicher Intelligenz (KI)
 - Software-Entwicklung mit KI
 - Angriffe gegen und mit KI
 - Sicherheitsaspekte und gesetzliche SchrankenIt then continues: "Es gibt auch eine Kaffee-Pause und eine abschließende Demo-Session mit Apéro ¹." and "Interessiert dich ein bestimmtes Thema aus dem Programm?"
- References:** "Weitere Informationen" followed by five numbered links: 1. cnlab.ch, 2. cnlab.ch, 3. bing.com, 4. cnlab.ch, 5. cnlab.ch.
- Interaction Bar:** Includes icons for thumbs up/down, copy, download, and voice, along with a "1 von 4" indicator.
- Quick Actions:** Three buttons: "Kann ich mich noch anmelden?", "Gibt es eine Live-Übertragung der Veranstaltung?", and "Welche Referenten sind dabei?"
- Input Area:** A blue "Neues Thema" button, a text input field with the placeholder "Fragen Sie mich etwas...", and icons for image, voice, and send.



Generierung

Suchmaschinen



Generierung

Programmcode

- Github Copilot
- Amazon Code Whisperer

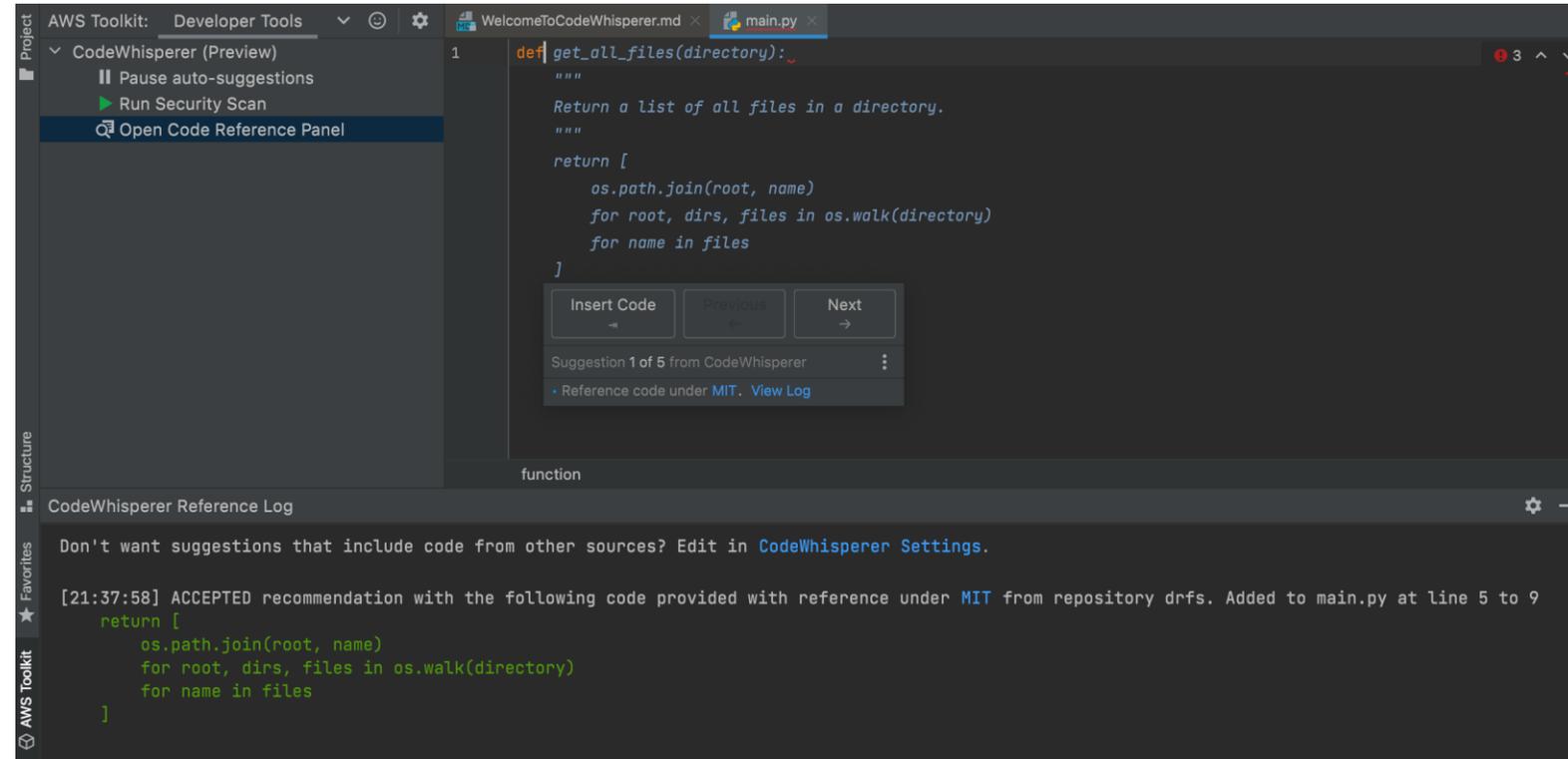


Bild: Amazon

Generierung

Bilder

- OpenAI Dall-E (Craiyon)
- StabilityAI Stable Diffusion
- Midjourney
- Adobe Firefly



Generierung

Deep Fakes

- Falscher Klitschko in Berlin
- Video-Identifizierung

Bild: Senatskanzlei Berlin

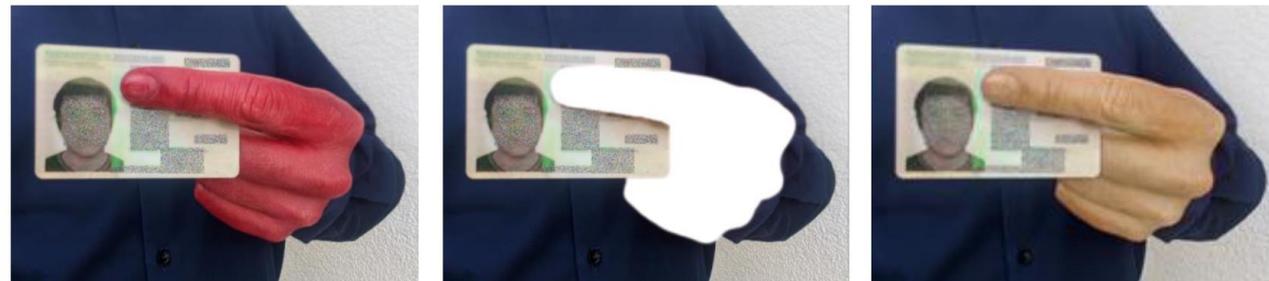
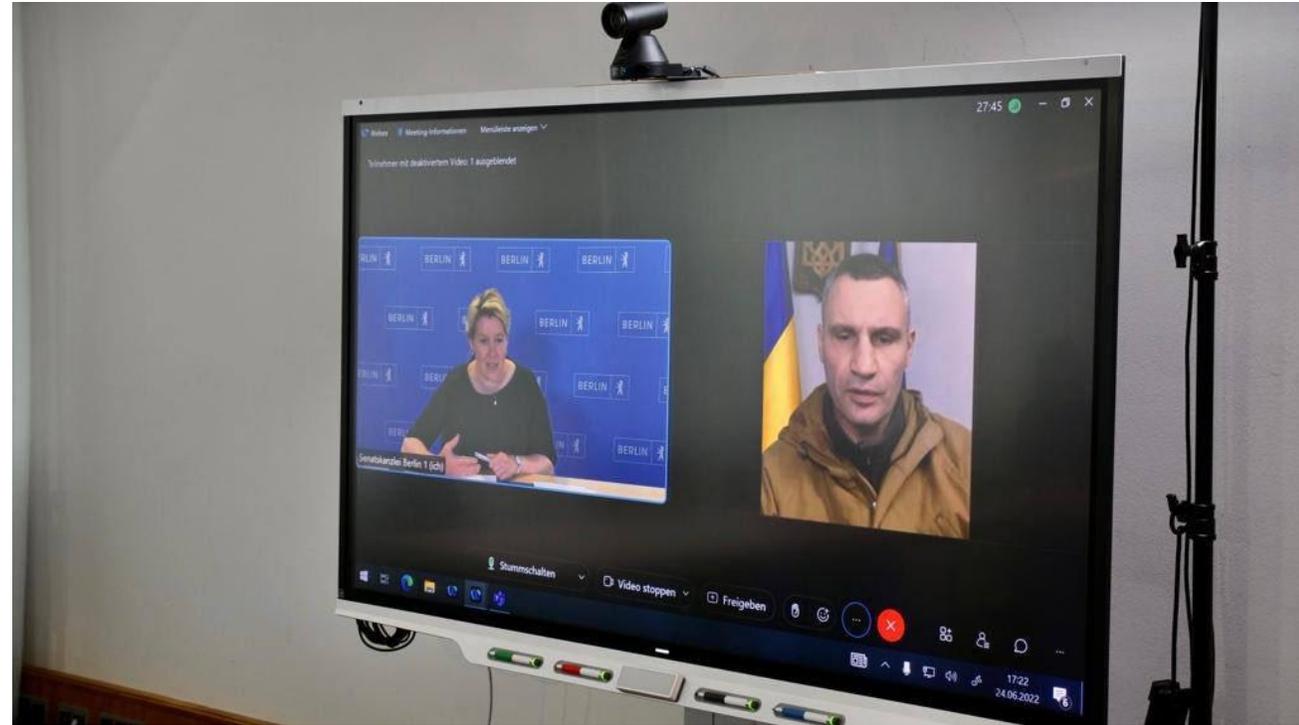


Bild: Chaos Computer Club

Abbildung 12: Segmentierung anhand rot eingefärbter Bereiche. Die Original-Hautfarbe wird durch Verschiebung des Farbtons innerhalb des Segmentes wiederhergestellt.

Generierung

Musik

- suno.com

SUNO

Make a song about anything

(You'll need to sign up for a free account)

Home

Create

Library

Explore

Search

A futuristic synthpop song about writing a face-melting guitar solo

Create a Song

Global Trending

Global

Now



Carambola-スターフ...

[Classical-Crossover], Ar...



Brutus



Dear, Oh Dear

{ industrial grindcore x h...



frowns



Get It Done, Son

Emo Cloud Psychedelic ...



Teemuth



Chasing Sunsets |...

Pop Rock Psychedelic F...



Teemuth

Subscribe

What's New? 20

Help

About

Careers

Sign In

The Summer of Suno, Vol. 2

Show more





Nutzen

I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes.

Joanna Maciejewska, Twitter, 29. März 2024



Rechtliches: Trainingsdaten

Trainingsdaten aus dem Internet

- ➔ Datenschutz wegen personenbezogener Daten
 - Mehrere Klagen eingereicht, noch keine Urteile
- ➔ Urheberrechtsprobleme (Nutzungsrechte und Attributierung)
 - Nutzer (Entwickler) haftet für Urheberrechtsverletzungen
 - Die grossen Anbieter versprechen Haftungsübernahme



Rechtliches: Qualität

Qualität unzuverlässig

→ Fehlerhafte Ergebnisse

- Haftung im Fall von fehlerhaften Entscheidungen beim Nutzer

→ Autonome Autos

- Haftung im Fall von Unfällen ist beim Betreiber des Dienstes

→ Fahrassistenten

- Haftung im Fall von Unfällen zurzeit beim Fahrer

Europäische Union: AI Act



Vielen Dank für Ihre
Aufmerksamkeit_

Stephan Verbücheln

security@cnlab.ch

+41 55 214 33 33

cnlab security AG

Obere Bahnhofstrasse 32b

CH-8640 Rapperswil-Jona

Switzerland