



Jan Fischer, Axel Wagenitz

## Generative KI und Cybersecurity: Wissenschaft trifft Praxis

**Generative KI und Deepfake-Technologien bringen neue Herausforderungen für die Cybersicherheit mit sich. Durch täuschend echte Video- und Audiofälschungen können Angreifer in die Identität von Führungskräften oder Mitarbeitern schlüpfen und unerlaubte Transaktionen auslösen oder sensible Daten abgreifen. Die Weiterentwicklung dieser Technologien macht es notwendig, durch gezielte Mitarbeiterschulungen und Sicherheitsprozesse ein Bewusstsein für diese Risiken zu schaffen.**

### Einleitung

Das Telefon klingelt, und der Buchhaltungsmitarbeiter seufzt – ein Anruf kurz vor Feierabend. Zögernd nimmt er ab; auf dem Bildschirm erscheint ein Video: Die Geschäftsführerin, ernst und gestresst. Sie sei im Ausland, es gebe ein Problem. Ein wichtiger Deal drohe zu platzen. „Wir müssen heute noch eine Überweisung machen.“

Die Verbindung knistert, das Bild flackert, aber die Stimme ist unverkennbar. Kein Zweifel, es ist die Chefin. Der Mitarbeiter nickt, versucht zu begreifen, was auf dem Spiel steht. „Wir reden hier von Millionen“, fügt die Geschäftsführerin hinzu, ihr Blick durchdringend. Alles hänge davon ab.

Das Gespräch endet abrupt, und sofort erscheint eine E-Mail. Der Mitarbeiter öffnet sie, seine Hände zittern leicht. Kontodaten, Betrag – alles ist klar und präzise aufgelistet. Es bleibt keine Zeit für Fragen, keine Zeit, das Gesagte zu hinterfragen. Die Stimme war authentisch, die Dringlichkeit echt – die Geschäftsführerin hat es angeordnet, und es gibt keine Zeit zu verlieren.

Was der Mitarbeiter nicht weiß: Am anderen Ende der Leitung saß nicht die Geschäftsführerin. Es war eine Künstliche Intelligenz (KI), ein Deepfake, der mit erschreckender Präzision Stimme und Gesicht imitiert hat. Ein perfekt orchestrierter Betrug, der in wenigen Minuten Millionen aus dem Unternehmen schleust.

Diese Geschichte ist fiktiv, doch genau so oder ähnlich kann es in der Realität passieren. Anfang 2024 wurde ein internationaler Konzern Opfer eines solchen Deepfake-Betrugs. Kriminelle inszenierten eine täuschend echte Videokonferenz, in der sie Stimmen und Gesichter von Führungskräften nachahmten. Der Druck war groß, die Situation schien dringend, und so gelang es ihnen, 24 Millionen Euro zu erbeuten. Der Mitarbeiter, der die Überweisung vornahm, hatte keine Chance, den Betrug zu erkennen.<sup>1</sup>

Doch nicht immer sind solche Angriffe erfolgreich. Im Frühjahr 2024 geriet Ferrari ins Visier von Betrügern. Ein Topmanager erhielt WhatsApp-Nachrichten und sogar einen Anruf von einer KI, die Stimme und Akzent des Ferrari-CEOs Benedetto Vigna perfekt imitierte. Doch der Manager war misstrauisch. Kleine Unstimmigkeiten in der Sprechweise ließen ihn aufhorchen, und er stellte eine Fangfrage. Der Betrug flog auf, bevor Schaden angerichtet werden konnte. Dieser Fall zeigt, wie täuschend echt solche Deepfake-Angriffe inzwischen sein können und wie knapp Unternehmen oft an einem Desaster vorbeischnappen.<sup>2</sup> Für kleine und mittlere Unternehmen (KMU) ist es daher entscheidend, sich der Risiken bewusst zu sein und geeignete Sicherheitsmaßnahmen zu ergreifen.<sup>3</sup>

Das „Business Innovation Lab“ (BIL) der HAW Hamburg, als Teil des Mittelstand-Digital Zentrums Hamburg, untersucht und testet die Potenziale und Gefahren von Technologien wie Deepfakes und anderen generativen KI-Anwendungen. Ziel ist es, Unternehmen fundierte Einschätzungen zu möglichen Risiken und Nutzen dieser Technologien zu geben. Neben den Gefahren durch Missbrauch gibt es aber auch positive Einsatzszenarien. So können KMU generative KI nutzen, um beispielsweise automatisierte Produktvideos oder multilinguale Onboarding-Videos zu entwickeln. Durch praxisnahe Forschung und realitätsnahe Szenarien bietet das BIL KMU wertvolle Einblicke und unterstützt sie dabei, sich besser gegen Bedrohungen zu schützen und gleichzeitig innovative Anwendungen gewinnbringend zu nutzen.

## Deepfakes: Die Kunst der digitalen Täuschung

Ursprünglich beschrieb der Begriff Deepfakes lediglich der Erstellung von KI-generierten Bildern. Mittlerweile sind damit jedoch auch täuschend echte Audio- und Videoaufnahmen gemeint. Deepfakes basieren auf Deep-Learning-Algorithmen, die Mimik, Gestik oder die Stimme einer Person analysieren und verblüffend realistisch reproduzieren können.<sup>4</sup> Kriminelle nutzen diese

Fortschritte, um in die Rolle von Geschäftsleitung, Mitarbeitern oder anderen Vertrauenspersonen zu schlüpfen. So können sie unbemerkt unbefugte Transaktionen durchführen oder auf sensible Daten zugreifen. Die Auswirkungen sind gravierend: Identitätsbetrug, Industriespionage und finanzieller Schaden sind nur einige der möglichen Konsequenzen.<sup>5</sup>

Früher waren minutenlanges Videomaterial und stundenweise Audioaufnahmen erforderlich, um realistische Fälschungen zu erstellen. Heute sind moderne Deep-Learning-Modelle in der Lage, mit nur einem oder wenigen Bildern einen einigermaßen realistischen Deepfake zu erzeugen. Ebenso können sie innerhalb weniger Sekunden die Stimme einer Person klonen. Die Erstellung hochwertiger Deepfakes war einst mit großem technischem Aufwand und hohen Kosten verbunden, die erhebliche Rechenleistung erforderten. Mittlerweile sind benutzerfreundliche Tools frei verfügbar, und kommerzielle Dienste bieten Deepfake-Software an, die auch auf normaler Verbraucher-Hardware läuft. Dadurch ist es nun auch technisch weniger versierten Personen möglich, täuschend echte Fälschungen zu erstellen.<sup>6</sup>

## Automatisierte Phishing-Angriffe: Massenhafte Täuschung

Generative KI eröffnet Cyberkriminellen neue Möglichkeiten bei klassischen Phishing- und Spear-Phishing-Angriffen. Sie können personalisierte Phishing-Angriffe in großem Maßstab automatisieren und so massenhaft täuschen. Die durch Large Language Models (LLMs) generierten E-Mails wirken täuschend echt, weil sie individuell auf die jeweiligen Empfänger zugeschnitten werden können.<sup>7</sup> Ein alarmierendes Beispiel ist der 135%ige Anstieg von Spam-Mails mit verbesserter Grammatik und Syntax, der auf den Einsatz generativer KI zurückzuführen ist.<sup>8</sup>

Neben E-Mails setzen Kriminelle auch automatisierte Telefonanrufe ein. Dabei verwenden sie synthetische Stimmen, um sensible Informationen zu erlangen - ein Vorgehen, das als Voice Phishing (Vishing) bezeichnet wird.<sup>9</sup>

## WormGPT: Large Language Models für schnelle Malware-Entwicklung

Die Produktivitätsgewinne durch Large Language Models wie ChatGPT kommen nicht nur Unternehmen zugute, sondern auch Cyberkriminellen. Bereits 2021

1 Hurtz (2024).

2 Brien (2024).

3 Bundesamt für Sicherheit in der Informationstechnik (2022).

4 Perov et al. (2022); Mirsky et al. (2021); Bovenschulte (2019).

5 Bundesamt für Sicherheit in der Informationstechnik (2022); Mirsky et al. (2021); Dash und Sharma (2023).

6 Bovenschulte (2019).

7 Falade (2023); Mirsky et al. (2021); Neupane et al. (2023).

8 Neupane et al. (2023).

9 Falade (2023); Neupane et al. (2023).

tauchte WormGPT in Darknet-Foren auf; es soll speziell für die Entwicklung von Schadsoftware entwickelt worden sein. Im Gegensatz zu ethisch ausgerichteten Modellen bietet es uneingeschränkte Möglichkeiten zur Erstellung schädlicher Inhalte.<sup>10</sup> Mit der Weiterentwicklung immer besserer freiverfügbarer Modelle ist davon auszugehen, dass auch die Qualität und Vielfalt der Malware steigen wird.

## Visuelle Deepfakes: Funktionsweise und Technologien

Die Erstellung visueller Deepfakes basiert auf verschiedenen KI-Technologien, darunter Autoencoder, Generative Adversarial Networks (GANs) und Diffusionsmodelle.

### Autoencoder

Autoencoder sind neuronale Netze, die darauf trainiert werden, Daten zu komprimieren und wiederherzustellen. Für Deepfakes werden zwei Autoencoder eingesetzt: einer mit Bildern der Zielperson, also der Person, deren Gesicht im Video erscheinen soll, und einer mit Bildern der Originalperson aus dem ursprünglichen Video.

Der Prozess funktioniert folgendermaßen: Der Encoder wandelt das Gesicht der Originalperson in eine komprimierte Darstellung um. Anstatt diese Daten mit dem eigenen Decoder zu rekonstruieren, wird der Decoder der Zielperson verwendet. So entsteht ein Gesichtstausch, bei dem das Gesicht der Originalperson durch das der Zielperson ersetzt wird. Das Ergebnis ist ein Video, in

<sup>10</sup> Falade (2023).

dem die Zielperson scheinbar die Mimik und Lippenbewegungen der Originalperson ausführt, was den täuschend echten Effekt des Deepfakes erzeugt.<sup>11</sup> Möchte man hochqualitative Deepfakes wie in der Filmindustrie erzeugen, benötigt man in der Regel eine große Menge an Trainingsdaten. Tausende von Bildern oder Stunden an Rohvideomaterial sind nötig, damit die KI die Mimik, Gestik und andere Details einer Person präzise erlernen kann. Zusätzlich ist eine Trainingszeit von mehreren Tagen bis Wochen erforderlich, um diese hochauflösenden Deepfakes zu erstellen. Aufgrund des enormen Rechenaufwands sind solche Deepfakes oft nicht in Echtzeit nutzbar.<sup>12</sup>

Für eine Online-Videokonferenz mit begrenzter Qualität ist der Aufwand jedoch geringer. Hier können oft nur wenige Minuten Videomaterial oder sogar ein einzelnes Bild ausreichen. Das Training eines solchen Modells kann in wenigen Stunden bis zu einem Tag abgeschlossen sein.

### Generative Adversarial Networks (GANs)

Es gibt auch Systeme, die kein spezielles weiterführendes Training benötigen. Einmal auf eine Vielzahl von Gesichtern trainiert, können diese Modelle in Echtzeit ein Deepfake eines bisher ungesesehenen Gesichts erstellen. GANs bestehen aus zwei neuronalen Netzen: dem Generator und dem Diskriminator. Der Generator versucht, realistische Daten zu erzeugen, während der Diskriminator echte von gefälschten Daten zu unterscheiden versucht. Dieser Wettbewerb führt dazu, dass der

<sup>11</sup> Perov et al. (2022).

<sup>12</sup> Perov et al. (2022).

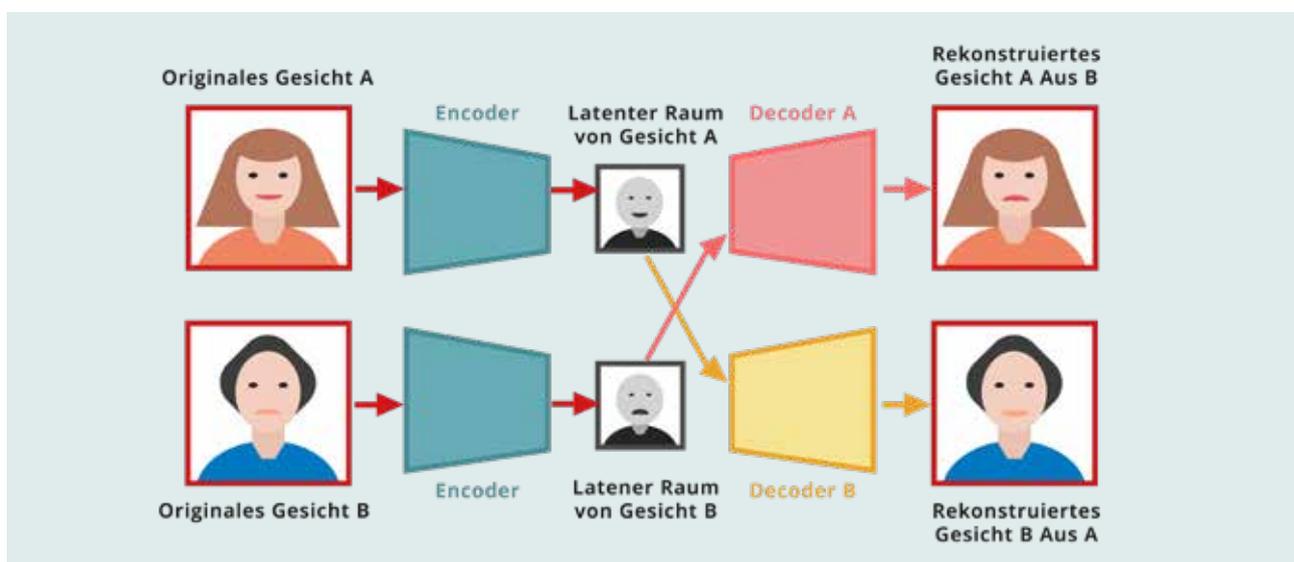


Abbildung 1: Deepfake mittels Autoencoder

Generator immer bessere und realistischere Deepfakes erstellt.<sup>13</sup> Aufgrund des fehlenden speziellen Trainings weist das erzeugte Bild meist weniger Ähnlichkeit zum Original auf. Um Echtzeit-Fakes erstellen zu können, ist die Auflösung dieser Modelle oft zu gering.

## Diffusionsmodelle

Diffusionsmodelle sind eine neuere Technologie, die hochrealistische Bilder erzeugen kann, indem sie den Prozess der Datenzerstörung umkehrt. Sie fügen zunächst Rauschen zu den Daten hinzu und trainieren dann ein neuronales Netzwerk, dieses Rauschen wieder in klare, realistische Bilder zu verwandeln. Diese Modelle sind in der Lage, äußerst komplexe und detailreiche Bild-Deepfakes zu erzeugen.<sup>14</sup> Allerdings ist der Prozess derzeit noch recht aufwendig, und es gibt bisher keine praktikablen, frei verfügbaren Echtzeit-Anwendungen für Videos. Es ist jedoch zu erwarten, dass sich diese Technologie weiterentwickeln wird und in Zukunft realistische Echtzeit-Videoanwendungen möglich sein könnten – ähnlich wie wir es bereits von Video-Generatoren wie OpenAI's Sora sehen.

## Funktionsweise von Voice Deepfakes

Voice Deepfakes oder genauer gesagt Voice Cloning nutzen KI-Modelle, die Sprachaufnahmen analysieren, um die einzigartigen Merkmale einer Stimme zu erlernen – wie Tonhöhe, Sprachmelodie und Rhythmus. Mit diesen Informationen erzeugen sie neue Sprachdaten, die der Originalstimme sehr ähnlich sind. Die meisten modernen Voice-Cloning-Systeme basieren auf einer speziellen Art von Autoencoder.

## Variational Autoencoder (VAE)

Variational Autoencoder bestehen ebenfalls aus einem Encoder-Decoder-Paar. Der Encoder nimmt die Sprachaufnahme und extrahiert daraus die charakteristischen Merkmale der Stimme. Diese Merkmale werden in einem komprimierten Code, dem sogenannten latenten Raum, gespeichert. Der Decoder nutzt diesen Code, um eine neue Audioausgabe zu erzeugen, die der Originalstimme entspricht. Der Encoder des VAE strukturiert den latenten Raum so, dass er nicht nur die Stimme klonen, sondern auch neue Variationen der Stimme erzeugen kann. Dadurch kann der VAE die Originalstimme nicht nur imitieren, sondern sie auch in verschiedenen Stilen und Nuancen reproduzieren, was ihm eine größere Flexibilität verleiht.<sup>15</sup>

<sup>13</sup> Dash und Sharma (2023).

<sup>14</sup> Bhattacharyya et al. (2024).

<sup>15</sup> Casanova et al. (2024); Bird und Lotfi (2023); Kim et al. (2021).

Früher benötigten solche Systeme viel Rechenzeit und Stunden an Aufnahmen, um eine Stimme realistisch zu klonen. Die Ausgaben klangen oft noch "mechanisch" oder verzerrt. Heute genügen bereits Minuten oder sogar wenige Sekunden, um täuschend echte Stimmklone zu erzeugen. Neben kommerziellen Anbietern, die Voice Cloning als kostengünstige Dienstleistung für automatische Voice-Overs oder Videoübersetzungen anbieten, gibt es auch frei verfügbare Software. Die synthetischen Stimmen können in Echtzeit erzeugt werden, direkt während eines Gesprächs in Telefonaten oder Videokonferenzen.<sup>16</sup>

## KI-gestützte Phishing-Angriffe: Funktionsweise

KI-gestützte Phishing-Angriffe nutzen Large Language Models wie ChatGPT, um personalisierte und überzeugende Phishing-Nachrichten zu erstellen. Automatisierte Systeme durchsuchen das Internet und soziale Netzwerke nach Informationen über Unternehmen und deren Mitarbeiter – wie Namen, Positionen und E-Mail-Adressen. Selbst scheinbar unwichtige Details werden verwendet, um spätere Phishing-Nachrichten authentischer zu gestalten.<sup>17</sup> Mit diesen Daten erstellen LLMs personalisierte E-Mails und Nachrichten, die täuschend echt wirken und mit passenden Ansprachen und Themen versehen sind.<sup>18</sup> Dadurch können Cyberkriminelle in Echtzeit massenhafte Phishing-Kampagnen durchführen, was die Effizienz und den potenziellen Schaden solcher Angriffe steigert.<sup>19</sup>

## Gegenmaßnahmen - Praktische Ansätze für KMU

Um sich vor den Gefahren von KI-unterstützten Cyberangriffen wie Deepfakes und Phishing zu schützen, müssen Unternehmen proaktiv handeln. Diese Bedrohungen lassen sich nicht allein durch technische Lösungen abwehren; sie erfordern gezielte Mitarbeiterschulungen und Sensibilisierung.

Es gibt zwar Softwarelösungen, die speziell zur Erkennung von Deepfakes entwickelt wurden.<sup>20</sup> Doch dies ist ein ständiges Katz-und-Maus-Spiel zwischen Angreifern und Detektoren, und die Wirksamkeit dieser Systeme ist in der Praxis oft unzuverlässig. Daher ist es ratsam, die öffentliche Verfügbarkeit von Bildern und Videos von Mitarbeitern und Führungskräften zu minimieren, um weniger Angriffsfläche für Deepfakes zu bieten. Unternehmenswebseiten und soziale Netzwerke sollten regelmäßig

<sup>16</sup> Bird und Lotfi (2023).

<sup>17</sup> Neupane et al. (2023).

<sup>18</sup> Neupane et al. (2023); Dash und Sharma (2023).

<sup>19</sup> Mirsky et al. (2021).

<sup>20</sup> Pei et al. (2024).

überprüft werden, um sensible Materialien zu entfernen oder den Zugang zu beschränken. Sensible Informationen sollten nicht öffentlich zugänglich sein, und der Zugriff auf solche Daten sollte streng kontrolliert und nur autorisierten Personen gewährt werden.<sup>21</sup> Ein bewusster Umgang mit Daten kann potenzielle Deepfake- oder Spear-Phishing-Angriffe bereits im Vorfeld erschweren.

Der wichtigste Schutz ist die Schulung der Mitarbeiter: Sie sollten über die Existenz, Funktionsweise und Risiken von Deepfakes und Phishing aufgeklärt werden. Durch regelmäßige Sensibilisierung sind sie besser darauf vorbereitet, verdächtige Inhalte zu erkennen. Ungewöhnliches Verhalten – wie ein abweichendes Profilbild, mangelnde Interaktion oder ein abruptes Beenden des Gesprächs – sollte als potenzielles Warnsignal erkannt werden. Im Zweifelsfall ist es ratsam, gezielte Fragen zu stellen, die nur die echte Person beantworten kann. E-Mails und Anrufe, vor allem aus unbekanntem Quellen, sollten kritisch betrachtet und nicht sofort vertraut werden. KMU sollten klare Prozesse zur Identitätsprüfung von Anrufern und Videokonferenzteilnehmern etablieren, um sich vor KI-unterstützten Angriffen zu schützen. Besonders bei sensiblen Themen oder finanziellen Transaktionen ist es wichtig, eine zusätzliche Sicherheitsebene einzuführen.<sup>22</sup> Beispielsweise können festgelegte Sicherheitsfragen oder Codewörter genutzt werden, um die Identität des Gesprächspartners zu verifizieren. Zusätzlich können einfache Methoden angewendet werden, um Deepfake-Systeme zu stören. Viele dieser Systeme nutzen Gesichtserkennung, bei der die Software die Position des Gesichts im Bild erkennt, um es anschließend manipulieren zu können. Sie funktionieren oft nicht richtig, wenn das Gesicht verdeckt wird oder man die gespreizten Finger langsam vor dem Gesicht vorbeiführt, im Zweifel müssen Mitarbeiter ihre Gesprächspartner dazu auffordern, um den Deepfake zu entlarven. Diese Methode, die auch bei Online-Identifizierungsverfahren von Banken eingesetzt wird, bietet jedoch keinen vollständigen Schutz, da fortgeschrittene Deepfake-Modelle solche Störungen umgehen könnten.

## Fazit

Die rasante Entwicklung von generativer KI und Deepfake-Technologien stellt eine erhebliche Gefahr für die Cybersicherheit dar. Von täuschend echten Video- und Audioaufnahmen bis hin zu KI-gestützten Phishing-Angriffen – die Möglichkeiten für Cyberkriminelle, Unternehmen zu schädigen, nehmen stetig zu. Diese Technologien haben nicht nur die Qualität von Angriffen verbessert, sondern auch die Eintrittsbarriere für weniger versierte Täter gesenkt.

Die Bedrohung durch Deepfakes und automatisierte Phishing-Angriffe zeigt, dass technische Lösungen allein nicht ausreichen. Besonders KMU müssen auf Mitarbeiterschulungen und ein erhöhtes Bewusstsein setzen. Nur durch gezielte Sensibilisierung und die Einführung klarer Sicherheitsprozesse kann man sich effektiv gegen diese Art von Angriffen schützen.

Obwohl spezialisierte Softwarelösungen zur Erkennung von Deepfakes existieren, bleibt der Kampf zwischen Angreifern und Verteidigern ein fortwährendes Katz- und-Maus-Spiel. Deshalb ist es entscheidend, proaktiv zu handeln – sei es durch die Kontrolle öffentlich verfügbarer Informationen oder durch die Entwicklung eines gesunden Misstrauens gegenüber ungewöhnlichen Anfragen. Letztlich liegt der Fokus darauf, die menschlichen Faktoren zu stärken. Geschulte und aufmerksame Mitarbeiter machen den entscheidenden Unterschied im Kampf gegen KI-gestützte Cyberangriffe aus.

Das Mittelstand-Digital Zentrum Hamburg bietet KMU Unterstützungsangebote zu generativen KI-Technologien. Diese umfassen Beratungen zur Identifikation von Risiken und Entwicklung von Schutzstrategien, aber auch Schulungen, um die Potenziale der Technologien im Bereich Kundenkommunikation und Marketing zu nutzen.

## Literatur

- Bhattacharyya C, Wang H, Zhang F, Kim S, Zhu X. Diffusion Deepfake [Internet]. arXiv.org. 2024 [cited 2024 Sep 19]. Available from: <https://arxiv.org/abs/2404.01579v1>
- Bird JJ, Lotfi A. Real-time Detection of AI-Generated Speech for DeepFake Voice Conversion [Internet]. arXiv.org. 2023 [cited 2024 Sep 19]. Available from: <https://arxiv.org/abs/2308.12734v1>
- Bovenschulte M. Deepfakes – Manipulation von Filmsequenzen. 2019 Jan 1 [cited 2024 Sep 19]; Available from: <https://publikationen.bibliothek.kit.edu/1000133910>
- Brien J. KI-Deepfake-Betrug gescheitert: Ferrari-Manager stellt entscheidende Frage [Internet]. t3n Magazin. t3n Magazin; 2024 [cited 2024 Sep 19]. Available from: <https://t3n.de/news/ki-deepfake-betrug-ferrari-manager-entscheidene-frage-1638005/>
- Bundesamt für Sicherheit in der Informationstechnik [Internet]. Bundesamt für Sicherheit in der Informationstechnik. BSIWEB; 2022 [cited 2024 Sep 19]. Available from: [https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes\\_node.html](https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes_node.html)
- Casanova E, Davis K, Gölge E, Gökner G, Gulea I, Hart L, et al. XTTS: a Massively Multilingual Zero-Shot Text-to-Speech Model [Internet]. arXiv.org. 2024. Available from: <https://arxiv.org/abs/2406.04904>

21 Neupane et al. (2023).

22 Bundesamt für Sicherheit in der Informationstechnik (2022).

Dash B, Sharma P. Are ChatGPT and Deepfake Algorithms Endangering the Cybersecurity Industry? A Review. IJEAS. 2023;10(1). doi: 10.31873/IJEAS.10.1.01.

Falade P. Decoding the Threat Landscape : ChatGPT, FraudGPT, and WormGPT in Social Engineering Attacks. J Sci Res Comput Sci Eng Inf Technol [Internet]. 2023;9(5):185-98. Available from: <https://arxiv.org/ftp/arxiv/papers/2310/2310.05595.pdf>

Hurtz S. Deepfake-Betrug: Angestellter überweist 24 Millionen Euro ins Nichts [Internet]. Süddeutsche.de. Süddeutsche Zeitung; 2024 [cited 2024 Sep 19]. Available from: <https://www.sueddeutsche.de/wirtschaft/deepfake-betrug-videokonferenz-hongkong-1.6344209>

Kim J, Kong J, Son J. Conditional Variational Autoencoder with Adversarial Learning for End-to-End Text-to-Speech [Internet]. arXiv.org. 2021 [cited 2024 Sep 19]. Available from: <https://arxiv.org/abs/2106.06103v1>

Mirsky Y, Demontis A, Kotak J, Shankar R, Gelei D, Yang L, et al. The Threat of Offensive AI to Organizations [Internet]. arXiv.org. 2021 [cited 2024 Sep 19]. Available from: <https://arxiv.org/abs/2106.15764>

Neupane S, Fernandez IA, Mittal S, Rahimi S. Impacts and Risk of Generative AI Technology on Cyber Defense [Internet]. arXiv.org. 2023 [cited 2024 Sep 19]. Available from: <https://arxiv.org/abs/2306.13033v1>

Pei G, Zhang J, Hu M, Zhang Z, Wang C, Wu Y, et al. Deepfake Generation and Detection: A Benchmark and Survey [Internet]. arXiv.org. 2024 [cited 2024 Sep 19]. Available from: <https://arxiv.org/abs/2403.17881v4>

Perov I, Gao D, Chervoniy N, Liu K, Marangonda S, Umé C, et al. DeepFaceLab: A simple, flexible and extensible face swapping framework. arXiv:200505535 [cs, eess] [Internet]. 2020 May 20; Available from: <https://arxiv.org/abs/2005.05535>

## Autoren



**Jan Fischer** M.Sc. ist wissenschaftlicher Mitarbeiter am Forschungs- und Transferzentrum „Business Innovation Lab (BIL)“ der HAW Hamburg. Nach seinem Studium der Wirtschaftsinformatik (M.Sc.) an der Universität Oldenburg arbeitet er seit 2017 intensiv an der Digitalisierung von kleinen und mittleren Unternehmen, mit Fokus auf Produktionsplanung, Supply Chain Management und den Einsatz künstlicher Intelligenz. Er ist an verschiedenen Projekten beteiligt, darunter das "Mittelstand-Digital Zentrum Hamburg" und der "European Digital Innovation Hub Hamburg". Im Rahmen seiner Tätigkeiten beschäftigt er sich zudem intensiv mit der praktischen Anwendung von (generativer) KI sowie deren Auswirkungen u.A. auch auf die Cybersicherheit.



**Prof. Dr.-Ing. Axel Wagenitz** ist Professor für Wirtschaftsinformatik an der HAW Hamburg und leitet das "Business Information Lab". Als Informatiker mit Promotion im Bereich Ingenieurwesen entwickelte er am Fraunhofer IML Ansätze zur Planung von Logistiknetzwerken. Dort war er auch an der Umsetzung von über 50 Forschungs- und Industrieprojekten beteiligt. Als Projektleiter am Mittelstand-Digital Zentrum Hamburg und Mitbegründer des Forschungs- und Transferzentrums berät er Unternehmen in den Bereichen Digitalisierung, Künstliche Intelligenz und Logistik. Zudem ist er weiterhin als wissenschaftlicher Berater im Supply Chain Engineering tätig.

Das **Mittelstand-Digital Zentrum Hamburg** unterstützt kleine und mittlere Unternehmen (KMU) in Hamburg und der Metropolregion dabei, ihre Wettbewerbsfähigkeit durch innovative Digitalisierungslösungen nachhaltig zu stärken. Das interdisziplinäre Team, ergänzt durch KI-Trainer und Klima-Coache, bietet kostenfreie, anbieterneutrale Beratungen sowie praxisorientierte Veranstaltungsformate, die auf die individuellen Bedürfnisse der Unternehmen abgestimmt sind.

Schwerpunkte liegen auf Themen wie

- ▶ Künstliche Intelligenz,
- ▶ Prozessoptimierung,
- ▶ digitale Geschäftsmodelle,
- ▶ Nachhaltigkeit,
- ▶ AR/VR,
- ▶ Cybersecurity.

Die KMU profitieren dabei von der Expertise des Zentrums sowie von einem regionalen und bundesweiten Netzwerk, das Zugang zu umfassendem Fachwissen und praktischen Umsetzungsbeispielen bietet.

<https://digitalzentrum-hamburg.de/>

