

# KI in der Medienbranche: Content-Erstellung, Mediendesign und Produktion

Dieser Artikel beschäftigt sich mit der grundlegenden Frage: Wie verändert Künstliche Intelligenz (KI) die Medienbranche? Zum jetzigen Zeitpunkt gibt es viele technologische Entwicklungen, die insbesondere durch neue Erkenntnisse im Bereich des maschinellen Lernens mit großen Datenmengen ermöglicht wurden.

Unsere Erwartung ist, dass diese intelligenten Technologien in kurzer Zeit weite Teile der Medienbranche fundamental verändern werden. Betrachtet man genauer, was aktuell erforscht und erprobt wird, fallen die folgenden zentralen Richtungen auf:

- Unterstützung kreativer Tätigkeiten durch KI-Werkzeuge und Anwendungen
- Nutzung von bestehenden Inhalten zur Erzeugung neuer Inhalte und Formate
- Verbesserung der Werkzeuge für die Medienproduktion und Verarbeitung

Zum jetzigen Zeitpunkt sind viele der Technologien und die darauf aufbauenden Werkzeuge und Anwendungen noch in den Kinderschuhen. Dennoch lassen sich Trends erkennen. Es erscheint ähnlich wie 1992, als viele in der Informatik und darüber hinaus vom WWW begeistert waren, obwohl die Funktionalität nur rudimentär vorhanden war. Aber trotz der Limitierungen in der Technologie waren bereits Anfang der 1990er die Richtung und das Veränderungspotenzial deutlich zu sehen. Ähnlich ist es heute. KI-Werkzeuge stehen am Anfang, aber die Richtung der Veränderung ist klar zu sehen, und man kann sich kaum vorstellen, dass die Veränderung aufgehalten wird.

Betrachtet man diese Veränderungen im Kontext der gesamten Medienbranche, wird klar, dass eine große Automatisierungswelle entlang der gesamten Wertschöpfungskette sehr wahrscheinlich ist. Das wird grundlegend die Frage stellen, wie sich in Zukunft Medienschaffende differenzieren können und wo originäre Wertschöpfung entsteht. Es steht außer Frage, dass durch neue Werkzeuge neue Formate entstehen. Der bevorstehende Umbruch durch KI ist in der Dimension ähnlich, wie die Umstellung von analogen Medien (Film und Tonband) auf digitale Technologien. Insbesondere die Frage, welche Qualifikationen und Fähigkeiten Mitarbeitende benötigen, um Medienformate zu erstellen, wird ganz grundlegend neu gestellt werden.



© Joas Strecker

## Prof. Dr. Albrecht Schmidt

Ludwig-Maximilians-Universität München  
Fakultät für Mathematik, Informatik und Statistik  
Institut für Informatik  
Human-Centered Ubiquitous Media  
albrecht.schmidt@ifi.lmu.de

**Schlüsselbegriffe:** Künstliche Intelligenz | KI-Werkzeuge | Medienbranche | Kreativität | ChatGPT | Sprachmodelle | Filmprojekte | Neue Medienformate | Medienproduktion

## 1 Unterstützung kreativer Tätigkeiten durch KI-Werkzeuge

Bisher schreiben kreative Menschen Texte und Drehbücher, zeichnen Bilder, fotografieren und filmen, komponieren Musik, spielen Theater, sind als Schauspieler vor der Kamera oder gestalten Werbekampagnen. Diese kreativen Aufgaben werden durch neue Werkzeuge verändert und in Teilen automatisiert.

Exemplarisch werden im Folgenden wesentliche Schritte in einem Filmprojekt betrachtet. Es werden Trends aufgezeigt, wie diese Schritte sich durch KI verändern können. Die folgenden Beispiele zeigen mögliche Richtungen auf, wie KI genutzt werden könnte. Diese Beispiele basieren auf Technologien und Forschungsprototypen, welche in den letzten Jahren veröffentlicht wurden. Diese Dinge sind bereits jetzt möglich. Die Qualität der Ergebnisse muss natürlich weiter diskutiert werden.

Beispiele für Aufgaben in einem Filmprojekt:

- Ideen entwickeln, eine Geschichte oder ein Drehbuch schreiben
- Ein Storyboard gestalten und einen Plan für die Aufnahmen erarbeiten
- Filmaufnahmen machen, Spezialeffekte und Computergrafik
- Szenen anordnen und schneiden
- Musik zum Film auswählen oder komponieren
- Anpassung für verschiedene Zielgruppen, Formate und Kanäle

### Geschichten schreiben

Große künstliche Sprachmodelle, wie GPT-3, sind in der Lage, Texte zu erzeugen<sup>1</sup>. Dies kann als Antwort auf eine Frage geschehen oder auch durch Ausformulieren von Stichwörtern. Die Qualität ist durchaus sehr unterschiedlich. Eine iterative Nutzung von ChatGPT<sup>2,3</sup>, wie in Abbildung 1 dargestellt, zeigt, wie man mit solchen Technologien eine Geschichte und ein Drehbuch entwickeln kann. Hier wird der kreative Prozess mit Werkzeugen unterstützt. Die Arbeit wird dadurch beschleunigt. Man kann sich davon auch nur inspirieren lassen oder einige Alternativen erzeugen.

<sup>1</sup> Elkins, K., & Chun, J. (2020). Can GPT-3 pass a Writer's turing test?. *Journal of Cultural Analytics*, 5(2).

<sup>2</sup> van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, 614(7947), 224-226. doi: <https://doi.org/10.1038/d41586-023-00288-7>

<sup>3</sup> <https://openai.com/blog/chatgpt>

## Abbildung 1: Iterativer Dialog zur Erstellung eines Drehbuchs

*Blau und kursiv sind die Prompts (Vorgaben) und dann folgen die Antworten von ChatGPT.*

*Welche inhaltlichen Elemente sollte ein Krimi haben, bei dem Zuschauer mitraten können?*

Ein Krimi, bei dem Zuschauer mitraten können, sollte in der Regel folgende inhaltliche Elemente enthalten:

1. Ein Verbrechen oder eine Straftat: Das ist die Ausgangslage für den Krimi und das Rätsel, das gelöst werden muss.
2. Verdächtige Personen: Es müssen mindestens ein paar Personen geben, die in Frage kommen könnten, das Verbrechen begangen zu haben.
3. Indizien und Beweise: [...]
4. Eine gut durchdachte Handlung: [...]
5. Überraschende Wendungen: [...]
6. Ein cleveres Auflösung: Am Ende sollte das Rätsel auf eine überzeugende und befriedigende Art gelöst werden.

*Beschreibe die Person einer jungen Ermittlerin. Sie studiert Informatik, mag Rätsel und ist viel im Internet unterwegs.*

Die junge Ermittlerin, die Informatik studiert, Rätsel mag und viel im Internet unterwegs ist, könnte folgendermaßen beschrieben werden:

1. Intellektuell begabt: Sie studiert Informatik an einer renommierten Universität und ist dadurch sehr klug und analytisch veranlagt.
2. Rätselfreudig: Sie mag Rätsel und Herausforderungen und verbringt viel Zeit damit, knifflige Probleme zu lösen.
3. Technikbegeistert: [...]
4. Detailorientiert: [...]
5. Unerschrocken: [...]
6. Unkonventionell: [...]

*Wie könnte eine Eröffnungsszene aussehen, die in München im Englischen Garten spielt und einen Zusammenhang mit dem Internet hat?*

[...] Es ist ein sonniger Tag im Englischen Garten und viele Menschen genießen das schöne Wetter. Die junge Ermittlerin sitzt auf einer Bank und blickt auf ihr Laptop, als plötzlich ein Mann auf sie zustürmt. Er ist außer Atem und scheint in Panik zu sein.

„Sie müssen mir helfen! Mein Sohn ist verschwunden!“, ruft er aufgeregt.

Die Ermittlerin blickt den Mann besorgt an. „Beruhigen Sie sich. Erzählen Sie mir, was passiert ist.“ →

Der Mann atmet tief durch und erklärt, dass sein Sohn vor ein paar Stunden in den Englischen Garten gegangen ist, um seine neue Drohne auszuprobieren. Seitdem hat er nichts von ihm gehört und auch sein Handy ist ausgeschaltet.

Die Ermittlerin [...] beginnt, Fragen zu stellen. Sie erfährt, dass der Sohn des Mannes ein begeisterter Drohnenpilot ist und im Internet eine große Community hat. Plötzlich fällt ihr etwas ein.

„Hat Ihr Sohn in letzter Zeit irgendwelche merkwürdigen Nachrichten bekommen? Oder hat er sich vielleicht mit jemandem im Internet getroffen?“

Der Mann überlegt kurz. „Ja, jetzt, wo Sie es sagen. Er hat vor ein paar Tagen von einem Unbekannten eine Nachricht bekommen, in der ihm ein wertvolles Ersatzteil für seine Drohne angeboten wurde. Er war sehr aufgeregt und hat sich sofort mit dem Absender getroffen. Seitdem habe ich nichts mehr von ihm gehört.“

Die Ermittlerin runzelt die Stirn. „Das könnte eine heiße Spur sein. Kommen Sie, wir müssen so schnell wie möglich herausfinden, wo sich Ihr Sohn aufhält.“

Sie stehen auf und gehen los, während die Ermittlerin bereits ihre nächsten Schritte plant. Sie ahnt, dass dieser Fall sie tief in die Welt des Internets und der Drohnen führen wird.

## Schreibe ein Drehbuch für diese Eingangsszene

[...] SZENE: Englischer Garten, München

Es ist ein sonniger Tag und viele Menschen genießen das schöne Wetter im Englischen Garten. Die junge Ermittlerin sitzt auf einer Bank und blickt auf ihren Laptop, als plötzlich ein Mann auf sie zustürmt.

MANN: (außer Atem) Sie müssen mir helfen! Mein Sohn ist verschwunden!

ERMITTLERIN: Beruhigen Sie sich. Erzählen Sie mir, was passiert ist.

MANN: (holt tief Luft) Mein Sohn ist vor ein paar Stunden in den Englischen Garten gegangen, um seine neue Drohne auszuprobieren. Seitdem habe ich nichts von ihm gehört und auch sein Handy ist ausgeschaltet.

ERMITTLERIN: (zieht Notizblock hervor) Okay, lassen Sie mich das notieren. Ihr Sohn ist also verschwunden, seit er in den Englischen Garten gegangen ist. Hat er irgendwelche Freunde dort getroffen?

MANN: Nein, er ist alleine gegangen. Er ist ein begeisterter Drohnenpilot und hat im Internet eine große Community.

ERMITTLERIN: (überlegt kurz) Hat Ihr Sohn in letzter Zeit irgendwelche merkwürdigen Nachrichten bekommen? Oder hat er sich vielleicht mit jemandem im Internet getroffen?

Quelle: ChatGPT

## Storyboards gestalten

Hat man ein Drehbuch, kann man daraus ein Storyboard erstellen. Auch in diesem Schritt kann man sich vorstellen, dass eine KI weiterhelfen kann. In Abbildung 2 hat Dall-E-2<sup>4</sup> das erste Bild der Eingangsszene der Geschichte als Strichzeichnung erstellt. Auch hier kann man iterativ vorgehen und weitere Versionen generieren.

## Filmen oder Inhalte erzeugen, Spezialeffekte und Computergrafik

Der Aufbau des Sets und das Aufzeichnen des Films ist ein Herzstück. Unterschiedliche Formen der Unterstützung zeichnen sich hier ab. Einerseits lässt sich mit neuen Methoden vieles sehr leicht nachbearbeiten. Hintergründe können ausgetauscht werden, Wetter- und Lichtbedingungen angepasst werden oder auch Teile des Hintergrunds verändert werden. Solche Verfahren bauen auf Algorithmen für die semantische Segmentierung von Bildern auf<sup>5</sup>. Auch die Gestaltung mit Blende, Belichtung und Fokus lässt sich nachträglich verändern. Man kann nun argumentieren, dass ein weichgezeichneter Hintergrund nicht den gleichen künstlerischen Wert hat, wie eine sorgfältig eingestellte Tiefenschärfe. Mittelfristig wird es aber schwer sein, im Endergebnis einen Unterschied wahrzunehmen. Man kann daher erwarten, dass der Arbeitsaufwand für das Filmen wie auch für die Nachbearbeitung geringer wird.

In einem späteren Abschnitt wird das Thema Deep Fake noch besprochen. Bezogen auf das Filmen könnte hier eine Utopie (oder Dystopie) sein, dass die Szenen mit generischen Schauspielerinnen und Schauspielern gedreht werden. Erst in der Nachbearbeitung werden diese dann mit den entsprechenden Stars ersetzt.

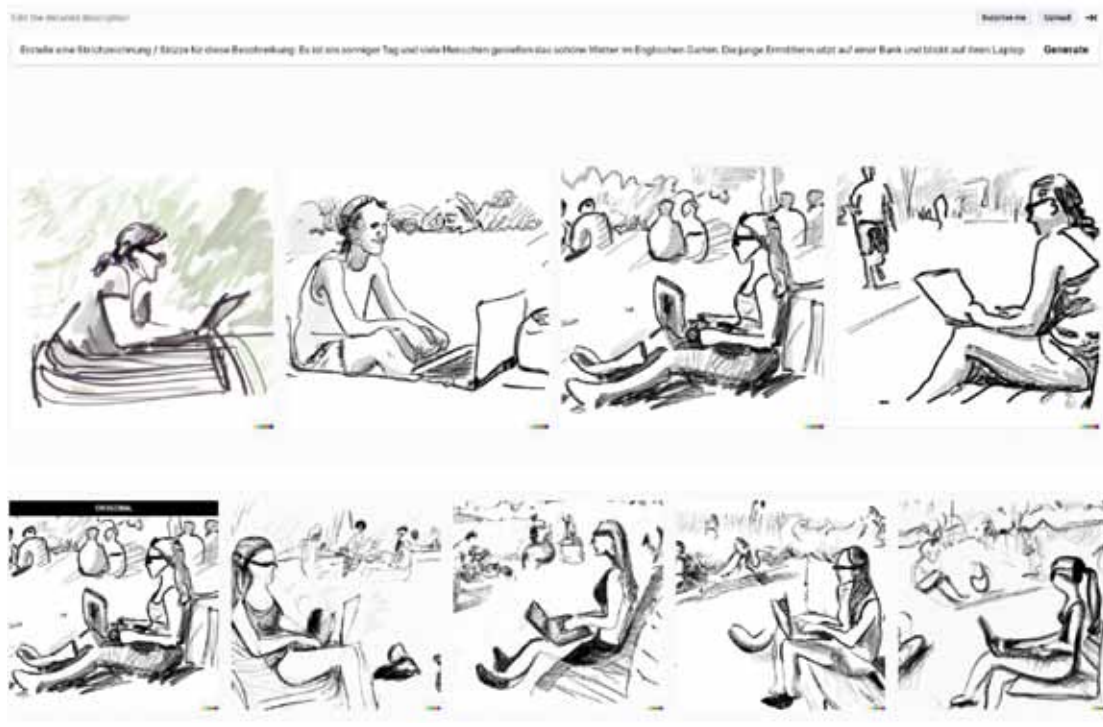
Das Erzeugen von einzelnen Bildern funktioniert für verschiedene Anwendungsbereiche bereits sehr gut. Dall-E-2 und Stable Diffusion<sup>6</sup> sind hier aktuelle Beispiele, die auf Basis einer textuellen Beschreibung ein Bild erzeugen können. Ganze Filmszenen aus einem Text erzeugen zu lassen, ist im Moment in guter Qualität noch sehr schwierig, es ist aber bereits mit aktuell vorhandenen Ansätzen grundsätzlich möglich. Man könnte so zum Beispiel für die Szenen die zentralen Bilder erzeugen (ähnlich wie mit Dall-E-2) und dann die Teile zwischen den Bildern generieren. So etwas ist natürlich für einen Zeichentrickfilm wesentlich einfacher als für einen Spielfilm mit Menschen. Die Unterscheidung, was ist Filmaufnahme, was ist Spezialeffekte und was ist Computergrafik, wird voraussichtlich verschwimmen.

<sup>4</sup> <https://openai.com/product/dall-e-2>

<sup>5</sup> Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440). <https://arxiv.org/abs/1411.4038>

<sup>6</sup> <https://stablediffusionweb.com/>

### Abbildung 2: Generiertes und verfeinertes Storyboard, erstellt mit DALL-E 2



Quelle: Dall-E-2; <https://openai.com/product/dall-e-2>

### Szenen anordnen und schneiden

Das Schneiden von Filmen oder Beiträgen ist zeitaufwendig. Hier können KI-Werkzeuge sehr hilfreich sein. Inzwischen ist ein Suchen in Bildern und Filmsequenzen technisch möglich, und auch die Dialoge können automatisch transkribiert werden<sup>7</sup>. Auch das Vergleichen und Bewerten von verschiedenen Aufnahmen einer Szene kann in der Zukunft zumindest teilweise automatisiert werden. Es ist somit vorstellbar, dass mit dem Drehbuch und den aufgenommenen Filmsequenzen automatisch ein Rohschnitt erzeugt werden kann. Auch für die Ausarbeitung der Details ist es möglich, KI zu verwenden. Für bestimmte Übergänge (z. B. Match Cut) könnten mit Bilderkennungsalgorithmen Positionen im Rohmaterial gesucht werden, die sich dafür eignen.

Auch beim Schnitt von Ton bieten die Anwendungen bereits heute algorithmische Unterstützung. Das Anordnen und Synchronisieren von verschiedenen Tonquellen mit der Filmsequenz kann über Erkennungsalgorithmen erledigt werden. Ähnlich kann auch der genaue Schnittpunkt (z. B. beim Audio-Material der Nulldurchgang des Signals) automatisiert

gewählt werden. Auch das Verbessern von Aufnahmen durch Entfernen von Füllwörtern wie „äh“ wird zunehmend automatisiert werden<sup>8</sup>. Hier ist der Übergang zwischen klassischen Werkzeugen und Algorithmen und KI-basierten Werkzeugen fließend.

### Musik zu Film auswählen oder komponieren

Ähnlich wie Dall-E-2 für Bilder gibt es auch Modelle für die Erzeugung von Musik. Mit musicLM<sup>9</sup> gibt es einen Ansatz und eine Implementierung, die auf Basis von textueller Eingabe eine Musik erzeugen kann. So kann mit der Eingabe „bedrohliche dunkle Musik, welche anschwillt und lauter wird mit Streichinstrumenten; Schlaginstrument, welches einen Takt vorgibt, der schneller wird“ ein entsprechendes Musikstück erzeugt werden. Aktuell funktioniert das System musicLM mit englischen Eingaben, aber durch die Kombination mit automatischer Übersetzung lassen sich Prompts in anderen Sprachen formulieren.

Auch Liedtexte und Melodien lassen sich generieren. Das Lied „You can't take my door“<sup>10</sup> ist ein Lied, was mit ei-

<sup>7</sup> Beispielsweise bieten Programme wie Zoom eine automatische Transkription der Sprache in Echtzeit an. Auch YouTube kann automatisiert Untertitel erzeugen, welche dann direkt genutzt oder auch heruntergeladen und weiter bearbeitet werden können.

<sup>8</sup> Zhu, G., Caceres, J. P., & Salamon, J. (2022). Filler Word Detection and Classification: A Dataset and Benchmark. arXiv preprint arXiv:2203.15135. <https://arxiv.org/abs/2203.15135>

<sup>9</sup> Agostinelli, A., Denk, T. I., Borsos, Z., Engel, J., Verzettl, M., Caillon, A., ... & Frank, C. (2023). Musiclm: Generating music from text. arXiv preprint arXiv:2301.11325. <https://google-research.github.io/seanet/musiclm/examples/>

<sup>10</sup> <https://open.spotify.com/track/3FlerPYLMqbCNqFzJO0E2h>

nem Model erzeugt wurde. Hier war die Aufforderung, einen Country Song zu erzeugen. Das Ergebnis hört sich durchaus wie ein Country Song an; der Text ist repetitiv und nur bedingt sinnhaftig, was aber wohl auch auf viele andere Liedtexte zutrifft.

## Anpassung für verschiedene Zielgruppen, Formate und Kanäle

Online-Plattformen wie YouTube bieten bereits seit einiger Zeit die Möglichkeit, Untertitel hinzuzufügen. Diese lassen sich natürlich dann auch automatisiert übersetzen. Mit Computer-Vision-Algorithmen, welche die Inhalte von Bildern erkennen, kann man sich vorstellen, dass solche Untertitel auch um Audiobeschreibungen von Szenen automatisch erweitert werden können. Aktuell sind diese Untertitel oft noch nicht optimal, aber natürlich besser als keine Untertitel zu haben, insbesondere wenn man nicht hören kann oder die Originalsprache nicht spricht.

Die Anpassung an verschiedene Formate (z. B. 21:9, 16:9, 4:3) und verschiedene Abspielgeräte (Telefon, Tablet, TV, Projektor) ist heute meist statisch. Man wählt zwischen dem Abschneiden von Inhalten oder der verkleinerten Darstellung mit schwarzen Balken. Hier wäre mit Bilderkennung und Bildgenerierung eine dynamische Anpassung an das Zielformat möglich. Es könnte dynamisch auf die Bereiche fokussiert werden, die aktuell relevant sind (z. B. in einer Fußballübertragung auf die Stelle, an der der Ball ist). Auch die Einstellungsgrößen (halbnah vs. nah) oder die Wahl des Bildausschnitts könnten für die jeweilige Betrachtungssituation mit KI optimiert werden.

## Welche Aufgaben bleiben beim Menschen?

Diese Beispiele zeigen, dass in sehr vielen Bereichen KI eingesetzt werden kann, um einen Mehrwert in der Herstellung (Vereinfachung, Effizienz, Kostenersparnis) und auch beim Konsumenten (Individuierung, Anpassung) zu erreichen. Das wird natürlich verändern, was wir als kreative Tätigkeit betrachten und wo detaillierte Kontrolle erwünscht ist.

Gleichzeitig ist schon mit den jetzt verfügbaren Werkzeugen klar, dass diese die Produktivität erhöhen werden. Es stellt sich daher eher die Frage, wie und wo diese Werkzeuge eingesetzt werden sollen und nicht, ob sie eingesetzt werden. Es wird aber sicher notwendig sein zu lernen, wie bestimmte Aufgaben am besten unterstützt werden.

## 2 Neue Medienformate werden entstehen

Die auf KI-Basis entstehenden Werkzeuge sind sehr mächtig und werden neue Medienformate ermöglichen. Ähnlich wie die Einführung von digitaler Filmaufzeichnung die Medienbranche stark verändert hat, werden auch hier gänzlich

neue Formate entstehen. Insbesondere die höhere Automatisierung bei der Erstellung und die Wiederverwendung (ähnlich wie Remixes in der Musik) spielen dabei voraussichtlich eine große Rolle. So ist es gut vorstellbar, dass Produktionen, welche hochautomatisiert sind, selbst für sehr kleine Zielgruppen (bis zur Einzelperson) wirtschaftlich umgesetzt werden können. Die Möglichkeit zur Personalisierung und individuellen Inhaltserstellung wird in einer neuen Dimension möglich werden. Der eigene Tatort nicht nur für München oder Berlin, sondern auch für eine Kleinstadt, könnte damit wirtschaftlich sinnvoll werden.

## Nutzung von bestehenden Inhalten zur Erzeugung neuer Inhalte und Formate

Die Menge an bestehenden Inhalten ist bisher schon sehr groß und unüberschaubar. Auf YouTube, Instagram und TikTok werden jede Sekunde große Mengen an neuen „Filmen“ erzeugt. YouTube gibt an, dass jede Minute über 500 Stunden Videos auf die Plattform hochgeladen werden<sup>11</sup>. Zunehmend erzeugen auch technische Systeme (Überwachungskameras, Kameras in selbstfahrenden Fahrzeugen) große Mengen an Videos, welche aber bisher nur bedingt verfügbar sind. Im WWW und auf sozialen Medien wie Twitter sind große Mengen an Text und Bildern vorhanden, und auch hier kommen jeden Moment neue Inhalte hinzu. Für die Erzeugung neuer Formate und Inhalte auf Basis der bestehenden gibt es grundsätzlich zwei Ansätze:

- Explizite Kombination bestehender Inhalte zu einem neuen Medienbeitrag
- Lernen von bestehenden Inhalten und die Erzeugung „neuer Inhalte“

## Explizite Kombination von Inhalten

Der Ansatz ist hierbei, dass bestehende Inhalte und insbesondere Filme genutzt werden, um einen Beitrag zu erstellen. Als Gedankenpiel: Wir möchten eine 60-minütige Dokumentation über die Entwicklung der deutsch-russischen Beziehung vom Fall der Mauer bis in die heutige Zeit erstellen. Wenn es nun Werkzeuge gibt, mit denen effizient in großen Medienmengen gesucht werden kann, ist es einfach vorstellbar, dass aus bestehendem Material eine spannende und neue Dokumentation erstellt werden kann. Bestehendes Material könnten dabei Nachrichtensendungen, aufgezeichnete Interviews, Dokumentationen zu spezifischen Ereignissen, aber auch Heimvideos oder Video-Kurzbeiträge in sozialen Medien sein. Wenn wir Zugriff auf all diese „Rohmaterialien“ haben und diese sinnvoll durchsuchen können, wird es möglich sein, eine „Schnittliste“ zu erstellen. Durch

<sup>11</sup> <https://blog.youtube/press/>



**Abbildung 3: Generierte zufällige Gesichter**



Quelle: Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4401-4410). <https://github.com/NVlabs/stylegan>

automatisierte Nachbearbeitung, Anpassung der Formate und Übersetzung kann somit eine neue Dokumentation erstellt werden.

Aktuell ist eine solche Vorgehensweise noch schwierig, da die technischen Werkzeuge noch nicht vollständig vorhanden sind. Hier zeichnet sich aber ab, dass es diese in den nächsten Jahren geben wird. Gleichzeitig ist ein solcher Ansatz rechtlich aktuell fast unmöglich. Wenn wir davon ausgehen, dass die einzelnen Beitragsteile im Durchschnitt 10 Sekunden lang sind, müsste man für eine 60-Minuten-Dokumentation die Rechte an 360 Filmstücken erhalten. Das ist manuell kaum vorstellbar, dennoch könnten hier mittelfristig auch KI-Systeme automatisiert Verträge aushandeln, um so etwas zu ermöglichen.

### Lernen von Inhalten für generative Modelle

Ein alternativer Ansatz, der sehr aktuell ist, sind lernende Systeme und generative Verfahren. Hierbei werden bestehende Inhalte nicht direkt verwendet, sondern diese werden in neuronalen Netzen „gelernt“. Auf Basis des Gelernten werden dann „neue Inhalte“ erzeugt. Beispiele hierfür sind Modelle zur Bildgenerierung (Dall-E-2) und Technologien für die Erzeugung von Musik (musicLM). Diese Werkzeuge haben gemeinsam, dass auf Basis einer textuellen Beschreibung neue Medienstücke erzeugt werden. Auch große Sprachmodelle (GPT-3, ChatGPT) verfolgen einen ähnlichen Ansatz. Hier sind große Textmengen „gelernt“, und es wird als Antwort auf eine Anfrage Text erzeugt.

Durch die Erzeugung neuer Grafiken und Bilder können Medieninhalte mit gänzlich neuen Inhalten entstehen. So können z. B. Personen erfunden werden. Es werden also Bilder erzeugt von Personen, die nicht in dieser Form existieren. So gibt es einen Generator für zufällige Gesichter<sup>12</sup> – keines der erzeugten Gesichter existiert, siehe Abbildung 3. Sie sehen aber trotzdem realistisch aus<sup>13</sup>.

Beim Lernen von bestehenden Inhalten und bei der Generierung neuer Medien ist die Rechtfertigung wesentlich komplexer<sup>14</sup>. So stellt sich z. B. die Frage, ob ein Bild, das mit dem Kommando „Zeichne einen Bären, der angelt und telefoniert im Stil von Walt Disney“ eigentlich von Walt Disney lizenziert werden müsste. Hier hat erst die Diskussion begonnen, und Medienschaffende möchten hier sinnvoll beteiligt werden. Die Frage der Urheberschaft hat neben dem rechtlichen Aspekt auch eine philosophisch-ethische Dimension. Ist die Erstellung von Inhalten mit KI eine kreative Leistung der Software oder nur eine veränderte Reproduktion der Inhalte, welche für das Training des Systems verwendet wurden?

<sup>12</sup> <https://this-person-does-not-exist.com/en>

<sup>13</sup> Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4401-4410). <https://github.com/NVlabs/stylegan>

<sup>14</sup> Franceschelli, G., & Musolesi, M. (2022). Copyright in generative deep learning. Data & Policy, 4, e17. <https://www.cambridge.org/core/journals/data-and-policy/article/copyright-in-generative-deep-learning/C401539FDF79A6AC-6CEE8C5256508B5E>

**Abbildung 4: Verschiedene Beispiele Echtzeit-Bildmanipulationen in Zoom**



Quelle: Screenshots vom Autor bei der Verwendung von Zoom.

## Retusche und Deep Fakes

Das Verändern von Inhalten, insbesondere das Manipulieren von Bildern, ist kein neues Phänomen. Jedoch wird der Aufwand für die Änderungen von Inhalten, für das Retuschieren von Bildern, durch Künstliche Intelligenz stark reduziert. Selbst das vollständige und automatische Ändern von umfangreichen Inhalten, wie z. B. dem Gesicht von Schauspielern in einem Film, ist inzwischen technisch möglich.

### Automatische Retusche und Greenscreens ohne Greenscreen

Aktuelle Bildbearbeitungsprogramme bieten inzwischen Funktionen, um Teile von Bildern einfach auszutauschen. So kann ein Strommast im Feld, der die Ästhetik der Aufnahme stört, oder eine andere Person im Tiefschneehang, die man nicht mehr auf dem Bild haben möchte, einfach durch einen von der KI erzeugten Hintergrund ersetzt werden. Ähnlich lässt sich auch der Himmel in einem Bild mit einer Funktion auswechseln. Je nach Bildinhalt funktioniert das besser oder schlechter. Im Vergleich zu händischem Ausschneiden (ob in früheren Bildbearbeitungsprogrammen oder noch mit der Rasierklinge) und dem nachfolgenden Ausmalen ist es aber wesentlich schneller und oft auch wesentlich besser.

Wie für einzelne Bilder Inhalte verändert werden können, ist es auch für Filme möglich. Es ist zum Beispiel einfach vorstellbar, dass ein Film, der in den 60er-Jahren spielt, aktuell in einer Landschaft mit Windrädern aufgenommen wird und

diese Windräder dann einfach automatisch wegetuschiert werden. Wird an verschiedenen Tagen gedreht und ist der Himmel und die Beleuchtung unterschiedlich, kann man sicher mittelfristig KI einsetzen, um diese Szenen anzugleichen.

In Videokonferenz-Programmen können wir inzwischen Hintergründe auch ohne Greenscreen austauschen. Dazu werden in Echtzeit die Person und der Hintergrund erkannt. Wenn die Person und auch noch die Lippenbewegung erkannt wird, kann man damit weitere visuelle Effekte erzeugen. Man kann dann z. B. das Gesicht durch einen Avatar ersetzen oder auch nur die Darstellung des Gesichts verändern (z. B. virtuell schminken), siehe Abbildung 4. Aktuell sind die Ergebnisse häufig noch wenig überzeugend, aber es ist absehbar, dass sich das in den nächsten Jahren mit zunehmender Auflösung und Rechenleistung verbessert.

### Bestehende Filme mit neuen Schauspielern besetzen

In den letzten Jahren haben Deep Fakes in verschiedenen Bereichen für Aufsehen gesorgt, von vermeintlichen Pornodarstellungen von Prominenten bis zu Aussagen von Politikern, die es nie in der Form gegeben hat.

Technisch werden dabei durch Verfahren des Maschinellen Lernens (insbesondere GANs und ähnlich wie beim Erzeu-

<sup>15</sup> Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM Computing Surveys (CSUR), 54(1), 1-41. <https://doi.org/10.1145/3425780>

Abbildung 5: Beispiele für Bild-zu-Bild-Transformationen



Quelle: Jun-Yan Zhu\*, Taesung Park\*, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", in IEEE International Conference on Computer Vision (ICCV), 2017. <https://arxiv.org/pdf/1703.10593.pdf>

gen von Bildern) Bereiche in einer Bildfolge durch andere Inhalte ersetzt<sup>15</sup>. So wird z. B. das Gesicht einer anderen Person dem ursprünglichen Gesicht überlagert, und dieses wird dann auch so animiert, dass es authentisch wirkt. Dieser Ansatz funktioniert auch für ganze Körper oder auch für Tiere, wie in Abbildung 5 zu sehen, oder auch für den Stil eines Bildes<sup>16</sup>. Man kann sich nun vorstellen, dass einzelne Personen oder auch alle Handelnden in einem Film mit neuen Schauspielerinnen und Schauspielern „besetzt“ werden, z. B. wenn die ursprünglichen politisch nicht mehr

akzeptabel sind. Man könnte auch Filme mit generischen Schauspielerinnen und Schauspielern produzieren und diese dann mit unterschiedlichen „Gesichtern“ für die jeweiligen Zielgruppen ersetzen.

### 3 Verbesserung der Werkzeuge für die Medienproduktion und Verarbeitung

Es gibt eine große Anzahl an Werkzeugen für die Medienproduktion, welche KI nutzen.

#### Beste Aufnahmen ohne Kenntnisse

Die Fähigkeit, mit einem Smartphone bei Nacht auf einem schaukelnden Boot scharfe Fotos zu machen, ist hierfür ein

Abbildung 6: Aufnahme bei Nacht auf einem Boot. Im „Smart Mode“ wird das Bild mit KI verbessert und erscheint scharf.



Quelle: Venedig bei Nacht (Eigenes Bild des Autors).



eindrückliches Beispiel, siehe Abbildung 6. Hierzu werden algorithmisch viele Einzelbilder mit wenig Licht zu einem Bild mit ausreichender Belichtung zusammengesetzt, und dabei wird die Bewegung kompensiert. Bildoptimierung durch KI ist inzwischen in Telefonen wie in Kameras häufig zu finden. Die Kamera wählt einen sinnvollen Ausschnitt, den richtigen Zeitpunkt und auch die technischen Parameter wie Blende und Belichtung. Damit wird das Erstellen von Bildern und Filmen vereinfacht, gleichzeitig gibt die Person, die eigentlich kreativ arbeitet, Kontrolle ab. Das Foto, das im „Smart-Modus“ aufgenommen wurde, ist dann lediglich eine KI-Interpretation dessen, was der Sensor aufgezeichnet hat. Es zeigt nicht zwangsläufig das, was vor der Kamera war. Das Bild ist fotorealistisch aber bildet eben nur bedingt die Realität ab.

Auch bei der Aufnahme von Tönen, Musik und Gesprächen zeichnen sich ähnliche Trends ab. Hier kann dann auch automatisch die Stimme des Interviewpartners hervorgehoben werden, und die Umgebungsgeräusche können ausgeblendet werden. Mit besseren Werkzeugen werden Basisfähigkeiten abgenommen. Das kann sowohl dazu führen, dass Medien von Menschen mit weniger Qualifikation in ähnlicher Qualität wie von Profis erstellt werden. Es eröffnet auch die Möglichkeit, sich auf andere Dinge in der Schaffung von Medien zu konzentrieren und öffnet Wege für neue Medienformate.

## Auflösung erfinden: KI Upscaling

Es gibt viel digitales Material, das nicht in 4K oder höheren Auflösungen verfügbar ist. Heimvideos oder DVDs haben im Vergleich zu aktuellen Formaten eine sehr geringe Qualität (z. B. PAL DVD mit  $720 \times 576$  bei 25 Bildern pro Sekunde). Eine einfache Vergrößerung, bei welcher die fehlenden Zwischenpixel interpoliert werden, führt zu unscharfen und verwaschenen Bildern. KI-Verfahren erzeugen hier wesentlich bessere Bilder<sup>17</sup>. Die Grundidee ist hierbei, dass mit einem Machine-Learning-Verfahren die Beziehung zwischen geringen Auflösungen und hohen Auflösungen gelernt wird. Für dieses Lernen steht beliebig viel Material zur Verfügung, da man aus hochauflösenden und qualitativ hochwertigen Bildern solche in geringer Größe und Qualität erzeugen kann. Hat man nun sehr viele dieser Paare von Bildern (eines in kleiner Auflösung und geringer Qualität und ein zugehöriges in großer Auflösung und hoher Qualität), kann man ein System trainieren, das bei Eingabe eines Bildes in geringer Auflösung und Qualität ein Bild in hoher Auflösung und Qualität voraussagt. Diese Ansätze werden sowohl in Software-Werkzeugen für die Bild- und Filmbearbeitung als auch in Fernsehgeräten eingesetzt.

## Medien und Realität

Die Beziehung zwischen Realität und der Darstellung von Realität ist ein Spannungsfeld, in welchem sich die Medien schon immer bewegen. In einer modernen Informationsgesellschaft nehmen wir vieles über unsere Umwelt und was wir über die Welt wissen durch digitale Medien wahr. Damit wird die Macht der Medien, Realität zu erschaffen, größer und dringt in viele Lebensbereiche vor. Gleichzeitig werden durch KI die Möglichkeiten zur Manipulation von Inhalten vielfältiger und auch wesentlich günstiger in der Umsetzung.

Die benannten Beispiele sind Schlaglichter und Ausschnitte aus einem sich schnell entwickelnden Feld. Ähnlich wie zu Beginn der 1990er-Jahre erwarten wir eine fundamentale Veränderung der Medienbranche. In noch größerem Maße als in den letzten 30 Jahren werden sich Aufgaben, Tätigkeiten, Formate, Verteilwege und nicht zuletzt Geschäftsmodelle grundlegend verändern. Es scheint, dass in den nächsten Jahren das Feld neu definiert wird. Es gilt mitzubestimmen, wie Medien in der Zukunft aussehen und welche Funktion sie haben werden.

<sup>17</sup> <https://www.geoffreylitt.com/2017/06/04/enhance-upscaling-images-with-generative-adversarial-neural-networks.html>