

„Täuschend echt“

Deepfakes aus theologischer, religionspädagogischer und ethischer Perspektive¹

Dr. Ann-Sophie Markert, ann-sophie.markert@fau.de, Professur für Praktische Theologie (Religions- und Gemeindepädagogik), Friedrich-Alexander-Universität Erlangen-Nürnberg

Dr. Max Tretter, max.tretter@fau.de, Lehrstuhl für Systematische Theologie (Ethik), Friedrich-Alexander-Universität Erlangen-Nürnberg

Christian Gürtler, christian.guertler@fau.de, Lehrstuhl für Medienkommunikation, Medienethik und Digitale Theologie, Friedrich-Alexander-Universität Erlangen-Nürnberg

Deepfakes haben in den letzten Jahren mehrfach mediale Aufmerksamkeit erregt. Bei diesen Deepfakes handelt es sich um KI-generierte Bilder, Tonaufnahmen oder Videosequenzen, häufig von bekannten Persönlichkeiten, die täuschend echt wirken, obwohl sie nicht real sind, sondern entweder von Grund auf künstlich generiert oder durch Manipulationen aus existierenden Inhalten erstellt wurden. Solche Deepfakes können lustig sein, wie bspw. Bilder von Papst Franziskus in einer Moncler-Puffer-Jacket oder Videoclips von Will Smith beim Spaghetti-Essen belegen. Es gibt jedoch auch Deepfakes, deren Inhalt äußerst verletzend für die dargestellten Personen ist (zu diesen zählen etwa die pornographischen Deepfakes von Taylor Swift) oder die enormes politisches Spreng- wie gesellschaftliches Spaltpotenzial besitzen (etwa Bilder einer vermeintlichen Verhaftung Donald Trumps). Nicht zuletzt können derlei Deepfakes auch für kriminelle Zwecke, etwa für raffinierte Varianten des sog. „Enkeltricks“, oder zu Propagandazwecken genutzt werden.

War es vor 10 Jahren noch nahezu unmöglich, hochrealistische Inhalte zu erstellen, lassen sich Deepfakes heute dank immer leistungsfähigerer Hard- und Software immer einfacher herstellen. Teils können sie sogar mit öffentlich zugänglichen KI-Tools wie DALL-E von OpenAI erstellt werden. Zwar haben die Betreibenden großer, öffentlich zugänglicher Generative-KI-Modelle versucht, dem einen Riegel vorzuschieben, indem sie inzwischen rigorose Vorsichtsmaßnahmen implementiert haben; so weigert sich DALL-E zum Beispiel, Bilder von oder auf der Grundlage von realen Personen zu erstellen. Doch diese Einschränkungen lassen sich mit etwas Kreativität und Fachkenntnis umgehen, sodass, wer wirklich Interesse daran hat, Deepfakes herzustellen, sich durch diese Barrieren nicht wird hindern lassen.

Diese Einfachheit, mit der sich Deepfakes erstellen lassen, lässt erwarten bzw. befürchten, dass uns derlei gefälschter und manipulierter Content in Zukunft immer häufiger begegnen wird. Dies ist besonders bedenklich, da sich Deepfakes häufig nicht als solche identifizieren lassen und es häufig gar keine Unterschiede mehr zu „realen“ Fotos, Aufzeichnungen oder Videomitschnitten gibt.

Zwar wurden vereinzelt bereits Gesetze zur Kennzeichnungspflicht von KI-generierten Inhalten erlassen und es gibt weitreichende Bestrebungen, diese Pflicht auszuweiten. Doch solange diese

¹ Lightning-Talk an der Tagung „Theologische Forschung, Lehre und Infrastruktur im Kontext von Digitalisierung und KI“ am 12./13.02.2026 in Tübingen.

Kennzeichnungsgesetze nicht rigoros eingefordert und forciert werden – und es stellt sich die Frage, wie eine solche Umsetzung real stattfinden soll –, stellen uns Deepfakes vor beträchtliche Herausforderungen und werfen grundlegende Fragen auf: Was ist echt, was nicht? Welchen Institutionen bzw. welchen Inhalten können wir – im Zeitalter von Fake News, alternativen Fakten und nun auch Deepfakes – noch vertrauen? Welche Auswirkungen werden Deepfakes auf unsere Gesellschaft und Politik haben?

Vor diesem Hintergrund haben wir im Sommersemester 2025 eine vom Innovationsfond Lehre der Friedrich-Alexander-Universität Erlangen-Nürnberg finanziell geförderte, interdisziplinäre Lehrveranstaltung durchgeführt, die sich explizit mit dem Thema *Deepfakes* befasst. Ziel war es, Studierenden ein fundiertes theoretisches Verständnis von Deepfakes zu vermitteln und sie zugleich praktisch zu einem souveränen, reflektierten und selbstbestimmten Umgang mit dieser Technologie zu befähigen – sowohl im Hinblick auf ihr eigenes Handeln als auch auf ihre zukünftige Rolle als Multiplikator:innen, etwa im Pfarramt oder im Schuldienst. Auf diese Weise wurde Digitalisierung nicht nur thematisch, sondern methodisch und kritisch-reflexiv in die theologische Lehre integriert.

Didaktisch folgte das Projekt einem konsekutiven Drei-Phasen-Modell:

- [1] In einer ersten Phase der *technischen Dekonstruktion* setzten sich die Studierenden *hands-on* mit den Funktionsweisen von Deepfakes auseinander. Sie lernten, wie KI-Modelle trainiert werden, welche Rolle Trainingsdaten spielen, und versuchten in praktischen Übungen mit KI-Bild- und Videogeneratoren sowie spezifischer Software selbst, Deepfakes zu erzeugen, um die Mechanismen der Täuschung „von innen heraus“ zu verstehen.
- [2] In einer zweiten Phase stand die *kritische Reflexion* im Mittelpunkt. Flankiert von Impulsen externer Expert:innen wurden ethische, philosophische und pädagogische Perspektiven auf Deepfakes erarbeitet. Leitend war dabei die Frage, wie sich Deepfakes auf Wahrheitsverständnisse, Vertrauen und das soziale Miteinander auswirken.
- [3] Die dritte Phase war der *Wissenschaftskommunikation* gewidmet: Die Studierenden entwarfen aus ihren Erkenntnissen eine Plakatausstellung, die in Kollaboration mit <https://www.ausstellung-leihen.de> veröffentlicht wurde. Indem die Studierenden die Lerninhalte des Semesters elementarisierten und sie – unter professioneller Anleitung – zu einer Ausstellung ausarbeiteten, wurden sie zu Vermittler:innen und steigerten und vertieften dadurch ihre bisher erlangten Kompetenzen.

Jenseits einer Kompetenzorientierung und der Neuverfügung von Wissensbereichen betonte unsere Lehrveranstaltung die Bildungsperspektive eigens: Im Rückgriff auf die digitale Mündigkeit als aufklärerisches Ziel und den Bildungsbegriff Wilhelm von Humboldts als wechselseitige Durchdringung von Ich und Welt, treten kritisches Denken und Persönlichkeitsbildung in den Vordergrund. Bildung in der digitalen Welt erschöpft sich nicht in erlerntem Wissen und neuen quantitativ messbaren Kompetenzen, sondern von Bedeutung ist eine reflexive, qualitative Veränderung der Welt- und Selbstbilder. So wurden den Studierenden wichtige Urteils- und Handlungsfähigkeiten zur Orientierung in einer digitalen Welt vermittelt. Darüber hinaus wurden sie befähigt, als Multiplikator:innen auch anderen Personen zentrale Kompetenzen zu vermitteln, die sie in digitalen Kontexten benötigen. Zudem setzten sich Studierende während des gesamten Lernweges eigenverantwortlich und selbstständig mit Deepfakes auseinander. Dadurch wurden sie nicht nur partizipativ und aktiv

in den Lehr- und Lernprozess eingebunden, sondern auch in hohem Maße zu kritischen Denk- und Reflexionsprozessen angeregt.