

LAB REPORT

KÜNSTLICHE INTELLIGENZ

LAB REPORT: KÜNSTLICHE INTELLIGENZ

Dieser Report des Media Lab Bayern enthält Beispiele, die zeigen, wie in München schon seit Jahren mit Technologien gearbeitet wird, die heute unter dem Begriff „Künstliche Intelligenz“ zusammengefasst werden.

In den meisten Fällen zeigen die Beispiele Produkte, die am Markt verfügbar sind und deren Gründer:innen jederzeit bei Fragen und Ideen ansprechbar sind. Die Profile in diesem Report zeigen auch, wie diese Gründer:innen KI-Technologien eingesetzt und angepasst haben, welche Hürden es dabei zu überwinden gab und was im Bereich KI vielleicht einfacher ist, als man es sich denken könnte.

Lerne in diesem Report:

- Wie junge Medienfirmen Probleme mit KI lösen
- Welche KI-Tools dein Unternehmen voranbringen
- Wie du selbst besser mit KI experimentieren kannst
- Wie das Media Lab Bayern dich dabei unterstützen kann

DAS MEDIA LAB BAYERN

Das Media Lab Bayern ist der Innovation Hub für digitale Medien in München. Das Lab unterstützt Gründer:innen, Startups und Medienhäuser aus ganz Deutschland dabei, innovative Ideen zu finden und umzusetzen.

Wir sind ein Team aus Innovationsexpert:innen, Medienmacher:innen und Eventmanager:innen und arbeiten jeden Tag daran, mehr Innovation in die Medienbranche zu bringen. Wir lieben es, Talente zu unterstützen und zusammenzubringen, starten neue Projekte lieber heute als morgen und finden zu (fast) jedem Problem eine Lösung.

→ media-lab.de

LAB REPORT: KÜNSTLICHE INTELLIGENZ	02
EXECUTIVE SUMMARY	03
VORWORT VON CHARLIE BECKETT	04
GLOSSAR	05

DIE WELT EIN STÜCKCHEN BESSER MACHEN MIT KI	06
SUMM AI	07
ROKIN	09
NESOMA HEALTH	10
KONSOLE LABS / DAILY CORONA NEWS	11
FOCAL ANALYTICS / DIVERSAL ANALYTICS	12

WENN DIE KI DIE INHALTE DOCH SELBST ÜBERNIMMT	13
ONEGUIDE	14
TELEVISION.AI	16
HYPERATE	17
SNIP AID	18
KOMPRENO	19
ANYMATE ME	20

BESSER ARBEITEN MIT, STATT GEGEN KI	21
OPEN RECOMMENDATION BUTLER & SPEECHCATCHER	22
VARIA RESEARCH	24
AUDIOINDEX	25
FUSIONBASE	26
LUMEDIO	27
FRÄNKISCHER TAG	28

IMPRESSUM

TEXT

Bernd Oswald,
Christian Simon &
Johannes Klingebiel

LAYOUT & DESIGN

Johannes Klingebiel

SCHLUSSREDAKTION

Jessica Weber, Alex Rühl,
Christian Simon

BILDER

Cindy Ngo (wenn nicht anders angegeben)

MIT DANK AN

Unsere Fellows, unsere Coaches und das Media Lab Bayern Team für die tatkräftige Unterstützung

V.I.S.D.P.

Lina Timm

Medien.Bayern GmbH
August-Everding-Straße 25
81671 München

© MEDIA LAB BAYERN 2023

Das Media Lab Bayern ist Teil der Medien.Bayern GmbH, einer Tochterfirma der BLM und wird gefördert durch die Bayerische Staatskanzlei.

**MEDIEN.
BAYERN**

BLM
Bayerische Landeszentrale
für neue Medien

Bayerische Staatskanzlei



WAS MAN AUS DIESEM REPORT LERNEN KANN

1

NIEMAND MUSS DAS RAD NEU ERFINDEN

Nur wenige der erfolgreichen KI-Beispiele aus dem Media Lab haben neue Modelle von Grund auf selbst gebaut. Denn dafür braucht es Unmengen an Trainingsdaten und Geld für Rechenleistung und Speicherplatz—und das unabhängig davon, ob man die Rechner selbst baut oder sich Rechenleistung von einem der großen Cloud-Anbieter mietet.

Stattdessen gibt es bereits zahlreiche Modelle, die meist über einfache APIs nutzbar sind. Large Language Models wie Falcon, Bird oder spaCy, Tools wie AWS Recognition oder Microsofts Azure Video Indexer und natürlich GPT4—für alle Medienformen und viele Bedürfnisse gibt es fertige Modelle, die man sich für die eigenen Anwendungen zunutze machen kann. Ohne größere Investments.

2

DATEN, DATEN, DATEN!

Egal, ob man Content mittels einer KI erstellen möchte oder „nur“ ein neues Tool bauen will: Strukturierte Daten sind die Grundlage für jede KI-Anwendung. Dafür braucht es Einheitlichkeit und vernünftige Metadaten. Oder man will ein bestehendes Modell mit eigenen Daten weiter trainieren; auch dann muss es diese Trainingsdaten erst einmal geben (oder sie müssen auffindbar sein).

Medienhäuser, die in den vergangenen Jahren saubere Archive geführt oder einfache Schnittstellen in ihrem CMS eingerichtet haben, haben hier die Nase vorn.

3

USE CASES SIND WICHTIGER ALS TECHNOLOGIEN

Wenn die eigentliche KI von Amazon, Google oder OpenAI kommt—was bleibt da überhaupt noch übrig für Medienfirmen? Die Antwort: Spezialisierung! Bis auf Weiteres können die bestehenden Modelle sehr viel und vieles ziemlich gut. Sie sind aber oft unspezifisch. Mit Anpassungen, eigenen Filterregeln oder Extra-Trainings können die Modelle einzelne Jobs noch viel besser erledigen. Und zwar auf Arten, die dann einzigartig für die jeweiligen „Trainierenden“ sind.

Gleichzeitig werden die Use Cases enorm wichtig. Welches Problem soll gelöst werden und warum? Wenn es eine Technologie gibt, die genau das kann: Wie verpackt man sie so, dass Menschen damit umgehen können? Wie gewinnt man das Vertrauen der Nutzer:innen? Hier liegt die große Chance für Medienfirmen, KI erfolgreich einzusetzen—auch, wenn sie keine eigene bauen wollen oder können.

4

KI IST SO VIEL MEHR ALS CHATGPT!

Die Diskussion um KI im Medienbereich dreht sich zu oft um ChatGPT und andere KIs, die synthetische Inhalte generieren, egal ob Text, Bild oder Audio. Außerdem steht oft die Frage im Raum, wie man Fakten in KI-Texten checken kann—oder sogar, ob KI im schlimmsten Fall zu Jobverlusten führt.

Alles wichtige Diskussionen, aber dieser Report zeigt: Die Welt der KI ist vielfältiger. Für jede automatisierte Content-Erstellung gibt es mindestens ein Tool, das Journalist:innen nervige Arbeitsschritte erleichtert, ihre Infrastruktur verbessert oder neue, bessere oder einfachere Recherchen ermöglicht. Und das alles auch noch, nachdem der Hype um ChatGPT wieder abgeflaut ist. Wer langfristig denkt, gewinnt.

„Ich freue mich über den *HYPE* um ChatGTP, DALL-E, Midjourney und all die anderen KI-Werkzeuge, die im letzten Jahr in unserer digitalen Welt aufgetaucht sind.“



Alle, die wollen, können dort eine einfache Eingabeaufforderung eintippen und sofort eine sinnvolle (oder unsinnige) Antwort erhalten. Seit fünf Jahren rühre ich die Werbetrommel dafür, die Bedeutung von KI in den Medien (und in fast allen anderen Bereichen) ernst zu nehmen—und jetzt hört man mir endlich zu. Vom Praktikanten bis zum CEO wissen wir jetzt alle mindestens ein bisschen etwas über KI.

Genau wie ich wirbt das Media Lab Bayern seit vielen Jahren für die Vorzüge digitaler Technologien. Nicht nur, um Startups aufzubauen und sie zu unterstützen, neue Produkte zu entwickeln, sondern auch, um die Welt zu verbessern. Die Technologien, die durch Algorithmen und maschinelles Lernen angetrieben werden und die wir „Künstliche Intelligenz“ nennen, sind die jüngste Welle technologischen Wandels. Zuerst sind wir mit dem Internet online gegangen, dann sind wir alle in sozialen Netzwerken aktiv geworden. Beides hat jeden Aspekt der Medieninnovation, -produktion und -nutzung verändert.

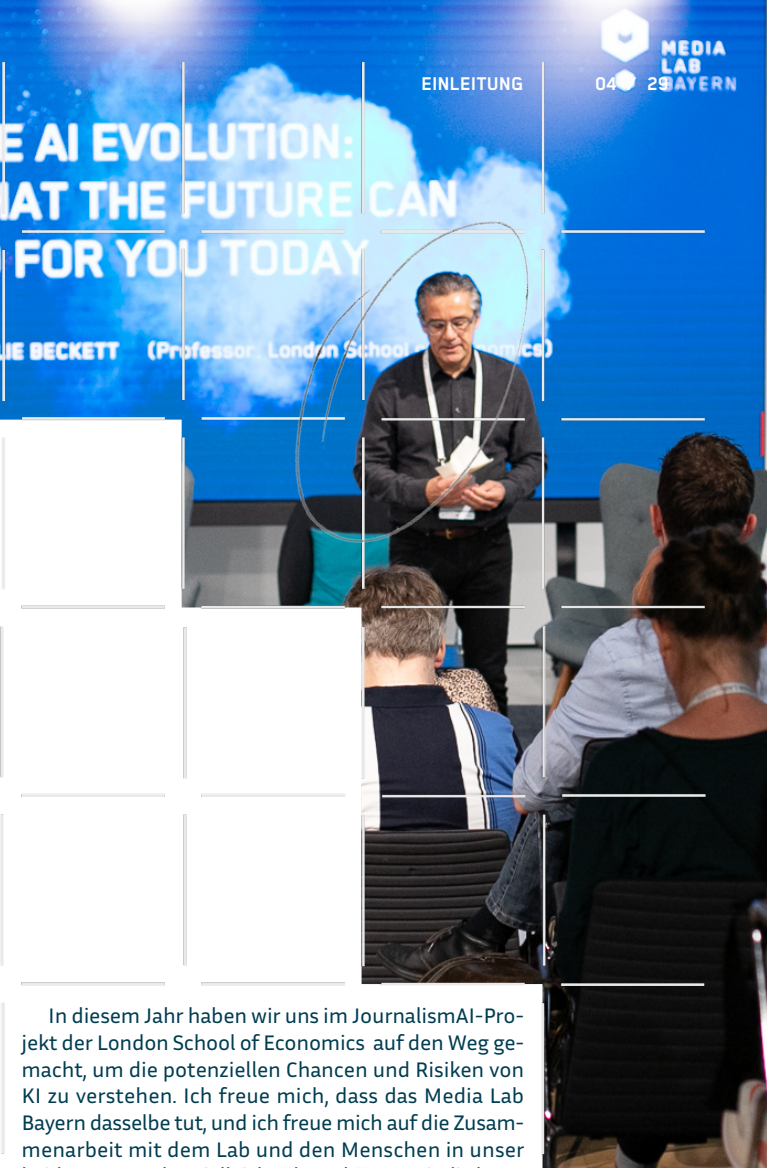
Spätestens ab dem Jahr 2018 konnte ich feststellen, dass KI begann, vor allem den Journalismus stark zu beeinflussen. Unsere globale Journalismus-Umfrage hat uns gezeigt, dass KI großes Potenzial bietet, Journalist:innen zu unterstützen, Redaktionen effizienter zu machen und eine bessere Beziehung zu den Nutzer:innen herzustellen. Wir haben aufgezeigt, wie Automatisierung, Personalisierung und die Arbeit mit vielen Daten die Nachrichtenrecherche, die Produktion von Inhalten, die Nutzererfahrung und die Umsatzsteigerung fördern.

GENAU WIE ICH WIRBT DAS MEDIA LAB BAYERN SEIT VIELEN JAHREN FÜR DIE VORZÜGE DIGITALER TECHNOLOGIEN. NICHT NUR, UM STARTUPS AUFZUBAUEN UND SIE ZU UNTERSTÜTZEN, NEUE PRODUKTE ZU ENTWICKELN, SONDERN AUCH, UM DIE WELT ZU VERBESSERN.

Wir haben Nachrichtenorganisationen geholfen, ihr Wissen zu erweitern, neue Fähigkeiten zu lernen und ihre Arbeitsabläufe so zu organisieren, dass sie sich an die Vorteile der KI anpassen können—statt in die Fallstricke zu laufen. Aber KI schien damals immer noch etwas für die "Techies" zu sein, nicht für alle anderen in einer Nachrichtenorganisation.

Mit generativer KI hat sich das nun geändert. Sie bietet mehr kreative Möglichkeiten, birgt aber auch größere Gefahren, sowohl für die Gesellschaft im Allgemeinen als auch für Medienunternehmen im Besonderen. Alle sollten mit generativer KI experimentieren und herausfinden, wie und wo sie seine oder ihre Arbeit unterstützen kann—und wie wir uns vor dem kommenden Sturm generativer Desinformation und gefälschter Produkte schützen können. Dabei gilt dieselbe Strategie, die man auch für "normale" KI empfehlen kann:

Stellt sicher, dass ihr wisst, womit genau ihr da arbeitet. Stellt sicher, dass ihr Mitarbeiter:innen habt, die ein Verständnis für die Entwicklung der generativen KI haben und dafür verantwortlich sind, eure Organisation darauf vorzubereiten. Sorgt dafür, dass alle in eurem Unternehmen an dieser Diskussion beteiligt sind. Fangt an, Richtlinien auszuarbeiten, die euren Kolleg:innen die Arbeit mit KI erleichtern, ohne euer Geschäftsmodell und Auftrag zu gefährden. Vor allem: Zeigt ein bisschen Fantasie! Und überprüft eure Arbeit und stellt sicher, dass der Mensch das letzte Wort hat und nicht der Algorithmus.



In diesem Jahr haben wir uns im JournalismAI-Projekt der London School of Economics auf den Weg gemacht, um die potenziellen Chancen und Risiken von KI zu verstehen. Ich freue mich, dass das Media Lab Bayern dasselbe tut, und ich freue mich auf die Zusammenarbeit mit dem Lab und den Menschen in unser beider Netzwerke. Vielleicht überschätzen wir die kurzfristigen Auswirkungen der generativen KI—aber ich bin überzeugt, dass sie langfristig fundamentale Auswirkungen haben wird.

Professor Charlie Beckett,
Director LSE JournalismAI project

BEGRIFFE & ABKÜRZUNGEN

API (Application Programming Interface)

Eine Schnittstelle, die das Zusammenspiel verschiedener Computerprogramme erleichtert.

AWS (Amazon Web Services)

Einer der größten Anbieter für Cloud-Computing Services. Stellt gegen Gebühr über das Internet Speicherplatz, Rechenleistung und andere Angebote zur Verfügung, die Nutzende nicht selbst bereitstellen wollen oder können.

CMS (Content Management System)

Eine Software zur gemeinschaftlichen Erstellung, Bearbeitung, Organisation und Darstellung digitaler Inhalte; zumeist zur Verwendung in Webseiten.

Embedding

Die Umwandlung von Text oder anderen Inhalten in maschinenlesbare Zahlenreihen, die dann durch statistische Verfahren (→ **Machine Learning**) miteinander verglichen werden können. Grundlage für viele KI-Anwendungen.

Framework

Ein Software-Tool, das verschiedene andere Programme (z. B. → **Sprachmodelle**) bündelt und die Zusammenarbeit zwischen ihnen erleichtert.

LLM (Large Language Model / Sprachmodell)

Machine Learning Modelle, die explizit für die Arbeit mit Sprache entwickelt wurden und so z. B. Freitext als Eingabe akzeptieren und als Antwort ausgeben kann. Benötigt große Mengen an → **Trainingsdaten**.

Machine Learning

Statistisches Verfahren, das es einem leistungsstarken Computer ermöglicht, selbstständig Muster und Wahrscheinlichkeiten in Daten (→ **Embedding**) zu erkennen (zu „erlernen“) und diese Muster auf unbekannte Daten anzuwenden. Grundlage jeder modernen Form der KI. Benötigt große Mengen an strukturierten → **Trainingsdaten**.

Metadaten

Informationen, die Daten beschreiben, sie strukturieren und so auffindbar und durchsuchbar machen. Beispiele für Metadaten sind Schlagworte, Dateigrößen oder Autor:innen.

Microservices

Microservices sind eine Art, komplexe Anwendungssoftware aus Prozessen zu bauen, die untereinander mit Programmierschnittstellen kommunizieren. Die Dienste fungieren wie Bausteine, die unabhängig voneinander Aufgaben erfüllen können.

Natural Language Processing / Generation

Beschreibt die Funktion eines → **Sprachmodells**, natürlich klingende Sprache/Texte als Eingabe zu akzeptieren, bzw. als Resultat auszugeben.

Named Entity Recognition

Named Entity Recognition oder Eigennamenerkennung bezeichnet die automatische Identifikation und Klassifikation von Eigennamen, wie z. B. Firmennamen.

Open Innovation Challenge

Ein unregelmäßiges Programm des Media Lab Bayern, bei dem innovative Lösungen zu aktuellen Branchenthemen gesucht und prämiert werden.

Open Source Software

Software, die unter einer Lizenz veröffentlicht wird, in der die Urheberrechtsinhaber:in den Benutzer:innen das Recht einräumt, den Code für jeden Zweck zu nutzen, zu ändern und zu verteilen.

Prompt

In natürlicher Sprache verfasste Handlungsaufforderung an ein Machine Learning Model (wie bspw. → **Large Language Model**). Populär gemacht durch das Chatinterface von ChatGPT.

Trainingsdaten (Data)

Bildet die Grundlage für eine KI. Beschreibt die Menge, den Korpus, an „Beispielen“, aus denen die KI mit Hilfe von → **Machine Learning Muster** und Wahrscheinlichkeiten „lernt“. Trainingsdaten müssen dafür sorgfältig ausgewählt und aufbereitet werden.

DIE FÖRDERPROGRAMME DES MEDIA LAB BAYERN



Media Company Fellowship (MCF)

Programm des Media Lab Bayern für Medienhäuser. In diesem kombinierten Förderprogramm werden Führungskräfte im Leadership Program an der City University of New York fortgebildet, während ein Team aus dem Unternehmen im Media Lab an einem innovativen Projekt arbeitet.



Media Startup Fellowship (MSF)

Acceleration-Programm des Media Lab Bayern für Medien Startups. Teilnehmende Teams werden neun Monate lang mit Geld und Coaching gefördert und beim Einstieg in den Markt begleitet. Mit über 100 geförderten Teams in 13 Batches ist es unser erfolgreichstes Förderprogramm.



Media Tech Lab (MTL)

Von künstlicher Intelligenz, über Augmented Reality, bis hin zur Web-Infrastruktur—das Media Tech Lab fördert Open-Source-Projekte für die Medienwelt.

DIE WELT EIN STÜCKCHEN BESSER MACHEN MIT KI

Die Sorge vor Jobverlust, Disruption und Skynet ist das eine—die zahlreichen Möglichkeiten, wie KI einen positiven sozialen Impact haben kann, das andere. Im Media Lab arbeiten zahlreiche Teams und Talente daran, mit KI die Welt ein bisschen besser zu machen.

Klischees und schnelle Urteile sind selten hilfreich, wenn man über die Zukunft nachdenkt. „KI wird den Arbeitsmarkt disruptieren und hunderttausende Menschen in die Arbeitslosigkeit treiben“, ist ein solches Klischee. „Startups geht es nur um den schnellen Profit“, ein anderes. Ja, KI wird Jobs in zahlreichen Branchen beeinflussen, verändern und teilweise überflüssig machen. Und manchen Startups geht es wirklich nur um den schnellen Exit.

Und trotzdem hat KI das Potenzial, positiven, gesellschaftlichen Impact zu haben. Und es gibt Startups, die genau daran arbeiten. SUMM aus dem Media Startup Fellowship Batch #11, die Machine Learning nutzen, um Inhalte im Internet barrierefreier zu machen, sind dafür ein Vorbild—auf den nächsten Seiten stellen wir das Team und ihr Produkt vor.

Doch die Umwandlung von komplizierten Texten in leicht zugängliche Sprache ist nur einer von vielen Anwendungsfällen, in denen KI, richtig eingesetzt, die Welt ein Stück fairer und besser macht—schon heute. Weitere Anwendungsfelder finden sich etwa im großen Bereich „Gesundheit“. Nesoma Health aus Batch #11 des Startup Fellowships arbeitet mit einem digitalen Therapeuten, um ihren Nutzer:innen zu helfen, ein besseres Sexleben zu haben—seriös, auf wissenschaftlicher Grundlage basierend und unterstützt durch KI.

Anwendungen wie diese sind keine Einzelfälle, und auch keine neue Entwicklung. Bereits im Batch #7 des Media Startup Fellowship, also 2020, nutzte das Team Rokin etwa Machine Learning, um Ingenieur:innen

bei der Suche nach geeigneten Technologien für ihre spezifischen Herausforderungen zu unterstützen—in diesem Fall also nicht, um Jobs „wegzurationalisieren“, sondern den Arbeitsalltag vieler Menschen in einer von Deutschlands wichtigsten Branchen zu verbessern.

Zum Abschluss noch zwei Beispiele außerhalb der Startup-Szene. In einer Open Innovation Challenge fragte das Media Lab 2020: „Wie kann Technologie helfen, für mehr Diversität in der Medienbranche zu sorgen?“ Eine der eingereichten Lösungen, „Diversal Analytics“, scannt mit Hilfe von KI automatisch Texte und vergibt einen Score: Wie viele People of Color werden im Text erwähnt? Nur junge Menschen? Oder auch Alte? Nur Männer? Oder auch Frauen? Der Impact hier ist klar: Nur wenn Journalist:innen wissen, wie divers das Ergebnis ihrer Arbeit ist, können sie darüber nachdenken, ob und wie sie die Diversität steigern wollen.

Und in der Corona-Krise half Daily Corona News dabei, Menschen mit den aktuellsten Zahlen und Fakten aus ihrem direkten Umfeld zu versorgen—genau dann, als solche Informationen für sie am wichtigsten waren.

Diese Auswahl zeigt: KI bedeutet keinen Weltuntergang; sie kann auch eine bessere Welt bedeuten. Die folgenden Seiten werden zeigen: Impact muss beim Thema KI kein Hobbythema sein. Und wir müssen auch nicht darauf warten, dass KI etwas Schlimmes anrichtet, bevor wir uns mit den guten Seiten beschäftigen: Die positiven Anwendungsfälle gibt es schon heute.



LEICHT VERSTÄNDLICHE TEXTE AUF KNOPFDRUCK

SUMM AI HAT EIN TOOL
ENTWICKELT, DAS
KOMPLIZIERTE TEXTE
SOFORT IN LEICHTE
SPRACHE ÜBERSETZT.

DIESE ANWENDUNG ZEIGT,
WIE KI EINEN GREIFBAREN
GESELLSCHAFTLICHEN
NUTZEN HABEN KANN.

Gründerin Flora
Geske auf dem Media
Innovation Festival
2023 des Media Lab
Bayern

Behördendeutsch zu verstehen ist eine echte Herausforderung—selbst für Leute mit akademischem Hintergrund. Aber wie ist das erst, wenn man den nicht hat und vielleicht noch durch eine Behinderung benachteiligt ist?

Personen mit körperlicher oder geistiger Behinderung, Migrationshintergrund oder einer Lernschwäche haben bei komplizierten Texten besonders große Verständnisschwierigkeiten. Für sie gibt es die Leichte Sprache, ein Regelwerk für eine leichter zu lesende Sprache: Kurze Sätze, einfache Wörter, Bilder.

Doch es ist aufwändig, Texte in Leichter Sprache zu erstellen. Das machen meistens speziell geschulte Übersetzer:innen. Und weil es davon nicht viele gibt, gibt es auch nur wenige Texte in Leichter Sprache. Dabei gehen Studien davon aus, dass bis zu zehn Millionen Menschen in Deutschland Bedarf an Texten in Leichter Sprache haben.

SUMM IST EIN KLASSISCHES BEISPIEL FÜR NATURAL LANGUAGE PROCESSING

Das brachte Flora Geske auf die Idee, ein Tool zu erfinden, das automatisch Texte von komplizierter in Leichte Sprache übersetzt. Schon im Studium des Finanz- und Informationsmanagements an der TU München hatte sich Geske mit Natural Language Processing, kurz NLP, befasst: Einem Teilgebiet der Künstlichen Intelligenz, das sich mit der automatischen Verarbeitung von natürlicher Sprache beschäftigt.

Zusammen mit ihren Studienkolleg:innen Vanessa Theel und Nicholas Wolf gründete sie im April 2022 Summ AI. Kernprodukt ist ein Tool, in das man einen komplizierten Text hinein kopiert und auf Knopfdruck eine Version in Leichter Sprache erhält.

NAME	SUMM AI
GRÜNDUNG	2022
FÖRDERUNG	MEDIA COMPANY FELLOWSHIP #11
WEBSEITE	SUMM-AI.COM
GRÜNDER:INNEN	FLORA GESKE, VANESSA THEEL, NICHOLAS WOLF

Summ besteht aus vier Modulen:

- dem Pre-Processing der Input-Texte
- einem Large Language Model für allgemeines Textverständnis und Textproduktion
- einem Leichte-Sprache-Layer für das Einhalten der Leichte-Sprache-Regeln sowie
- dem Post-Processing der Output-Texte.

Summ verwendet die Large Language Models von Anbietern wie Bloom, Aleph Alpha oder Open AI. Die anderen drei Module hat Summ selbst entwickelt. Wie viele andere KI-Startups praktiziert auch Summ Feintuning an bestehenden Sprachmodellen: „Das Feintuning funktioniert nach dem Vorbild eines lernenden Menschen. Wir verwenden ein LLM, das allgemein Texte versteht und Texte produzieren kann. Dabei vollziehen wir mit dem Tool eine Leichte Sprache-Ausbildung, indem wir Beispiele von komplizierten Texten und ihren Übersetzungen einpflegen.“, erklärt Flora Geske.

Diese Methode wendet SUMM vor allem bei sozialen Hilfsangeboten einer Stadt oder weiteren Bereichen an, in denen Leichte Sprache benötigt wird. Dabei wiederholt SUMM den Prozess so lange, bis das Ergebnis robust ist und konstant gute Ergebnisse liefert.

„Wir bekommen oft die Rückmeldung, dass unser Tool hilft, Barrierefreiheit zu gewährleisten und so ein riesengroßes Problem löst.“

AUTOMATISIERTE ÜBERSETZUNG GEHT VIEL SCHNELLER

Die Kundschaft brauchen dann ihre Texte nur noch ins Eingabefeld kopieren und auf „Übersetzen“ klicken und erhalten binnen Sekunden eine Version in Leichter Sprache. Aus dem Satz „Die Erteilung der Gewerbeauskunft erfolgt nach Entrichtung der Gebühren“ macht das SUMM-Tool „Sie wollen die Gewerbeauskunft? Dann müssen Sie Geld bezahlen.“

Auf Leichte Sprache spezialisierte Übersetzungsdienste würden dafür ein Vielfaches an Zeit brauchen und je länger der Text, desto größer die Zeitersparnis. Hier gibt es also einen ganz klaren Skalierungsfaktor.

Abgerechnet wird über ein Lizenzmodell: SUMM bietet seinen Kundschaft Pakete mit unterschiedlichem Zeichenvolumen pro Monat: „Wenn man die Lizenz auf die genutzten Seiten umrechnen würde, ist SUMM um mindestens 90 Prozent günstiger als ein Übersetzer“, sagt Flora Geske. Zu den Kundschaft zählen derzeit vor allem Behörden: Aschaffenburg übersetzt seine Pressemitteilungen mit dem Web-Interface von SUMM, Hamburg hat sogar die SUMM-Programmierschnittstelle in sein Content Management System eingebunden und bietet inzwischen eine ganze Reihe von Texten auf seiner Website in Leichter Sprache an.

Dafür haben die Stadt Hamburg und SUMM gerade eine Auszeichnung beim Ideenwettbewerb „Gemeinsam wird es KI“ des Bundesarbeitsministeriums bekommen. Nur einer von mehreren Preisen, mit denen SUMM prämiert wurde.

SUMM-TOOL HILFT DEN KUNDSCHAFT, IHRE WEBSITES BARRIEREFREI ZU GESTALTEN

Auch das Feedback der Kundschaft ist sehr gut, berichtet Flora Geske: „Wir bekommen oft die Rückmeldung, dass unser Tool hilft, Barrierefreiheit zu gewährleisten und so ein riesengroßes Problem löst“. Das Münchner Startup ist auch in Gesprächen mit Medienhäusern, hier ist es aber etwas schwieriger als bei Behörden. „Ich habe das Gefühl, dass Journalistinnen und Journalisten mehr an ihren Texten hängen“, sagt Geske.

Bislang ist alles sehr gut gelaufen, erzählt Geske: Seit der Gründung in 2022 hat das Startup über 40 Kundschaft gewonnen. In der Gründungsphase hat Summ stark vom Startup-Fellowship des Media Lab Bayern profitiert: „Am wichtigsten war und ist für uns der Einblick und die Vernetzung in den Medienbereich. Von außen ist es sehr schwer, die Player und Dynamiken zu verstehen und zu evaluieren, ob man ein sinnvolles Angebot für die Branche macht. Dabei hat uns das Fellowship weitergeholfen“, freut sich Geske. Und natürlich sei auch die finanzielle Unterstützung in der Startup-Anfangsphase wertvoll gewesen. Beim weiteren Wachstum hilft das Grow Fellowship des Media Labs. Das größte Problem sei es gerade, Marketing-Expert:innen zu finden, denn hier gebe es mit den großen Playern wie Google oder Microsoft starke Konkurrenz.

Für die Zukunft haben Geske, Theel und Wolf schon einige Expansionsideen: So wollen sie auch Privatunternehmen als Kundschaft gewinnen. Dabei kommt SUMM das 2025 in Kraft tretende Barrierefreiheitsstärkungsgesetz zugute, das auch Unternehmen dazu verpflichtet, Texte in Leichter Sprache anzubieten. Geske denkt hier an Shopping-Plattformen oder Versicherungen. Und schließlich ist Leichte Sprache ja kein deutsches Phänomen, sondern ist in anderen Sprachen genauso nötig und nützlich. Englisch und Französisch stehen hier auf der Roadmap ganz oben.

ORIGINALTEXT

Wer einen positiven Schnelltest hat, muss sich isolieren und einen PCR-Test durchführen lassen.

Wenn sich die Infektion nach einem Laborbefund bestätigt, muss selbstständig–auch ohne weitere Anordnung durch das Gesundheitsamt–die Isolation zu Hause angetreten werden.

Infizierte Personen müssen diese Isolation regelhaft für fünf Tage einhalten.

LEICHTE SPRACHE

Sie haben einen positiven Schnell-Test? Dann dürfen Sie niemanden treffen. Und Sie müssen einen PCR-Test machen.

Der PCR-Test ist auch positiv? Dann müssen Sie 5 Tage zu Hause bleiben. Auch ohne Vorschrift vom Gesundheits-Amt.



DIE SUCHMASCHINE FÜR INGENIEURE

NAME	ROKIN
GRÜNDUNG	2020
FÖRDERUNG	MEDIA LAB FELLOWSHIP #7
WEBSEITE	ROKIN.TECH
GRÜNDER	JOHANNES KINKELDEI & ROBERTO DI GIROLAMO

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Rokin ist eine Startup-Ausgründung des VDMA Bayern aus dem Projekt „Industrie 4.0 – Produktion und Geschäftsmodelle von morgen“ des bayerischen Wirtschaftsministeriums. **Wir haben den digitalen Ingenieur entwickelt, eine Art ChatGPT für die Ingenieurbranche, der Ingenieur:innen hilft, eine Lösung für ihre technische Problemstellung zu finden.**

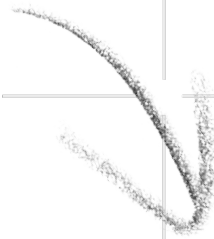
WAS MACHT EUER PRODUKT EINZIGARTIG?

Unser Fokus auf das Ingenieurwesen. Wir haben ein eigenes Sprachmodell trainiert, das speziell für Ingenieure bzw. technische Themen funktioniert.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Alles dreht sich um den Suchbegriff. Entweder allgemein, also etwa „Augmented Reality Technologien für die Produktion“ oder auch detaillierter: „Verfahren, um Oberflächen von bestimmten Stahllegierungen korrosionsresistent zu machen.“ Daraufhin durchsuchen unsere KI-Algorithmen eine Menge verschiedener Quellen, z. B. wissenschaftliche Publikationen, Datenbanken von Fachpublikationen oder Technologie-Marktplätze, identifizieren die relevantesten Infos und spielen sie als Treffer aus.

Wir haben den digitalen Ingenieur entwickelt, eine Art ChatGPT für die Ingenieurbranche, der Ingenieur:innen hilft, eine Lösung für ihre technische Problemstellung zu finden.



WIE HABT IHR DAS TECHNISCH UMGESETZT?

Wir haben viele Daten gesammelt, darunter Fachzeitschriften, wissenschaftliche Publikationen und Google-Suchergebnisse. Diese Daten haben wir verwendet, um verschiedene Sprachmodelle zu trainieren. Wir haben Modelle für einzelne Sprachen oder auch multilinguale Modelle erstellt aus der GPT-Familie, aus der BERT-Familie und mehr.

Wir arbeiten mit verschiedenen Modellen. Eines ist für die Entity Recognition trainiert, erkennt also, um welche Firmen und Produkte es in einem Text geht und welche Eigenschaften diese Produkte haben. Ein anderes Modell extrahiert aus wissenschaftlichen Publikationen Informationen zu bestimmten Themen. Alle unsere KI-Modelle arbeiten zusammen, um aus einer Vielzahl von Quellen Informationen zu extrahieren und in einer Datenbank abzulegen.

Diese Modelle sind an unsere Suchmaschine angebunden. Wenn ein Ingenieur:in eine Suchanfrage stellt, erkennen die KI-Algorithmen die technischen Zusammenhänge und filtern die Informationen heraus, die zur Lösung der Problemstellungen entscheidend sind.

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Es war relativ leicht, Leute zu finden, die einen unterstützten: KI ist gut geeignet, um Werkstudent:innen zu finden, die in diesem Bereich arbeiten wollen.

Schwieriger war es hingegen, Ingenieur:innen mit ein bisschen Programmier-Background zu finden und sie in die Code-Entwicklung einzuarbeiten. Und wir hatten wenig Geld, aber hohe Kosten für das Hosting und das Trainieren der Algorithmen. Unser Plan, uns über Werbung zu finanzieren, ist nicht aufgegangen.

Bei Rokin klickt nur ein Bruchteil der Kundschaft auf gesponsorte Links. Ingenieure wollen Fakten und keine Werbung. Wir haben es leider nicht geschafft, ein tragbares Geschäftsmodell über unsere Suchmaschine aufzubauen und sind dabei, nach Alternativen zu suchen.

WAS MÜSSTE SICH ÄNDERN, UM EUER PRODUKT ZUKUNFTSSICHER ZU MACHEN?

Die Idee einer speziellen Suchmaschine fürs Ingenieurwesen ist schon zukunftsfähig. Aber man braucht dafür ein ordentliches Investment und das kriegt man nur, wenn man beweisen kann, dass man ein tragfähiges Geschäftsmodell hat.

KI-PERSONALISIERTE SEXUALBERATUNG FÜR ALLE

NAME	NESOMA HEALTH
GRÜNDUNG	2023
FÖRDERUNG	MEDIA COMPANY FELLOWSHIP #11, GROW FELLOWSHIP
WEBSEITE	NESOMA-HEALTH.COM
GRÜNDERIN	SARAH ELISABETH ALT

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Nesoma ist eine KI-basierte Sexual-Wellness-App. Wir helfen unseren Nutzer:innen, ihre Sexualität noch schöner zu machen oder das Schöne aufrechtzuerhalten. Uns ist wichtig, dass alle Nutzenden kein sexuelles „Problem“ haben müssen, sondern wir wollen ihr Leben bereichern.

WAS MACHT EUER PRODUKT EINZIGARTIG?

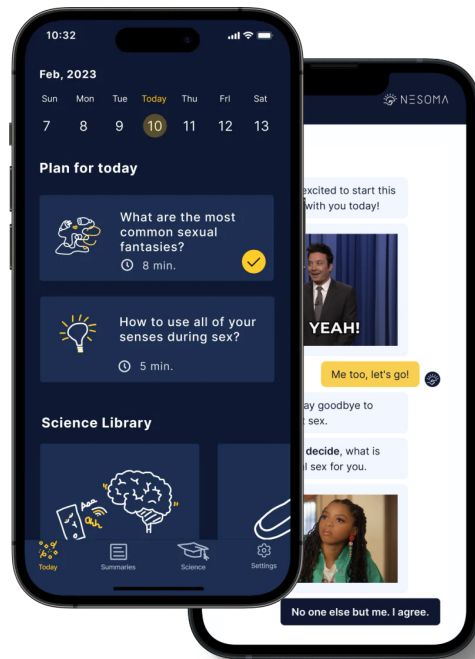
Es gibt nach wie vor großen Aufklärungsbedarf im Sexualbereich. Gerade auch bei Erwachsenen. Viele Lösungen richten sich an jüngere Menschen oder zum Beispiel nur an Frauen. Wir wollen geschlechtsneutral auftreten und diese Lücke füllen. Dass wir KI nutzen, um individuelle Beratung anzubieten, ist ebenfalls ein USP von uns. Und dann haben wir auch noch eine besondere Art, mit unseren Nutzer:innen zu kommunizieren, nämlich primär durch Chats.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Die Nutzenden bekommen täglich bis zu drei verschiedene Sessions zugeschickt, die individuell für sie ausgewählt werden und die sie dann bearbeiten können. Das Ganze funktioniert über einen Fragebogen, der zuerst ein paar Grundlegende Infos abfragt, zum Beispiel die sexuelle Orientierung. Danach können wir dann anpassen und fragen entsprechende Punkte zur sexuellen Zufriedenheit ab und erstellen dann für jede Nutzer:in eine individuelle Reise, über die sie mit unseren Inhalten und Themen in Kontakt kommen.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Unsere App ist ganz normal in React Native gebaut. KI nutzen wir dann an verschiedenen Stellen: Zum Beispiel für ein Empfehlungssystem, das Nutzenden direkt nach dem Ausfüllen des ersten Fragebogens passende Inhalte empfiehlt. Das funktioniert über Collaborative Filtering, das die Nutzenden in passende Gruppen einteilt. Diese KI hilft uns auch dabei, die individuellen Journeys zu erstellen.



Bilder: Nesoma Health

Für die tatsächliche Betreuung der Nutzenden verwenden wir LangChain, eines der besten Frameworks für Large Language Models, also ein Tool, dass hilft, mit diesen Modellen zu arbeiten. Als konkretes Sprachmodell nutzen wir Vicuna, ein Open-Source-Modell. Für die Ausgabe von Inhalten, also für die „Gesprächspartner“ im Chat, haben wir das mit Typebot kombiniert. Das ist ein Chatbuilder, ebenfalls Open Source, der sehr einfach in der Anwendung ist und zum Beispiel leicht mit ChatGPT verbunden werden kann.

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Eine Hürde ist, dass man erst wirklich etwas mit KI anfangen kann, wenn man die nötigen Daten hat. Gerade für Recommendation Systems braucht man große Datensätze. Sprachmodelle können oft schon mit wenigen Daten sehr gutem Finetuning unterzogen werden, auch, wenn man vorsichtig sein muss: Sie liefern manchmal eben doch sehr überzeugend klingende, aber falsche Antworten. Das ist gerade im medizinischen Bereich ein riesiges Problem.

Auf der anderen Seite sind die Frameworks oft fast kinderleicht anwendbar. Eine einfache Anwendung mit LangChain, zum Beispiel für eine Slack-App, kann man auch ohne größere Vorerfahrungen in wenigen Stunden einrichten.

IST DAS ZUKUNFTSSICHER?

Wir glauben nicht, dass zum Beispiel ChatGPT irgendwann die Funktion unseres Beratungs-Chats übernehmen kann. Bei einer Beratung geht es zwar um reine Fakten, die vielleicht auch andere Anbieter sammeln kann, aber es geht auch so sehr um die Haltung der Therapeut:innen oder Kursleiter:innen. Es geht darum, verschiedene Kontexte und Situationen von Sexualität einordnen zu können. Und dafür muss die KI extra entwickelt werden. Es ist sehr unwahrscheinlich, dass sich bei OpenAI jemand hinsetzt, und ChatGPT genau auf so einen Case trainiert.

MIT KI UND DATEN SCHNELL AUF KRISEN REAGIEREN

NAME	KONSOLE LABS
FÖRDERUNG	OPEN INNOVATION CHALLENGE 2020
WEBSEITE	KONSOLE-LABS.COM

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Daily Corona News war ein Projekt von Konsole Labs, das während der Hochphase der Corona-Pandemie Medienhäuser unterstützt hat, hyperlokale Inhalte aus Corona-Daten zu erstellen. **Die Corona-Zahlen waren ja oft in jedem kleinen Landkreis anders. Unser Ziel war es, einen Service anzubieten, der automatisiert Content aus aktuellen Daten erstellt und Nutzer:innen informiert.**

WAS MACHT EUER PRODUKT EINZIGARTIG?

Unsere Lösung war so aufgebaut, dass sie beliebig skalierbar und trotzdem preiswert und für die Medienhäuser einfach in der Anwendung war. Wir brauchten nur einen Dateneingang, und konnten beliebig viele Nutzer:innen damit beliefern. Und gleichzeitig ist das alles unter dem Eindruck der damaligen Ereignisse entstanden und setzte auch für uns als Anbieter keine Raketenwissenschaft voraus.



Bild: Konsole Labs

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Es gab eine Anmeldemaske, in der man sich als Medienhaus registrieren konnte. Dort konnte man dann die Regionen auswählen, in denen die eigene Leser- oder Hörerschaft zu Hause war, dafür gab es eine Suchmaske. Und dann bekamen die Nutzenden aus diesen Regionen und auf Basis der aktuellen Corona-Zahlen sowohl fertige Texte, als auch synthetisch vertonte Versionen dieser Texte zur Nutzung auf ihren Kanälen geliefert.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Konsole Labs arbeitet schon seit 2020 an ihrer Content Creation Intelligence, CCI. Das ist ein Browser-CMS für die Erstellung von automatisierten Texten und Audio-Inhalten mittels synthetischer Stimmen. Das ist, was die Datenquellen angeht, sehr flexibel. Für die Stimmen nutzen wir unter anderem Angebote von bottalk.io, Microsoft oder Google und für die Texte arbeiten wir mit AX Semantics und der Schnittstelle von OpenAI. Die so generierten Audios werden dann per FTP oder API in die Angebote oder Sendesysteme von Radiosendern exportiert.

Für Daily Corona News war die Herausforderung besonders, die neue Datenbasis zu schaffen und die neuen Datenquellen zu integrieren. Dazu haben wir mit news.de zusammengearbeitet, die die öffentlich zugänglichen Daten vom Robert-Koch-Institut und anderen gesammelt hat.

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Technisch ist das alles kein Hexenwerk—besonders, weil wir an der technischen Basis schon länger gearbeitet hatten. Wir führen Datenquellen zusammen und bauen ein kohärentes Tool aus verschiedenen Angeboten und Microservices zusammen. Die größere Hürde war bei diesem Projekt eigentlich, die Zeit für die Umsetzung zu finden und die richtige Idee zur richtigen Zeit für die richtigen Leute umsetzen zu können. Und dass die Lösung gleichzeitig reibungslos nutzbar ist.

IST DAS ZUKUNFTSSICHER?

Für spezialisierte Anwendungsfälle braucht es immer auch spezialisierte Daten, die zum Beispiel OpenAI nicht hat, die wir aber liefern können. Und es wird auch immer Platz für kleinere Dienstleister geben, die die Werkzeuge der großen Player nutzen, um aus verschiedenen Lösungen ein kohärentes Produkt zu bauen, das für einen konkreten Anwendungsfall passt.

Unsere Lösung war so aufgebaut, dass sie beliebig skalierbar und trotzdem preiswert und, für die Medienhäuser, einfach in der Anwendung

FEHLENDE DIVERSITÄT MUSS MAN ERST ERKENNEN

NAME	FOCAL ANALYTICS
FÖRDERUNG	MEDIA STARTUP FELLOWSHIP #6
WEBSEITE	DIVERSAL.FOCAL-ANALYTICS.COM
GRÜNDER	DR. MAKUS HÖRMANN & DR. ULRICH HÖRMANN

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Diversal Analytics beschäftigt sich mit dem Problem, dass Journalismus unserer Meinung nach nicht divers genug ist—nicht nur in der Produktion, sondern auch in den Inhalten selbst. **Junge Frauen und People of Color werden viel seltener zitiert und abgebildet als ältere, weiße Männer. Dieses Problem können wir natürlich nicht direkt lösen, aber unsere Text- und Bildanalysen helfen dabei, das Problem—und damit auch Wege aus dem Problem—sichtbarer zu machen.**

WAS MACHT EUER PRODUKT EINZIGARTIG?

Als wir unsere Marktanalyse gemacht haben, gab es noch kein Tool, das diese Infos an einem Ort gesammelt und aufbereitet hat. Man musste sich das selbe Wissen aus verschiedenen Tools zusammensuchen und oft auch noch die Daten mühsam selbst einspeisen. Bei Diversal ist die komplette Pipeline von Dateneingabe bis Visualisierung der Ergebnisse in einem Dashboard abgedeckt.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Diversal ist eine webbasierte Anwendung. Am Anfang braucht es erstmal die Daten, also die Texte der Nachrichtenseite oder des Blogs. Wir haben einen Crawler dafür, der individuell angepasst werden muss—das übernehmen wir. Für die Nutzenden ist dann die Arbeit schon erledigt. Nach dem Login sieht man ein Dashboard, das verschiedene Fragen zur Diversität in den eigenen Texten beantwortet—bisher hauptsächlich Fragen, wie viele Frauen und nicht-weiße Menschen erwähnt werden, aber auch, wie die Altersverteilung der Abgebildeten ist.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Eine Herausforderung ist es, die Daten richtig zu crawlen. Zu erkennen, was zum Beispiel auf einer Nachrichtenseite ein relevanter Text oder Link ist, und was nur ein Link zu einem anderen Ressort oder zu einer Anzeige, ist für einen Computer schwierig. Wenn man vom Verlag eine API hat, ist es natürlich einfacher. Aber da das viele nicht haben, ist eben die Anpassung des Crawlers nötig.

Dann nutzen wir ein Named Entity Recognition Modell, das Personen in Texten und ihre Beziehung zueinander erkennt. Dafür haben wir eine fertige Bibliothek angeschlossen, es gibt da schon ein paar gut funktionierende Modelle. Wir nutzen spaCy, ein weitverbreitetes Tool für Anwendungen im Bereich Natural Language Processing, und haben einen Microservice drumherum gebaut, der das für unseren Use Case nutzt.

Junge Frauen und People of Color werden viel seltener zitiert und abgebildet als ältere, weiße Männer. Unsere Text- und Bildanalysen helfen dabei, das Problem—und damit auch Wege aus dem Problem—sichtbarer zu machen.

Die Personen, die wir in Texten identifizieren, gleichen wir mit WikiData ab und bekommen so die Daten zu Alter, Geschlecht und Hautfarbe.

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Damals war die Arbeit mit fertigen Modellen noch schwieriger und teurer. Trotzdem war für uns die Datenaquise und Datenaufbereitung die größere Hürde als die tatsächliche Anwendung der Modelle. Und das ist heute ja auch noch viel einfacher geworden. Wenn man eine gute Datenbasis hat, gibt es so viele APIs zu gut funktionierenden Modellen, dass das wirklich nicht mehr besonders schwierig ist.

IST DAS ZUKUNFTSSICHER?

Man könnte das Tool sicher noch ausbauen und den Schritt zu selbsttrainierten Modellen machen oder wenigstens die bestehenden Modelle besser auf den eigenen Use Case anpassen.

Gleichzeitig wird es vermutlich nur wenige bekannte Player geben, die wirklich große, wirklich umfassende Modelle bauen. Dafür braucht es viel zu viel Rechenpower und Know-How. Die Chance für alle anderen Unternehmen besteht darin, diese Modelle in Form zu bringen und anzupassen für eine Million denkbare, spezialisierte Anwendungen. Sei es durch Anpassung oder dadurch, dass man verschiedene Modelle zu neuen Lösungen zusammenführt. Da spielt dann am Ende User Interface Design fast die größere Rolle als die dahinterstehende KI.

WENN DIE KI DIE INHALTE DOCH *SELBST* ÜBERNIMMT

Der Chatbot on OpenAI ist für viele synonym mit „KI“. Dieser Report zeigt, wie viele andere Anwendungsmöglichkeiten es trotzdem gibt – auch im Bereich der automatisierten Content-Erstellung.

Synthetische Inhalte sind für Medien schon länger kein Nischenthema mehr. Die BBC experimentiert in ihrem „Blue Room“ schon seit einigen Jahren damit, News beispielsweise von computergenerierten Avataren vortragen zu lassen. Und Computer, die Inhalte nicht nur vortragen, sondern auch erstellen, sind seit langem die Grundlage der Vision, die gemeinhin als „Roboterjournalismus“ bezeichnet wird.

Der Hype hinter ChatGPT hat diesen Ideen neuen Auftrieb verschafft. Dabei geht mit synthetischer KI noch viel mehr, als sie nur Artikel schreiben zu lassen.

Ein gutes Beispiel dafür ist OneGuide. Wer bisher eine Städtetour machen wollte, war dafür auf vorgefertigte Guides und Reiseführer angewiesen, die zwar viele Informationen enthielten, aber für alle Nutzenden gleich waren. One Guide erstellt Städtetouren, die Rücksicht auf die eigenen Interessen, die eigenen Vorstellungen und den eigenen Geldbeutel nehmen, angereichert mit Tipps, Empfehlungen und Erklärinhalten – alles unterstützt durch AI.

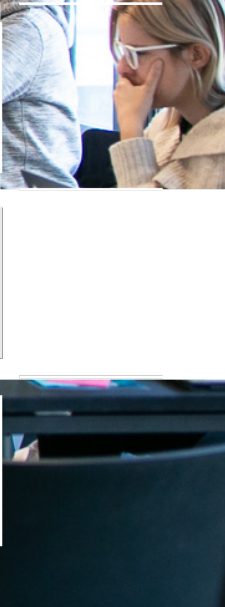
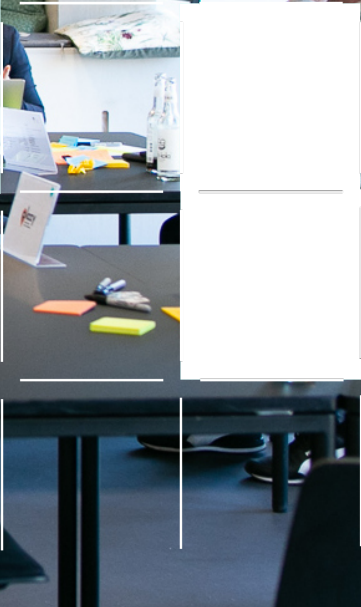
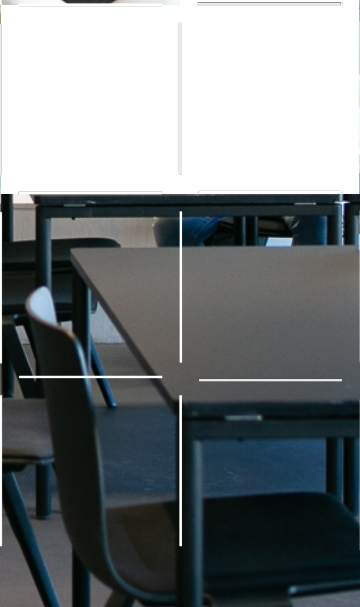
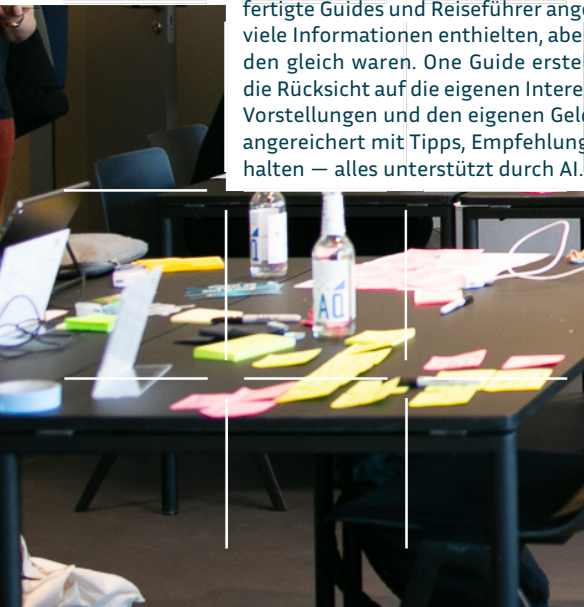
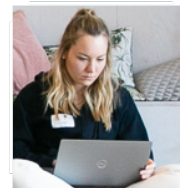
Bei Television.AI unterstützt eine KI Journalist:innen dabei, Videoclips und Nachrichteneinspieler zu erstellen – nur mithilfe des Nachrichtentextes und ihres Rohmaterials. Die Zeit von der Ankunft im Schnittraum bis zum fertigen Clip kann damit von oft über 90 auf knapp 15 Minuten reduziert werden. Die zusätzliche Zeit können Journalist:innen für ihre Kernaufgabe nutzen: ihren Journalismus noch besser zu machen.

Dasselbe Ziel, nämlich durch synthetischen Content mehr Zeit für die Kernaufgaben zu schaffen, verfolgen auch andere KI-Wegbereitende aus dem Media Lab. Zum Beispiel HypeRate: Tausende Videostreamer:innen arbeiten schon mit der Software, die Emotionserkennung und biometrische Daten nutzt, um die besten Momente eines Streams automatisch in handliche Clips für die großen Social-Plattformen umzuwandeln. Und SnipAid erstellt Überschriften, Teaser und andere Text-Snippets aus journalistischen Texten. Eine starke Unterstützung für das Klein-Klein der Textarbeit.

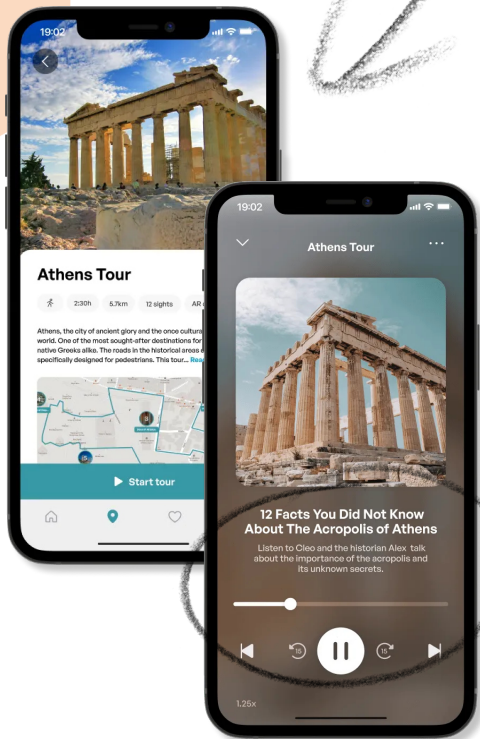
Gleichzeitig kann KI auch Inhalte erstellen, die Endnutzer:innen zugutekommen. Kompreno etwa ist mehr als eine Übersetzungssoftware: Durch die Kuration und Übersetzung von Nachrichtenartikeln aus ganz

Europa machen sie Inhalte, die bisher hinter Sprachbarrieren versteckt waren, für mehr Menschen nutzbar und tragen so dazu bei, eine europäische Medienöffentlichkeit zu schaffen.

Die folgenden Seiten zeigen: Wenn es um synthetische Inhalte geht, ist ein Chatbot, egal wie klug er sein mag, nicht die einzige denkbare Lösung. KI kann Menschen dabei helfen, Inhalte effizienter zu erstellen, und sie vor allem individueller zu machen: So können alle Nutzenden genau die Empfehlungen, Daten oder Inhalte bekommen, die für sie zugeschnitten sind.



DER REISEFÜHRER, BEI DEM ALLE INHALTE VON DER KI KOMMEN



ONEGUIDE ERSTELLT DIE INHALTE FÜR SEINE AUDIO- UND AUGMENTED-REALITY-REISEFÜHRER MIT EINEM SELBST ENTWICKELTEN KI-MODELL.

Bild: OneGuide

NAME	ONEGUIDE
GRÜNDUNG	2022
FÖRDERUNG	MEDIA COMPANY FELLOWSHIP #12
WEBSEITE	ONE-GUIDE.DE
GRÜNDER:INNEN	WIEBKE NADZEIKA & MARLON LÜCKERT

Alle Urlauber:innen kennen sie, viele nutzen sie: die Reiseführer von Lonely Planet gelten als verlässliche Quelle, liefern Tipps und Inspirationen für alle, die die Welt entdecken wollen.

Auch Wiebke Nadzeika hat den Lonely Planet früher selbst genutzt, doch das hat sich geändert: "Reiseführer sind umständlich und heute will niemand mehr groß lesen, um eine Stadt zu entdecken oder mehr über eine Sehenswürdigkeit erfahren will", sagt die Mitgründerin von OneGuide. Zwar bieten viele Sehenswürdigkeiten und Museen inzwischen auch Audioguides an, aber „da sind häufig die Inhalte viel zu lang und von schlechter Qualität“.

Und genau da setzen Nadzeika und ihr Kollege Marlon Lückert mit OneGuide an: Ihre App bietet Reisenden Sightseeing-Touren in Form von gut portionierten, als Dialog aufgezogenen Audioguides, ergänzt durch AR-Inhalte.

DIALOGE WERDEN VON EINEM SPEZIELLEN OPENAI-MODELL GESCHRIEBEN

Die Audio-Inhalte werden in Form von Dialogen zwischen zwei Sprecher:innen angeboten. Gestartet ist OneGuide mit von Menschen erstellten Inhalten, die in einem ersten Prototypen A/B getestet wurden, um Erkenntnisse über den perfekten Inhalt zu gewinnen.

Jetzt kommen die Dialoge von einem eigenen KI-Modell, das auf einem angepasstem OpenAI-Modell basiert. Das Training erfolgt durch ein Trainingsdaten-set bestehend aus Prompts und einer idealen Antwort. Bei Large Language Models kommt es häufig zu sogenannten Halluzinationen, bei der KI ihre Wissenslücken mit falschen Informationen auffüllt. Um dem zu begegnen, entwickelten Nadzeika und Lückert ein Verfahren, dass es der KI erlaubt, nur die Fakten aus ausgewählten, zuverlässigen Datenquellen zu verwenden.

OneGuide erhebt auch eigene Quellen, wie Wissen von Tourguides oder Interviews mit Zeitzeug:innen. Die Herausforderung dabei: Ein Prompt und dessen Antwort haben eine maximale Zeichenlänge, die nicht überschritten werden darf. „Deswegen haben wir ein System entwickelt, welches automatisiert nur die relevanten Information in den Prompt packt“, sagt Lückert, der bei OneGuide auch CTO ist. Durch das Training entsteht ein eigenes OneGuide-Modell.

„Es genügt eine Stimmprobe von etwa fünf Minuten und die geklonten Stimmen klingen einfach besser“

NUTZENDE TRAINIEREN MIT

Menschliche Autor:innen schreiben eine ideale Version von Einleitung und Dialog. OpenAI zufolge braucht es 100 bis 100.000 solcher Beispiele für ein optimales Feintuning. Je mehr Trainingsdaten, desto besser wird das Ergebnis.

Um das Training zu beschleunigen, verwendet das Hamburger Startup eine zweite Methode zu Finetuning: „Reinforcement Learning from Human Feedback“. Dabei lernt das Modell auf Basis von direktem menschlichen Feedback. Zum einen vom OneGuide-Team. Zum anderen von den App-Nutzer:innen, indem sie jede Folge mit eins bis fünf Sternen bewerten können. Auch die Verweildauer ist eine Form der Bewertung. „Wir haben das Training also aufs Produkt verlagert. Dadurch lernt unsere KI dauerhaft und wird auf diese Weise immer besser.“

In Zukunft können wir so sogar personalisierte Inhalte erstellen“, sagt Lückert. Eine Beta-Version der App ist bereits in Berlin verfügbar mit aktuell 8000

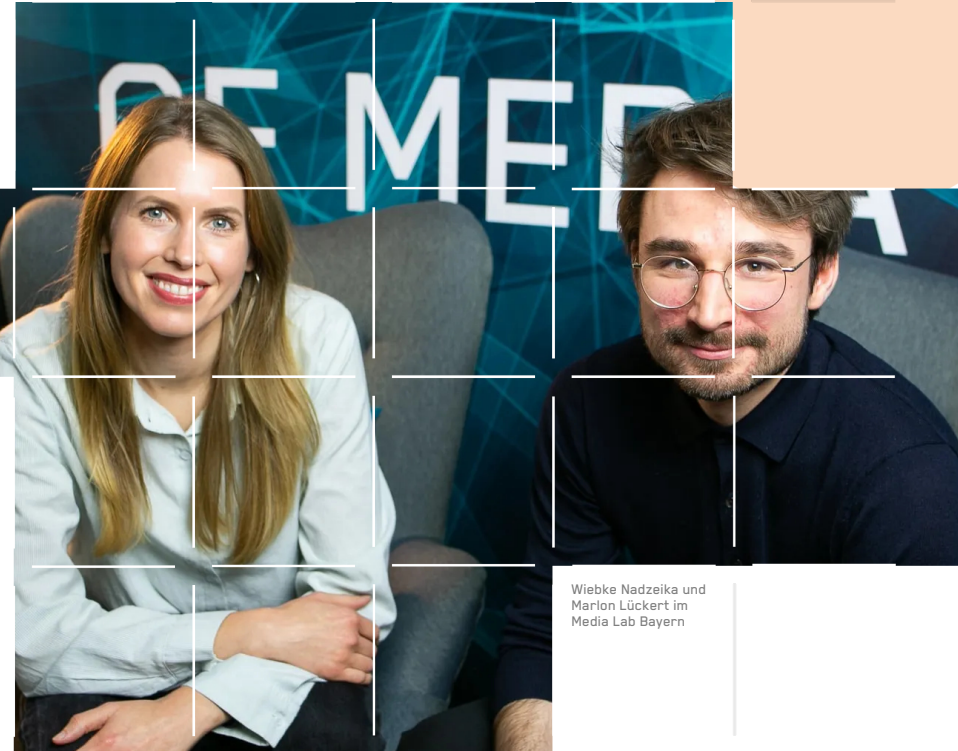
Nutzer:innen pro Monat. Es gibt momentan 37 Audio Stories in vier Sprachen und passend dazu vier AR-Experiences zu ausgewählten Berliner Sehenswürdigkeiten wie der East Side Gallery. „Das Produkt schon in der Frühphase zu testen und dabei auch die Nutzer:innen einzubeziehen, war ein Learning aus dem Media Lab Startup Fellowship“, berichtet Lückert.

Nach dem Training wird mittels einer Verkettung mehrerer Prompts das Audio-Guide-Skript erstellt. Hierzu hat OneGuide verschiedene Prompt-Templates entwickelt. Dann generiert das OneGuide-Modell den gewünschten Dialog.

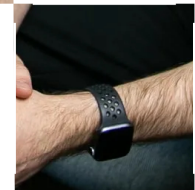
STIMMEN WERDEN MIT EINEM TEXT-TO-VOICE-TOOL GENERIERT

Wenn der Dialog fertig ist, folgt Schritt zwei: der geschriebene Guide wird in Audio umgewandelt. Dafür wird der Text in ein Text-to-Voice-Tool eingespeist. Dabei verwendet OneGuide nicht öffentlich verfügbare, synthetische Stimmen, sondern klonnt menschliche. Teilweise die eigenen oder die von anderen Menschen mit „schönem Klang“, die ihre Rechte an der Nutzung ihrer Stimmen an OneGuide verkaufen. „Es genügt eine Stimmprobe von etwa fünf Minuten“, sagt Lückert, „und die geklonten Stimmen klingen einfach besser.“ Innerhalb weniger Minuten entsteht so ein fertig vertonter KI-Audioguide.

Auch wenn momentan die Tonspur im Vordergrund steht: OneGuide will künftig noch mehr AR-Inhalte in seiner App anbieten. „Mithilfe der AR sieht man beispielsweise die Mauer am Brandenburger Tor oder die Wachtürme an der Gedenkstätte Berliner Mauer wieder. Auch hat man die Möglichkeit, in der East Side Gallery ein beliebtes Trabi-Bild aus der Wand fliegen zu sehen“, veranschaulicht Lückert.



Wiebke Nadzeika und Marlon Lückert im Media Lab Bayern



Auch die AR-Inhalte sollen durch KI generiert werden. Noch sei die Adaptionrate für AR-Inhalte gering, sagt Wiebke Nadzeika. „Aber wir glauben, dass in der Zukunft alltagstaugliche AR-Brillen die Smartphones ersetzen werden. Und dann können wir Geschichte und Kultur noch viel besser und immersiver erleben.“ Nadzeikas Vision ist, dass OneGuide der neue Lonely Planet wird—nur eben als App.

WENIGER SCHNITT FÜR MEHR JOURNALISMUS

NAME	TELEVISION.AI
GRÜNDUNG	2019
FÖRDERUNG	MEDIA COMPANY FELLOWSHIP #8
WEBSEITE	TELEVISION.AI
GRÜNDER	OLCAY BUYAN

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

In Fernsehsendern, Broadcastern und Verlagen ist es sehr aufwändig und oft mit großem Zeitdruck verbunden, Video-Rohmaterial zu fertigen Clips oder Nachrichteneinspielern zusammenzubauen. **Television.AI nutzt KI, um das Material zu sichten, zu analysieren und anhand eines Nachrichtentextes fertige Videos zu generieren. Was vorher eine Journalist:in und eine Cutter:in 90 Minuten gekostet hat, erledigt unsere Technologie in 15 Minuten.** Man kann also sagen, wir machen Filmschnitt einfacher, schneller und effizienter und schaffen mehr Zeit für die eigentliche journalistische Arbeit.

WAS MACHT EUER PRODUKT EINZIGARTIG?

Im Gegensatz zu vielen anderen KI-Lösungen haben wir beschlossen, keine eigenen Inhalte zu erstellen, sondern die Arbeit mit bestehenden Inhalten zu vereinfachen. Wer heute schon viel mit KI-generierten Inhalten arbeitet, steht irgendwann vor dem Flaschenhals, vor dem Redakteur:innen und Redakteure schon früher standen: Aus viel zu viel Material die wichtigsten Bilder oder Inhalte herauszufiltern. Und dabei helfen wir.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Als Redakteur:in loggt man sich bei uns auf der Website ein und lädt das Rohmaterial hoch. Die Redakteur:in kann in den Prozess manuell eingreifen, indem zum Beispiel bestimmte Dateien markiert werden, die die KI auf jeden Fall einbeziehen soll.

Das fertige Video kann anschließend in unserem Web-Editor weiter bearbeiten werden. Du brauchst weder klassische Schnittprogramme wie Adobe oder FinalCut, noch einen leistungsstarken Computer. Das Ergebnis kann dann einfach heruntergeladen werden.

Wie kann man mit einer Masse an Inhalten effizient arbeiten?

Wer die bekannten Workflows mit klassischen Programmen aber weiter nutzen will, kann diese über unsere API integrieren. Die erzeugten Clips landen dann in der gewohnten Schnittoberfläche und können dort weiter bearbeitet werden.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Zurzeit nutzen wir KI-Technologien in drei Kategorien: Ein Teil ist komplett selbstgebaut, ein zweiter Teil sind Open-Source-Technologien, die wir selbst angepasst haben, und ein dritter Teil sind Open-Source-Lösungen. Am Anfang haben wir fast komplett mit Out-Of-The-Box-Software gearbeitet, etwa mit AWS Recognition und vor allem mit Microsofts Azure Video Indexer. Wir haben uns auch das entsprechende Google-Tool angeschaut, aber da hatten wir Bedenken bei Datenschutz und Datensicherheit. Gerade für unsere Zielgruppe in journalistischen Organisationen ist das enorm wichtig, und da ist Microsoft stärker.

Nach und nach haben wir dann die Nutzung dieser fertigen Lösungen reduziert und immer mehr selbst angepasst und gebaut. Damit konnten wir nicht nur die Leistung unserer Software für genau unseren Use Case steigern, sondern auch die Kosten für die Nutzung der Modelle und Server enorm reduzieren.

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Unsere größte Hürde war es, das richtige Geschäftsmodell zu finden. Es hat eine Weile gedauert, bis wir uns auf unser heutiges Pricing-Modell festgelegt haben, das einen günstigen Einstieg ermöglicht und dann mit demonstriertem Mehrwert die Nutzen einfängt. Wir sind komplett selbst finanziert, und gerade als wir noch die hohen Kosten für Technik im Rücken hatten, war das schwierig.

Die technische Seite war dagegen fast einfach, weil man hier viel ausprobieren und iterativ immer ein kleines bisschen besser werden kann. Gerade auch, weil man mit fertigen Lösungen anfangen kann. Im engen Austausch mit unserer Kundschaft konnten wir dann unser Produkt nach und nach verbessern.

IST DAS ZUKUNFTSSICHER?

Nichts ist für immer zukunftssicher, schon gar nicht in einem Bereich, der sich so schnell verändert. Wir glauben aber, zumindest für die nächsten Jahre auf dem richtigen Zug zu sein. Es zeichnet sich ja schon ab, dass es immer mehr und mehr Inhalte geben wird, ob von den Menschen gemacht oder von KI generiert. Wir setzen einen Schritt danach an, und haben uns gefragt: Wie kann man mit einer Masse an Inhalten effizient arbeiten? Je mehr KI-Tools zur Inhaltserstellung es gibt, desto sicherer sind wir uns, dass es eine Lösung wie unsere braucht.

HYPERATE

HERZRASEN FÜR RICHTIG GUTEN CONTENT

NAME	HYPERATE
GRÜNDUNG	2021
FÖRDERUNG	MEDIA COMPANY FELLOWSHIP #8
WEBSEITE	TELEVISION.AI
GRÜNDER	HENDRIKJE WAGNER & ALEX HOLZREITER

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

HypeRate hilft Streamer:innen, Sichtbarkeit und Reichweite zu generieren. Wir tun das, indem wir ihnen helfen, automatisch Clips aus ihren Streams zu erstellen. Dazu nutzen wir zum einen die Herzrate des Streamenden, zum anderen aber auch KI-gestützte Emotionserkennung. Beides hilft uns dabei, die besten Momente eines Streams zu identifizieren und in Content zu verwandeln, der dann über alle gängigen Plattformen geteilt werden kann.

WAS MACHT EUER PRODUKT EINZIGARTIG?

Die Nutzung der Emotionserkennung, besonders aber auch die Nutzung von biometrischen Daten wie der Herzrate. Andere Tools verwenden als Signal hauptsächlich den Content selbst. Wie z.B. im Gaming-Bereich, wenn ein Highscore oder Ähnliches erreicht wird. Das ist dann vielleicht ein Highlight des Spiels für den Spielenden, aber nicht unbedingt ein Highlight des Streams für die Zuschauerinnen und Zuschauer. Für die geht es um Momente, in denen es emotional wurde, aufregend, oder in denen man zusammen gelacht hat.

Unsere Stärke ist, dass wir emotionale Verbindungen messen und nutzbar machen.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Die einfachste Art, HypeRate zu nutzen, ist durch unsere App für Smartphone und Smartwatch. Mit der Smartwatch messen wir die Herzrate, und die Streamer:innen können sie dann live in ihrem Stream einblenden.

Was dazukommt, ist dann die Möglichkeit, Clips erstellen zu lassen. Dazu brauchen die Streamer zusätzlich nur ein Plugin für OBS, die meistgenutzte Streamingsoftware, und ein Smartphone, dessen Kamera sie auf ihr Gesicht richten müssen. Das Smartphone filmt zum einen die Reaktionen der Streamenden für die Clips und erkennt auch, wenn der Streamer zum Beispiel laut lacht. Wenn dann Herzrate und Emotion aus der Gesichtserkennung beide anzeigen „Hier passiert gerade etwas!“, wird im OBS automatisch ein Clip dieses Momentes erstellt.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Die Steuerungszentrale ist das Smartphone. Es empfängt die Daten aus der Herzratenmessung und berechnet einen Durchschnittswert, um Ausreißer besser erkennen zu können. Es sendet auch das Signal „Jetzt einen Clip erstellen!“ an OBS. Und dank Apples Vision Framework haben iPhones auch schon die KI-Gesichtserkennung dabei. Es findet also alles dezentral statt, wir müssen keine Clips in die Cloud laden, um dort die Emotionen zu tracken; das ist unter Datenschutzaspekten super.

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Der Weg bis zur aktuellen Lösung war ziemlich lang. Ursprünglich wollten wir die Clips live von Twitch abgreifen und dann mit einer KI analysieren—wir hatten über Amazons AWS Cloud eine KI mit mehr als 14.000 Clips selbst trainiert. Wir haben das dann aber eingestellt, weil diese Cloud-Lösung sehr teuer ist, gerade, wenn man so viele Nutzende mit viel Videomaterial hat. Und für dieses Investment waren uns die Ergebnisse dann auch nicht gut genug.

Als nächstes haben wir versucht, die KI mit Hilfe eines Python-Skripts direkt in das OBS-Plugin zu packen. Dafür gibt es öffentlich zugängliche Libraries, die Emotionserkennung anbieten. Dabei gibt es aber hohe technische Anforderungen an die Rechner der Nutzenden.

Wir hatten dann aber fast zufällig einen sehr guten iOS-Entwickler im Team. Und durch ihn sind wir dann darauf gekommen, das Ganze so leicht wie möglich zu bauen, ohne Cloud, und einfach auf bestehende Kits von Apple zu setzen. Und das funktioniert jetzt hervorragend, effizient und dezentral.

IST DAS ZUKUNFTSSICHER?

Die HypeClips sind unser großes Produkt, aber der Kern des Ganzen sind eigentlich die biometrischen Daten, die wir messen: also die Herzrate. Die Emotionserkennung in Verbindung mit Filmschnitt kann eine andere KI vielleicht schon bald problemlos umsetzen, aber wir können andere, zusätzliche Inputs direkt von den Menschen hinter dem Content nutzen, um Erlebnisse zu individualisieren. Das kann man auch im Game Design einsetzen. Oder im User Testing. Oder bei Events. Unsere Stärke ist, dass wir emotionale Verbindungen messen und nutzbar machen.

SIEBEN SNIPPETS AUF EINEN SCHLAG

NAME	SNIP AID
FÖRDERUNG	MEDIA TECH LAB
WEBSEITE	GITHUB.COM/SNIP-AID-NLG
ENTWICKLER:INNEN	HANNAH GREVEN & BENJAMIN RECH

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Wir sind ein zweiköpfiges Team und haben mit Snip Aid ein Tool entwickelt, das es erlaubt, im Handumdrehen verschiedene Snippets für journalistische Texte zu generieren. Wenn Journalist:innen Überschriften, Teaser, Zusammenfassungen, Keywords, Title-Tags, Meta-Descriptions und Tweets selber formulieren müssen, kann das Stunden dauern. Mit unserem Modell bekommt man diese sieben Snippets in unter drei Minuten—es ist also eine enorme Zeitersparnis.

WAS MACHT EUER PRODUKT EINZIGARTIG?

Es gibt sonst kein Open Source Tool, mit dem man binnen Sekunden sieben verschiedene Snippets für journalistische Texte erstellen kann.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Man gibt den Text, zu dem Snip Aid ein Snippet generieren soll, in ein Fenster ein. In einem Dropdown-Menü kann man aus drei Sprachmodellen auswählen. Für deutsch funktioniert „snip-igel“ am besten. Dann klickt man auf „generieren“, anschließend schreibt das Modell alle sieben Snippets. Wichtig ist, dass man checkt, ob das Ergebnis inhaltlich korrekt ist, denn wie wir alle wissen, kann generative KI Fehler machen.

Wenn Journalist:innen Überschriften, Teaser, Zusammenfassungen, Keywords, Title-Tags, Meta-Descriptions und Tweets selber formulieren müssen, kann das Stunden dauern. Mit unserem Modell bekommt man diese sieben Snippets in unter drei Minuten

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Im Frühjahr 2023 kamen erste deutsche Sprachmodelle auf den Markt, die anhand von Prompts, kurzen Aufgabenbeschreibungen, passenden Text generieren können. Darauf konnten wir aufbauen und gezielt die Fähigkeiten eines dieser Modelle zur Generierung von News Snippets verbessern.

Dafür haben wir aus einem großen Trainingsdatensatz mit Nachrichtenartikeln und zugehörigen Zusammenfassungen rund 500 Artikel herausgegriffen. Für diese Nachrichtenartikel haben wir GPT-3.5, Keywords und Tweet generieren lassen. Im nächsten Schritt haben wir GPT 25 Prompts für jeden Snippet-Typ generieren lassen. So entstand ein Trainingsdatensatz mit knapp 3000 Zeilen. Für jeden der 500 Artikel jeweils 6 Snippets samt Prompts: Title, Teaser, Keywords, SERP-Title-, SERP-Description und Tweets. Mit diesem Datensatz haben wir das ursprüngliche Modell namens IGEL (steht für „Instruction-tuned German Large Language Model“) weitertrainiert, woraus dann das neue Modell Snip-IGEL entstand.

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Das größte Problem war, ein deutschsprachiges Modell zu finden, das klein genug ist, dass man es auch selbst noch hosten kann. Zu Beginn gab es nur englischsprachige Modelle. Und es war sehr aufwändig, sie zu trainieren. Dann mussten wir den Umweg über eine Übersetzung von Englisch in Deutsch gehen.

Das Sprachproblem hat sich erst im Frühjahr 2023 gelöst, als auch deutschsprachige Modelle auf den Markt kamen. Das IGEL-Modell erwies sich als geeignet, um es ressourceneffizient finetunen zu können. Schon mit 500 deutschsprachigen Artikeln erzielten wir sehr gute Ergebnisse. Dieses Training hat auch nur etwa 5 Dollar gekostet. Das Resultat ist das Snip-IGEL-Modell.

WARUM IST EUER PRODUKT ZUKUNFTSSICHER?

Das Snip-IGEL-Modell ist klein genug, dass man es noch selbst auf einem Server mit GPU hosten kann. Und es lässt sich sehr günstig und ressourceneffizient trainieren. Dieses Modell steht als Open Source zur Verfügung, jedes interessierte Medienhaus kann es installieren und nach seinen Bedürfnissen weiterentwickeln.

AUF DEM WEG ZUR EUROPÄISCHEN ÖFFENTLICHKEIT

NAME	KOMPRENO
GRÜNDUNG	2020
FÖRDERUNG	R&D FELLOWSHIP
WEBSEITE	KOMPRENO.EU
GRÜNDER	JOCHEN ADLER

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Kompreno ist ein Netzwerk für europäischen Qualitätsjournalismus. Zu unseren Partnern zählen Medienhäuser aus verschiedenen europäischen Ländern wie Le Monde (Frankreich), The Independent (Großbritannien), oder Die Zeit aus Deutschland. Wir haben aber auch Publikationen aus Ländern wie Dänemark, Bulgarien, Albanien oder der Ukraine dabei. Wir lösen zwei Probleme:

Erstens bieten wir in der heutigen komplexen Nachrichtenlandschaft ausgerichtete Informationen an. Wir legen Wert auf tiefgründigen Journalismus, um unseren Zielgruppen ein besseres Verständnis der Welt zu ermöglichen. Wenn ich verstehen will, was in Russland los ist und warum Prigoschin mit Putin im Clinch liegt, dann reicht es nicht, nur die Tagesschau zu kucken. Da braucht es Analysen, Kommentare und Reportagen—am besten aus verschiedenen verlässlichen Quellen. Und die bieten wir.

Zweitens möchten wir den europäischen Journalismus aus der Sprachblase herausholen. Der europäische Sprachraum ist sehr fragmentiert und mit unseren übersetzten Artikel wollen wir helfen, das zu überwinden. Unser Ziel ist es, Themen wie die Energiewende in einer Mischung aus verschiedenen europäischen Sprachen anzubieten, um ein breiteres Publikum anzusprechen.



kompreno

Bild: Kompreno

WAS MACHT EUER PRODUKT EINZIGARTIG?

Kompreno ist auf Skalierung ausgelegt: Bislang haben wir mehr als 30 Partner-Publikationen, es können aber auch 300 oder sogar 3000 sein. Mit der Zahl der Partner wächst auch die thematische Breite. Und wir können einen Nachrichtenüberblick bieten—und zwar aus verschiedenen Quellen. Wir können eine Vielzahl von Perspektiven anbieten: konservativ, progressiv, links, rechts; wir decken alles ab. Für Verlage ist Kompreno interessant, weil sie so ihre Reichweite vervielfachen können.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Nutzer erhalten von unserem Algorithmus Artikelvorschläge, die ihren Interessen entsprechen. Zusätzlich zu unseren Vorschlägen können die Nutzer zu den Themen, die sie interessieren, weiter recherchieren. Dabei werden relevante Treffer aus verschiedenen Quellen angezeigt, die qualitativ hochwertig und fundiert sind—besser als bei Google.

Kompreno übersetzt die Artikel der Medienpartner mit Hilfe von DeepL in fünf Sprachen: Englisch, Französisch, Spanisch, Italienisch und Deutsch, wobei die Inhalte in der Sprache des Nutzers angezeigt werden.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

In enger Zusammenarbeit mit unseren Verlagspartnern haben wir ein mehrsprachiges Content Management System (CMS) entwickelt, das auch das Lizenzmanagement regelt. Die Artikel lassen wir mit DeepL übersetzen, das über eine API in unser CMS eingebunden ist.

Unser Empfehlungsalgorithmus ist eine Eigenentwicklung auf der technischen Basis unseres CMS. Er basiert auf semantischen Wörteranalysen, die wir selbst pflegen und vornehmen, was sowohl thematisch als auch standortbezogen funktioniert: Wenn jemand sich für Migration in Südeuropa interessiert, bekommt er/sie Artikel vorgeschlagen, die für diese beiden Begriffe hohe Punktzahlen erzielen.

Dabei ist es unerheblich, ob der Artikel unseres Partners ursprünglich aus Bulgarien oder Belgien kam—also auf bulgarisch oder niederländisch vorlag. Eine deutsche Leser:in findet immer den relevanten deutschsprachigen Artikel. Im Volltext!

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Eine der größten Herausforderungen war das Henne-Ei-Problem: ausreichend Inhalte zu haben, um genügend Abonnenten anzulocken, aber auch genügend Abonnent:innen zu haben, um von den Verlagen ausreichend Inhalte zu bekommen. Der Markt für Abonnements ist umkämpft. Besser als gedacht läuft hingegen die Akquise neuer Medienpartner. Bei den Medienhäusern rennen wir offene Türen ein, denn die sind alle auf der Suche nach neuen Leserschaften. Der Umsatz aus einer fremdsprachigen Zielgruppe ist sehr verlockend für sie.

WARUM IST EUER PRODUKT ZUKUNFTSSICHER?

Die Welt wird immer komplexer und immer vernetzter. Die Menschen sind zunehmend mit globalen Ereignissen und Themen konfrontiert, die sie früher möglicherweise nicht direkt betroffen haben. Was zwischen China und Taiwan passiert, betrifft mich plötzlich. Da steigt der Bedarf an hochwertigen Informationen, Zusammenhängen und Hintergrundwissen. Slow Journalism statt Breaking News ist ein Trend, der uns begünstigt.

VIEL VIDEO IN WENIG ZEIT DANK SYNTHETISCHER AVATARE

NAME	ANYMATE ME
FÖRDERUNG	MEDIA STARTUP FELLOWSHIP #10
WEBSEITE	ANYMATEME.COM
GRÜNDER	JULIA LEDUC

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Anymate Me ist ein No-Code Web Studio zur automatisierten Produktion von synthetischen Videos durch Texteingabe (Text-to-Video). Unsere Lösung bietet eine Alternative zur zeitaufwändigen und kostspieligen traditionellen Produktion von Unternehmensvideos.

WAS MACHT EUER PRODUKT EINZIGARTIG?

Unsere realistischen Avatare verfügen über hochakurate Lippsynchronität für die deutsche Sprache, was für unseren Zielmarkt in der DACH-Region besonders relevant ist. Zudem können unsere Avatare Bewegungen abbilden und als KI-Gebärdensprachdolmetscher*innen für barrierefreie Videoinhalte fungieren.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Zunächst importiert man ein Dokument wie eine PPT oder ein PDF mit dem Unternehmensdesign. Das Design wird als Videohintergrund übernommen. Danach gibt man das Skript ein, das von einem Avatar vorgetragen werden soll. Dabei stehen mehr als 65 Sprachen zur Verfügung. Nach wenigen Minuten kann das Video heruntergeladen werden.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Unsere Lösung beruht auf der Nutzung mehrerer Machine-Learning-Technologien in den Bereichen Bild- und Sprachverarbeitung. Beispielsweise haben wir ein Bildverarbeitungsmodell, vergleichbar zu StyleGAN von Nvidia, erstellt und trainiert.

Für die Audios basierend auf der Texteingabe setzen wir Text-to-Speech-Technologie ein. Hier haben wir mit verschiedenen Open-Source-Modellen wie SV2TTS, experimentiert, um unser Modell zu trainieren. Für die Sprachen, die wir noch nicht abdecken, greifen wir auch auf Drittanbieter-Lösungen wie Microsoft TTS zurück.

WELCHE HÜRDEN GAB ES BEI DER (TECHNISCHEN) UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Generative AI ist ein schnell wachsendes Umfeld. Die kontinuierliche Forschung und Entwicklung verursachen auch bei uns einen hohen Aufwand und hohe Kosten. Zudem hat unsere Lösung einen starken Modellcharakter mit vielen denkbaren Nutzungsszenarien.

Das macht die Vermarktung besonders herausfordernd. Als Startup war es nicht einfach, auf die unterschiedlichen Kundenanfragen aus verschiedenen Branchen einzugehen und uns auf die für uns "idealen Kundschaft" festzulegen.

IST DAS ZUKUNFTSSICHER?

Aktuell sind fast alle unsere Kundschaft durch Inbound Sales gekommen, und wir haben sogar Kundenanfragen aus dem Ausland, mit denen wir nicht gerechnet hatten. Daher haben wir ein gutes Gefühl, dass wir uns in die richtige Richtung entwickelt haben.

Allerdings können wir nicht behaupten, dass unser Geschäftsmodell vollkommen zukunftsicher ist, da jederzeit ein Konzern oder ein neuer Player unseren Stand der Technik erreichen könnte. Wir müssen auf jeden Fall weiterhin aktiv bleiben und unsere Lösung weiterentwickeln!

Unsere realistischen Avatare verfügen über hochakurate Lippsynchronität für die deutsche Sprache. Zudem können unsere Avatare Bewegungen abbilden und als KI-Gebärdensprachdolmetscher*innen für barrierefreie Videoinhalte fungieren.

BESSER ARBEITEN MIT, *STATT* GEGEN KI

Der Einsatz von Machine Learning führt nur in Ausnahmen zur reinen Automatisierung. Stattdessen gibt es zahllose Tools, die den Alltag von Medienschaffenden im Hintergrund effizienter und stressfreier machen können. Das Ziel: Mehr Zeit für das Wesentliche.

Schlagworte vergeben nervt, egal, ob man Texte, Bilder oder Audios vertaggen will. Wenn man konsequent und fleißig ist und diese Arbeit macht, während das Material gerade frisch entstanden ist, mag es vielleicht noch gehen. Aber wer freut sich schon darauf, tausende Stunden Archivmaterial zu sichten, nur um sie am Ende mit ein paar Dutzend detaillierter Schlagworten zu verbinden?

Zum Glück gibt es heute Tools, die diese Arbeit übernehmen können.—gestützt durch KI. Speechcatcher und Open Recommendation Butler aus dem Media Tech Lab sind Beispiele für Tools, die das im Audibereich übernehmen können. Language Processing und Inhaltsanalysen sind die Methoden, die das möglich machen.

Denn eins ist klar: Wer die eigenen Inhalte für KI-Anwendungen, egal welcher Art nutzen will, braucht eine vernünftige Datenbasis — und dazu zählen eben auch die Metadaten. Wie gut, dass die nicht mehr ausschließlich von Menschen vergeben werden müssen!

Ein weiteres Tool, um eine gute Datenbasis erst aufzubauen, ist Fusionbase aus dem Media Startup Fellowship. Die Founder machen nicht nur unstrukturierte Daten aus öffentlichen Quellen maschinenlesbar und damit verarbeitbar, ihre KI ermöglicht es auch, Datensätze nach Anwendungsmöglichkeiten zu durchsuchen. Auch hier wird die anstrengendste Arbeit, das Sichten von zu vielen Informationen, durch künstliche Intelligenz unterstützt.

Varia unterstützt Recherchen mit Monitoring und KI-gestützten Zusammenfassungen. Und Lukas Kohlenbach, Gründer von Mediapod, hat ChatGPT mithilfe ein paar simpler Prompts dazu gebracht, ihn beim Faktencheck medizinischer Informationen zu unterstützen.

All diesen Anbietern ist vor allem eins wichtig: Medienschaffende und Journalist:innen sollen nicht ersetzt werden. Sie sollen ihr Handwerk ausüben, kreativ sein und gute, wichtige Arbeit machen, aufbauend auf einer modernen Infrastruktur. Vor allem aber sollen sie mehr Zeit haben, es zu tun—und dafür auf aufwändige, aber notwendige Aufgaben effizienter erledigen können.



KI-SUCHMASCHINE + TRANSKRIPTION = DURCHSUCHBARER PODCAST

ZWEI KI-ANWENDUNGEN, KLUG KOMBINIERT, ERGEBEN EIN KOMPLETT NEUEN USE-CASE—DER AUCH FÜR MEDIENHÄUSER UND NUTZER EINEN ECHTEN MEHRWERT BIETET.

WIE ES DER PODCAST DER „ONLINE MARKETING ROCKSTARS“ ZEIGT.

NAME	SPEECHCATCHER
LAUNCH	2023
FÖRDERUNG	MEDIA TECH LAB
WEBSEITE	GITHUB.COM/SPEECHCATCHER-ASR
ENTWICKLER	DR. BENJAMIN MILDE

NAME	OPEN RECOMMENDATION BUTLER
LAUNCH	2023
FÖRDERUNG	MEDIA TECH LAB
WEBSEITE	GITHUB.COM/OPEN-RECOMMENDATION-BUTLER
ENTWICKLER	MATTHIAS MEYER

Seit 2022 bietet das Media Lab Bayern ein Programm an, das sich speziell an Open-Source-Entwickler:innen richtet: Das Media Tech Lab. Im Batch 2 setzten alle drei Projekte KI ein: Snip Aid, Open Recommendation Butler und Speechcatcher. Die beiden letzteren wurden sogar kombiniert, um einen ganz neuen Use Case zu gestalten: Podcasts durchsuchbar zu machen. Doch der Reihe nach.

OPEN RECOMMENDATION BUTLER: KI-GESTÜTZTE SUCHMASCHINE FÜR MEDIENHÄUSER

Die interne Suche lässt auf vielen Websites—auch von Verlagshäusern—zu wünschen übrig. Sie liefert oft nur Treffer, die exakt dem Suchbegriff entsprechen. Nicht selten kommt dabei das Suchmaschinenprogramm Elasticsearch zum Einsatz. Informatiker Matthias Meyer wollte die Suche besser machen—speziell für Medienhäuser. Also hat er Elastic Search um Googles Sprachmodell BERT erweitert. BERT ist keine generative KI, eignet sich also nicht, um Texte zu schreiben. Stattdessen wurde BERT für Natural Language Understanding entwickelt. So ermöglicht das Modell eine semantische Suche. Wer zum Beispiel nach „Bundeskanzler“ sucht, bekommt nicht nur Treffer mit dem Wort Bundeskanzler, sondern auch Artikel, in denen es um Olaf Scholz geht. Außerdem

ermöglicht das Modell das so genannte „Topic Modeling“, bei dem Treffer zu einem Suchbegriff thematisch gruppiert werden. Wer „Scholz“ sucht, bekommt Treffer zur Sommer-Pressekonferenz des Kanzlers ebenso gebündelt angezeigt wie Treffer zu Scholz’ Äußerungen zum Ehegattensplitting.

Meyer nahm anschließend ein Feintuning des BERT-Modells vor: Er fütterte es mit einem im Netz verfügbaren Trainingsdatensatz, der 1000 Artikel enthält: „Das Training besteht darin, dass man zwei beliebige Artikel herausgreift und dem Modell sagt, ob die Artikel ähnlich sind oder nicht. Zwei Artikel zur Nato sind ähnlich, ein Artikel zu Nato und einer zu E-Autos sind nicht ähnlich.“ Jeder Artikel durchläuft das Sprachmodell und wird in einen Textvektor umgewandelt. Wenn zwei Textvektoren ähnlich sind, dann sind auch die Artikel ähnlich. „So ein Feintuning kann im Prinzip jeder machen, es würde zum Beispiel für ein auf Technologie-Themen spezialisiertes Medienhaus Sinn ergeben“, sagt Meyer.

Elasticsearch plus angepasstem Sprachmodell BERT ergibt also den Open Recommendation Butler. Das Softwarepaket packte Meyer in einen Docker-Container, der den Installationsprozess der verschiedenen Komponenten vereinfacht.

SPEECHCATCHER: DAS SPRACHERKENNUNGS-MODELL, DAS EIN ASS IN DEUTSCH IST

Parallel arbeitete Benjamin Milde, Doktor der Informatik, an einem Tool, das Audio- und Videodateien auf deutsch übersetzen, untertiteln und transkribieren kann: Speechcatcher. Als er im Oktober 2022 im Media Tech Lab anfang, hatte Open AI gerade sein großes Spracherkennungsmodell Whisper veröffentlicht—und zwar als Open Source. Whisper erzielte zwar hervorragende Transkriptionsergebnisse, funktionierte auf englisch aber besser als auf deutsch. Weitere Nachteile: Es konnte nicht in Echtzeit transkribieren und war (und ist) so ressourcenhungrig, dass es nur auf leistungsfähigen GPUs schnell läuft.

Milde schrieb ein Tool, um 30.000 Stunden Audio-material aus deutschen Podcasts herunterzuladen. 26.000 Stunden ließ er von Whisper transkribieren. Die dafür nötige Rechenpower organisierte er sich auf der Plattform vast.ai, wo man sich überschüssige Rechenkapazität mieten kann. „Die Transkripte waren nicht perfekt, aber sehr sehr gut“, erinnert sich Milde. „Manchmal hat Whisper halluziniert, das haben wir dann aus den Beispieldaten rausgeworfen.“

Mit den Transkripten aus Whisper trainierte er mehrere Wochen lang die Spracherkennungs-Engine aus dem Sprachverarbeitungs-Toolkit Espnet. Das Resultat: vier deutsche Spracherkennungsmodelle in den Größen S, M, L und XL. Alle sind viel kleiner als Whisper und lassen sich auch auf weniger leistungsstarken Rechnern installieren. Weil Speechcatcher einen anderen Modellierungsansatz benutzt, bei dem die Latenz geringer ist, eignen sich die Speechcatcher-Modelle auch für die Live-Transkription von Audios und Videos. „Ich bin super happy, Speechcatcher ist das beste Spracherkennungsmodell für Deutsch, das ich bisher trainiert habe“, freut sich Milde.



Wenn man nun einen Suchbegriff eingibt, spuckt der Open Recommendation Butler alle Podcast-Episoden aus, die diesen Begriff enthalten. Der Clou: Für jede Folge werden alle Zeitstempel angezeigt, an denen der gesuchte Begriff vorkommt.

DER SYNERGIE-EFFEKT: DURCHSUCHBARE PODCASTS

Im Verlauf des Projektes stellten er und Meyer fest, dass es gute Synergie-Effekte zwischen beiden Tools geben könnte: „Wenn ich meine Transkripte in den Open Recommendation Butler lade, könnte man eine Art Suchmaschine draus machen“, erinnert sich Milde. Im gemeinsamen Gespräch mit Erkan Kasap, CTO im Media Tech Lab, und Tech-Coach Martin Brüggemann entstand dann die Idee, ein Tool zu entwickeln, das Podcasts durchsuchbar macht. Brüggemann stellte den Kontakt zur Marketing-Plattform OMR her, die auch einen eigenen Podcast anbietet.

Unter den 30.000 Stunden Podcast-Material, die Benjamin Milde heruntergeladen hatte, befand sich praktischerweise auch der OMR-Podcast. Milde stellte die Transkripte davon Meyer zur Verfügung, der sie in seinen Open Recommendation Butler einspeiste. Wenn man nun einen Suchbegriff eingibt, spuckt der Open Recommendation Butler alle Podcast-Episoden aus, die diesen Begriff enthalten. Der Clou: Für jede Folge werden alle Zeitstempel angezeigt, an denen der gesuchte Begriff vorkommt. Man muss nur einmal auf den Play-Button drücken und die Wiedergabe startet an der jeweiligen Sprungmarke.

Noch befindet sich diese Anwendung in einer OMR-Testumgebung, könnte aber jederzeit live geschaltet werden. „Die OMR-Podcast-Suche zeigt, welcher Mehrwert entstehen kann, wenn man zwei KI-Tools geschickt kombiniert“, sagt Erkan Kasap.

Da es sich bei Speechcatcher und Open Recommendation Butler um Open-Source-Tools handelt, können alle Interessierte sie verwenden und auf ihre Bedürfnisse anpassen. Übrigens: Auch für Videos würde diese Lösung funktionieren.



Program-Manager Erkan Kasap moderiert das Media Tech Lab Meetup

ONLINE-RECHERCHE MIT KI-UNTERSTÜTZUNG

NAME	VARIA RESEARCH
GRÜNDUNG	2018
FÖRDERUNG	MEDIA COMPANY FELLOWSHIP, BATCH #5
WEBSEITE	VARIA.MEDIA
GRÜNDER	GEORG HORN, STEFAN HENSEL

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Wir ermöglichen mit unserem Produkt **Varia Research** eine effiziente online Recherche für Journalist:innen und alle anderen, die viel online unterwegs sind und recherchieren.

WAS MACHT EUER PRODUKT EINZIGARTIG?

Wir haben uns den journalistischen Online-Recherche-Prozess als Ganzes angeschaut und wollen ihn auch als Ganzes bedienen und optimieren. Das unterscheidet und von anderen Lösungen, welche meist nur auf einen Aspekt (e.g. Note Taking, File-Organisation, oder News Monitoring) fokussieren.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Varia Research besteht aus zwei Hauptkomponenten, die dem Nutzenden in einer SaaS-Applikation zur Verfügung stehen: Media Monitoring & Recherche-Organisation. Wer mit Varia Research arbeitet, kann den für die jeweilige Recherche relevanten News-Quellen über benutzerdefinierte Feeds folgen. Man kann aber dann auch gefundene, interessante Inhalte mit einem Mausklick in Dossiers strukturiert organisieren.

Damit wird dem Kopieren von Links in Word Dateien ein Ende gesetzt, alle Informationen (Links, Notizen, Dateien, Kontakte, etc.) können an einem Ort gefunden werden. Außerdem werden alle abgespeicherten Artikel-Inhalte mittels Machine Learning automatisiert analysiert und zusammengefasst. So kann bei der Recherche viel Zeit gespart werden.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Grundsätzlich ist Varia Research eine SaaS-Applikation und läuft auf Servern in Deutschland. Wir haben bei der Umsetzung Datenschutz und Sicherheit von Anfang an priorisiert, da das für unsere wichtigste Zielgruppen von essenzieller Bedeutung ist.

Die Machine-Learning-Komponenten kommen teilweise aus unserem Haus, teilweise basieren sie auch auf externen APIs. Insbesondere Microsoft Bing, Google und OpenAI spielen auch bei uns eine immer wichtigere Rolle. Für NLP-Aufgaben wie Sentiment-Analyse, Entity Recognition, oder automatisierte Zusammenfassungen nutzten wir bisher ausschließlich In-House-Ansätze, die auf einem Large Language Model und einer Transformer Architektur basieren – ganz ähnlich wie GPT-X, nur ohne RLHF (Reinforcement Learning Through Human Feedback).

Wir kennen also die LLM-Thematik bereits lange, haben uns auch vor der nun dominierenden Transformer-Modell-Architektur mit verschiedensten Ansätzen der Vektorisierung von Text auseinandergesetzt. Für die Zukunft setzen wir stärker auf APIs, damit unsere Lösung modularer, damit flexibler und auch besser skalierbar bleibt – sowohl im Hinblick auf mehr Nutzenden als auch auf mehr Sprachen, die wir abbilden wollen. Aktuell sind das Deutsch und Englisch, weitere sollen folgen.

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Hürden gibt es kommerzieller und technischer Natur. Es ist nicht leicht, ein auf Journalismus ausgerichtetes Produkt finanzieren zu können. Außer ein oder zwei Acceleratoren und Fördertöpfen gibt es leider wenig institutionelle oder private Kapitalgeber, die den Sektor spannend bzw. attraktiv genug finden. Auf technischer Seite war vielleicht unser Timing etwas ungünstlich.

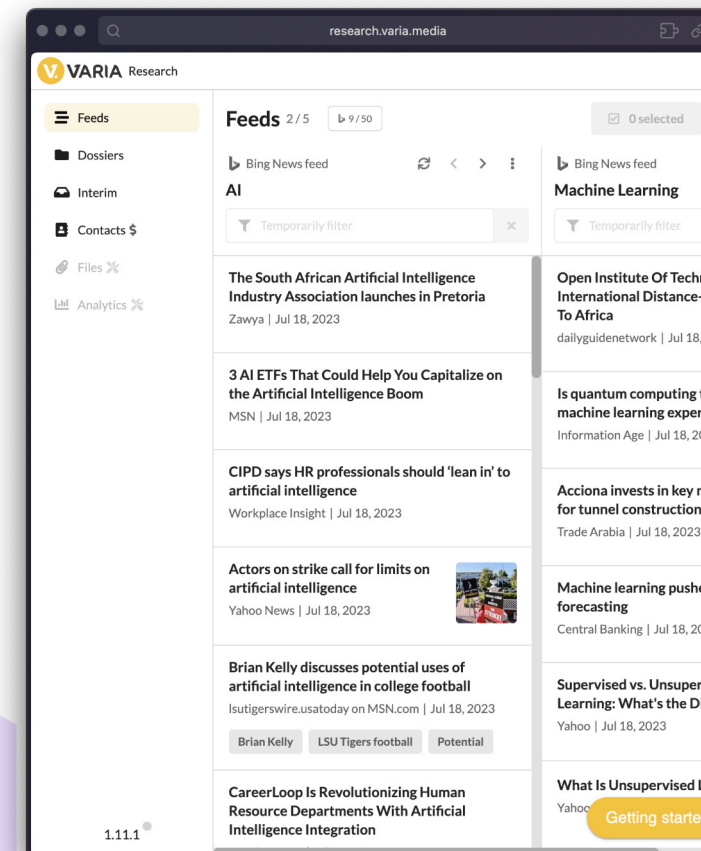
Als wir 2018 losgelaufen sind – mit einer initialen, anderen Produktidee – waren NLP-Modelle noch auf einem viel niedrigerem Niveau und die verfügbaren APIs deutlich eingeschränkter als heute. Wir haben daher relativ viel in eigene NLP-Kompetenz investiert.

Wenn wir auf dem Stand des heutigen NLP-Ökosystems starten würden, wären wir ziemlich sicher einen anderen Weg gegangen.

WARUM IST EUER PRODUKT ZUKUNFTSSICHER?

Recherchieren wird man immer. Teile der Arbeit lassen sich automatisieren, aber um eine strukturierte Organisation rund um eine Recherche kommt man bei aller Automatisierung nicht herum. Wir wollen auf dem Weg zur Teil-Automatisierung der Recherche eine entscheidende Rolle spielen.

Bild: Varia Research



AUDIOINDEX

DAS TRANSKRIPTIONS-TOOL, DAS DIALEKT VERSTEHT

NAME	AUDIOINDEX
GRÜNDUNG	2020
FÖRDERUNG	OPEN INNOVATION CHALLENGE 2022
WEBSEITE	AUDIOINDEX.AI
GRÜNDER	JOHANNES BUSCHING

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Mit AudioIndex lassen sich Audio und Video automatisiert verschriftlichen, und zwar in höchster Qualität. Die Transkription geht innerhalb von Minuten, die manuelle Transkription dauert oft Stunden.

WAS MACHT EUER PRODUKT EINZIGARTIG?

Die höhere Qualität bei der Transkription von deutschsprachigem Material. AudioIndex beherrscht im Gegensatz zu anderen Transkriptions-Tools auch deutsche und österreichische Dialekte sehr gut, und kommt auch mit Fachvokabeln gut klar. In der Politik gibt es ja zum Beispiel ein sehr spezifisches Vokabular. Außerdem legen wir einen Fokus auf Datenschutz. Alle Daten liegen auf unseren Servern in Österreich, Deutschland und anderen EU-Ländern, wir geben keine Daten an andere Unternehmen und in andere Länder wie die USA weiter.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Man loggt sich ein und lädt seine Audio- oder Videodateien hoch. Dann startet der Transkriptionsprozess automatisch, er dauert ungefähr ein Zehntel der Aufnahmezeit. Anschließend erscheint die Transkription in Textform. Das Ergebnis kann man einfach exportieren. Wir bieten auch eine API an, mit der zum Beispiel Redaktionen die Transkripte direkt in ihr CMS importieren können.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Wir bauen auf so ziemlich allen Open-Source-Spracherkennungs-Modellen und (Trainings-)Datensätzen auf, die es gibt und die wir (lizenzrechtlich) verwenden dürfen. Beim Nachtrainieren kommen zwei Methoden zum Einsatz. Zum einen das Reinforcement Learning: Nutzer:innen sagen dem System, was falsch war und korrigieren es direkt. So lernt es für die Zukunft. Zum anderen verwenden wir eigene Datensätze für deutsche und österreichische Dialekte und Vokabeln für gewisse Domänen.

Zudem helfen uns sogenannte „Knowledge Graphs“. Vorstellen kann man sich das etwa als riesige „Wortwolke“ zur jeweiligen Branche, in der alle Fachbegriffe und Konzepte miteinander verknüpft sind. Damit können wir falsch erkannte Begriffe auf der Textebene erkennen und verbessern.

In unseren Trainingsdatensätzen müssen wir gesprochenes und geschriebenes Wort synchronisieren, das nennt man „Forced Alignment“. Das gesprochene Wort wird in Form eines Spektrogramms dargestellt. Um es mit dem passenden Wort im Text zu synchronisieren, braucht man auf Wortebene Zeitstempel. Die werden während der Spracherkennung generiert.

**Alle Daten liegen auf
unseren Servern in
Österreich, Deutschland
und anderen EU-Ländern,
wir geben keine Daten an
andere Unternehmen und
in andere Länder wie die
USA weiter.**

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Hürden gibt es einige im Laufe des Aufbaus eines Startups. Manche (technische) Prozesse brauchen einfach ihre Zeit und es braucht das richtige Maß an Durchhaltevermögen und Flexibilität/Anpassung. Auch wenn man mit einem technischen Hintergrund an die Sache rangeht, darf man nie vergessen, früh mit Kundenfeedback zu arbeiten bzw. darauf zu achten, was die Leute wirklich brauchen. Zudem es ist wichtig, Kontakte in die Industrie, Forschung und zu potentiellen Kundschaft aufzubauen. Am Besten so früh wie möglich.

WARUM IST EUER PRODUKT ZUKUNFTSSICHER?

Spracherkennungsmodelle werden sich von den meisten Anbietern alle auf einem gewissen hohen Level einpendeln, es gibt also einen "qualitativen Grundstock", aber der Mehrwert für die Anwender kommt zunehmend über "die letzten paar Prozent", die in ihrer jeweiligen Domäne/Branche richtig erkannt werden. Und diesen Mehrwert liefert AudioIndex für deutschsprachige Audios und Videos. Daneben geht es um Zusatzfeatures, die den Kundschaft die Arbeit erleichtern.

Wir bieten mit AudioIndex bspw. automatisierte Zusammenfassungen, oder Metadaten (Kategorien, Tags, Erkennung von Personen, etc.) zur Personalisierung. So können automatisiert die passenden, im Medienhaus vorhandenen Audioinhalte zu den Texten angezeigt werden, die sich eine Nutzer:in auf der Website angesehen hat. Wer z. B. einen Artikel zum Heizungsgesetz gelesen hat, bekommt dazu passende Podcasts oder Interviews vorgeschlagen. In der Personalisierung von Inhalten und dem Wiederverwerten von bestehendem Material sehe ich eine große Zukunft.

BESSERE DATEN FÜR BESSEREN JOURNALISMUS



NAME	FUSIONBASE
GRÜNDUNG	2019
FÖRDERUNG	MEDIA COMPANY FELLOWSHIP #7, GROW FELLOWSHIP
WEBSEITE	FUSIONBASE.COM
GRÜNDER:INNEN	KEVIN GOSSLING & PATRICK HOLL

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Wir nehmen Public Data aus verschiedenen Quellen, tragen sie zusammen und machen sie maschinenlesbar. Fusionbase ist damit ein Data Hub für Daten mit verschiedenen Content-Features. Unser Ziel ist es, die Arbeit mit Daten für Datenexpert:innen so einfach wie möglich zu machen.

WAS MACHT EUER PRODUKT EINZIGARTIG?

Öffentliche Daten werden von Anbietern wie Behörden oder Regierungen oft als Excel-Tabellen, PDF-Dateien oder in anderen, unstrukturierten Formen angeboten. Diese Daten sind nur unter enormem Aufwand nutzbar, wenn man sie analysieren oder für eigene Modelle benutzen will. Wir machen diese Daten maschinenlesbar. Damit gestalten wir nicht nur deren Auffindbarkeit, sondern auch deren Einsatz so komfortabel wie möglich.

Öffentliche Daten werden von Anbietern wie Behörden oder Regierungen oft als Excel-Tabellen, PDF-Dateien oder in anderen, unstrukturierten Formen angeboten. Diese Daten sind nur unter enormem Aufwand nutzbar, wenn man sie analysieren will oder für eigene Modelle benutzen will.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Fusionbase ist erstmal eine Web-App, in der man nach Datensätzen suchen kann, genauso wie man auf Amazon nach Produkten sucht. Dabei setzen wir auch auf eine KI-gestützte, semantische Analyse, so dass man nicht nur nach Datensätzen direkt, sondern auch nach Use Cases suchen kann. Man bekommt dann Datensätze vorgeschlagen, die zu diesem Suchbegriff passen könnten. Zusätzlich bieten wir drumherum noch eine Python-Anwendung, die es ermöglicht, gefundenen Daten auch direkt in den eigenen Code zu integrieren. Das ist aber eher etwas für Data Scientists.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Wir generieren aus allen Datensatz-Eigenschaften—Texten, Beschreibungen, Namen und so weiter—so genannte Embeddings. Ein Embedding ist die Repräsentation eines Textes als Vektor-Zahl. Suchbegriffe werden ebenfalls in Embeddings umgewandelt und aus diesen beiden Zahlen lassen sich dann Wahrscheinlichkeiten errechnen, wie wahrscheinlich zwei Embeddings zueinander passen könnten. Daraus ergeben sich die Suchergebnisse.

Das funktioniert unter anderem mit Open-Source-Modellen wie Bert oder T5 von Google oder aktuellen Large Language Modellen wie Falcon, die wir dann aber auf unseren Fall anpassen. Jedes KI-Modell arbeitet mit Embeddings, aber nicht alle eignen sich dann für eine semantische Suche. In unserem Fall funktioniert besonders SentenceBert sehr gut.

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Eine große Hürde sind die Kosten für die Hardware, die man braucht, um überhaupt eigene Modelle zu trainieren. Für eigene GPUs kann man schnell Zehntausende Euro ausgeben, und selbst die Mieten bei Cloud-Anbietern wie AWS sind teuer. Und dann braucht man auch noch eine solide Datenbasis, wenn man ein eigenes Modell nicht nur im Detail anpassen, sondern wirklich selbst bauen möchte. Da mussten wir erstmal investieren.

Das Finetuning für unseren Use Case, wenn also ein Foundation Model—ein großes und kommerziell nutzbares Machine Learning Modell—schon existiert, ist dann aber oft einfacher als gedacht. Das Tool Transformers von Hugging Face hat inzwischen die Hürde für den Einsatz des Tools zum Beispiel deutlich gesenkt.

IST DAS ZUKUNFTSSICHER?

Realistischerweise liefert Finetuning und Anpassung auf einzelne Use Cases im Moment noch die besseren Ergebnisse—wenn man sich aber die Entwicklung von GPT4 oder Open-Source-Modellen wie Falcon anschaut, wird es am Ende in die Richtung „One size fits all“ gehen und nicht mehr viel Finetuning brauchen.

Für Fusionbase ist das aber kein Problem, wir würden uns sogar wünschen, dass die Tools generischer und einfacher in der Anwendung werden. Wir stellen weiterhin maschinenlesbare Daten zur Verfügung, die andere nicht haben—und ohne gute Daten bekommt man keine guten Ergebnisse, egal, in welches Language Model man sie füttert.

LUMEDIO

MIT WENIG AUFWAND ZUM EIGENEN KI-HELFERLEIN

NAME LUMEDIO
FÖRDERUNG MEDIA STARTUP FELLOWSHIP #12
GRÜNDER LUKAS KOHLBACH & ALEXANDER RASSBACH
Lumedio wurde zudem vom Innovationsfond der Wissenschaftspreskonferenz (WPK) unterstützt

WER SEID IHR UND WELCHES PROBLEM LÖST IHR?

Aus einem studentischen Podcast-Projekt heraus haben wir im Rahmen des WPK Innovationsfonds und des Media Startup Fellowships eine Plattform entwickelt, die Medizin-Podcasts kuratiert, Fakten checkt und die Qualität der Podcasts mit einem Ampelsystem bewertet. **Weil die Überprüfung durch Expert:innen viel Zeit kostet, haben wir damit experimentiert, wie wir KI einsetzen können, um uns die Arbeit zu erleichtern.**

WAS MACHT EUER PRODUKT EINZIGARTIG?

Unsere Plattform besetzt unserer Meinung nach noch eine Nische: Es gibt zwar bereits verschiedene „Siegel“ oder Hinweise für medizinische Inhalte, etwa im Rahmen von YouTube Health. Dort werden aber nur einmal die Macher:innen und nicht jede einzelne Folge auf Seriosität überprüft.

GROB SKIZZIERT: WIE FUNKTIONIERT EUER TOOL?

Wir haben für unsere Experimente GPT-4 genutzt. Das füttern wir dann über die Weboberfläche ChatGPT mit den Transkripten einer zu überprüfenden Podcastfolge (die auch von einer KI erstellt werden, nämlich von Trint). Man kann ChatGPT den Befehl geben, die gerade geposteten Inhalte für die Dauer des Chats zu „lernen“. Und dann haben wir verschiedene Prompts entwickelt, die uns eine Tabelle mit allen medizinischen Aussagen im Transkript liefern, und dann auch direkt einen Hinweis abgeben, ob diese Stelle möglicherweise kritisch sein könnte.

WIE HABT IHR DAS TECHNISCH UMGESETZT?

Zuerst haben wir experimentiert, ob unsere Prompts auf englisch oder deutsch besser sind, und haben festgestellt, dass deutsche Prompts und deutsche Inhalte gut zusammen funktionieren. Dann haben wir ChatGPT die Transkripte „lernen“ lassen, das geht einfach mit einem Prompt wie „Lerne den folgenden Text“. Nicht zuletzt haben wir begonnen, nach medizinischen Aussagen in den Texten zu fragen, und unsere Fragestellungen immer weiter angepasst.

Das geht übrigens auch super mit ChatGPT: Man kann der KI einfach einen Prompt zeigen und sie danach fragen, wie man ihn anpassen müsste, um für die KI selbst noch verständlicher zu sein. Und so werden die Prompts dann besser und besser. Man landet dann am Ende bei lediglich ein paar Zeilen.

WELCHE HÜRDEN GAB ES BEI DER UMSETZUNG? WAS GING LEICHTER ALS GEDACHT?

Wir haben das Ganze bisher nur in wenigen Fällen ausprobiert, und zwar in Fällen, die vorher schon komplett von unseren Gutachter:innen überprüft wurden. Und bei dieser kleinen Stichprobe können wir sagen: Die Technik funktioniert besser als gedacht. ChatGPT hat mit Hilfe unserer Prompts dieselben Stellen als „medizinische Aussagen“ identifiziert wie unsere Gutachter:innen, und auch dieselben Stellen als überprüfenswert eingestuft.

Die Hürde kam eher dadurch, dass man über die Weboberfläche nur eine begrenzte Zeichenzahl zum „lernen“ an ChatGPT geben kann. Das ist bei langen Transkripten etwas nervig, ließe sich aber schon durch den Einsatz der offiziellen API umgehen.

Was ChatGPT aber nicht kann, ist eine gute von einer schlechten Studie zu unterscheiden. Dafür wird es auch in Zukunft Expert:

innen brauchen, die Sprachmodelle auf besondere Anwendungsfälle anpassen.

IST DAS ZUKUNFTSSICHER?

ChatGPT wird immer besser. Und es hat natürlich enorm viel Material zum trainieren bekommen. Was es aber nicht kann, ist zum Beispiel, eine gute von einer schlechten Studie zu unterscheiden. Dafür wird es auch in Zukunft Expert:innen brauchen, die Sprachmodelle auf besondere Anwendungsfälle anpassen. Wir denken für die Zukunft auch darüber nach, ChatGPT nur als eines von mehreren Sprachmodellen zu nutzen, selbst Modelle zu hosten und diese dann mit den Studien, die wir in der Recherche finden, anpassen.

MIT JEDEM EXPERIMENT EINEN SCHRITT WEITER



Wenn Unternehmen einen gewissen Organisationsgrad erreichen und eine gewisse Mitarbeiterzahl, sinkt die Geschwindigkeit von Projekten rapide.

Neben Startups und kleinen Projekt-Teams unterstützt das Media Lab Bayern auch etablierte Medienhäuser. Im Media Company Fellowship begleitete das Team über die letzten 12 Monate unter anderem den Fränkischen Tag bei einem Innovationsprojekt—bei dem auch KI eine Rolle spielte.

Wenn Unternehmen einen gewissen Organisationsgrad erreichen und eine gewisse Mitarbeiterzahl, sinkt die Geschwindigkeit von Projekten rapide. Sie entwickeln eine organisatorische Trägheit oder werden, wie es Paul Froemel, Redakteur bei der Mediengruppe Oberfranken, ausdrückt: „ein schwerfälliger Tanker.“ Das stellt Redaktionen vor eine Herausforderung: Wie können sie trotz dieser Schwerfälligkeit auf die Geschwindigkeit neuer Technologien reagieren?

DEN REDAKTIONELLEN ZEITFRESSERN AUF DER SPUR

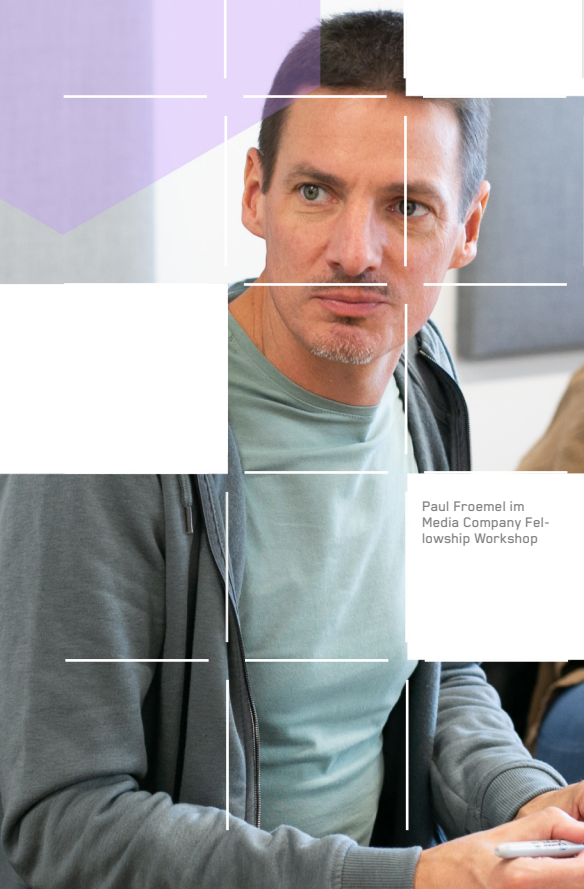
Um hier zu helfen, hat das Media Lab Bayern 2021 das Media Company Fellowship ins Leben gerufen, ein Förderprojekt, das Fortbildung und Projektarbeit vereint. Ziel ist es, die nötigen Fähigkeiten für gute Innovationsarbeit fest in Medienhäusern zu verankern. Dazu gehören unter anderem die Recherche nach Innen und nach Außen, ein solides Verständnis der eigenen Organisation, diplomatisches Geschick, eine Portion Pragmatismus und die Lust, Dinge einfach einmal auszuprobieren.

Für die zweite Runde des Programms bewarb sich das Team des Fränkischen Tags aus Paul Froemel, Maresa Schlemmer und Ellen Mützel mit einem bekannten Problem: Wie machen wir unsere 2.000-köpfige Organisation transparenter? „Neue Mitarbeiter wissen nicht, wer für was zuständig ist“, skizziert Froemel die Herausforderung. Die Konsequenz: Lange Fragerunden, die schließlich bis ins Management laufen und so wertvolle Zeit kosten. Zeit, die gerade in einer Zeitungsredaktion wortwörtlich Geld ist.

Nach mehreren Interviews und einer ausführlichen Betrachtung des Problems entschied sich das Team schließlich für eine interne Projektübersicht, inkl. Organigramm. Wenn man das Unternehmen schon nicht übersichtlicher machen kann, dann vielleicht ja lesbar—so zumindest die Hoffnung hinter dem ersten Prototypen.

Dennoch ließ das Team trotz des ersten Erfolgs das Thema „Zeitersparnis“ nicht in Ruhe und so war ein zweiter Zeitfresser schnell identifiziert: Leserzuschriften.

Paul Froemel im
Media Company Fel-
lowship Workshop



Es ist nicht der Fall, dass die Zeitung keine Veranstaltungshinweise, Nachberichte, zusätzliche Bilder oder Tipps haben will—im Gegenteil. Das Problem ist vielmehr, dass diese an ganz verschiedenen Orten im Haus ankommen: Mal bei einzelnen Redakteur:innen, mal bei mehreren, mal bei der offiziellen Kontaktadresse. Eine unüberschaubare Flut an Informationen, die wertvoll ist, aber schwer in die Bahnen der Tagesproduktion zu lenken ist. Auch hier war der erste Ansatz verhältnismäßig simpel, aber umso effektiver: ein Uploadportal.

Eine zentrale Anlaufstelle, über die Vereine, Kirchen, Kommunen und Einzelpersonen schnell und einfach ihre Informationen strukturiert an die Redaktion schicken können. Andere Zeitungen wie die Mainpost aus Würzburg, der Neue Tag aus Weiden oder Schwäbisch Media haben hier bereits gute Erfahrungen gemacht: bis zu fünf Mal schneller sei die Produktion.

Und hier kam dem Team eine Idee: könnte KI nicht vielleicht sogar die ersten Texte aus diesen Informationen schreiben? Was, wenn sich der berüchtigte Text über das Jahrestreffen des Kaninchenzüchtervereins in Zukunft selbst schreibt? Ein Segen für jede:n Volontär:in!

KI wird bei uns im Haus relativ positiv gesehen. Die Angst, KI nimmt die Jobs weg, geht bei uns nicht um. Was wir aber gelernt haben ist uns auf zwei bis drei Dinge zu fokussieren und die richtig gut zu machen.



Was folgte, waren mehrere Wochen Experimente: „Wir haben angefangen, rumzuspielen, zu testen, zu gucken. Wir haben mit Resumer, Mindverse und anderen Tools experimentiert. Fürs Zusammenfassen waren Paraphraser und Shorty von der Schickler Unternehmensberatung gar nicht schlecht. Am meisten haben wir mit GPT 3.5 experimentiert“, fasst Froemel zusammen. Die Zuverlässigkeit ließ allerdings zu wünschen übrig: „Inhaltlich mussten wir sehr intensiv kontrollieren, wodurch die Zeitersparnis verloren ging. Die Systeme halluzinieren, erfinden Quellen, selbst bei einfachen Aufgaben wie der, einen Text zu kürzen.“

Die Ergebnisse seien zwar grundsätzlich nutzbar und klingen journalistisch, müssen jedoch Absatz für Absatz kontrolliert werden. Es ist schließlich peinlich, wenn die KI Regionalligaspieler Tore schießen lässt, die schlicht nicht existieren. Trotz vieler Experimente und Feinschliff blieben die Ergebnisse mittelmäßig.

KEINE ANGST UM JOBVERLUSTE

Das Uploadportal soll trotzdem kommen, aber mit weniger Automatisierung und mehr Handarbeit. Trotzdem hat das Team viel gelernt und die Experimente nicht aufgegeben: Überschriften entwerfen, das kann GPT zum Beispiel sehr gut. „Blattmacher müssen für Zeitungsseiten am Tag 20-30 Stück entwerfen. Da eine Schnittstelle ins CMS, rechter Mausklick, mach mir Überschriftenvorschläge. Das würde eine Menge Zeit sparen.“ Gleiches gilt für Ankündigungen, Werbetexte oder E-Mails.

„KI wird bei uns im Haus relativ positiv gesehen. Die Angst, KI nimmt die Jobs weg, geht bei uns nicht um. Was wir aber gelernt haben ist uns auf zwei bis drei Dinge zu fokussieren und die richtig gut zu machen,“ resümiert Froemel. Auch das ist ein Erfolg des Fellowships.

NAME	FRÄNKISCHE TAG
GRÜNDUNG	1946
FÖRDERUNG	MEDIA COMPANY FELLOWSHIP #2
WEBSEITE	FRAENKISCHERTAG.DE