



# Natural Language Generation: Wenn aus Daten Texte werden

Faktenpapier

Aus der Serie: AI: Science over Fiction

[www.bitkom.org](http://www.bitkom.org)

**bitkom**

## Herausgeber

Bitkom  
Bundesverband Informationswirtschaft,  
Telekommunikation und neue Medien e. V.  
Albrechtstraße 10 | 10117 Berlin  
T 030 27576-0  
bitkom@bitkom.org  
www.bitkom.org

## Verantwortliches Bitkom-Gremium

AK Artificial Intelligence

## Projektleitung

Dr. Nabil Alsabah | Bitkom e. V.

## Autoren

Kristin Strauch | Retresco GmbH  
Dr. Nabil Alsabah | Bitkom e. V.

## Lektorat

Linda van Rennings | Bitkom e. V.

## Satz & Layout

Katrin Krause | Bitkom e. V.

## Titelbild

© Tony Reid | unsplash.com

## Copyright

Bitkom 2019

Diese Publikation stellt eine allgemeine unverbindliche Information dar. Die Inhalte spiegeln die Auffassung im Bitkom zum Zeitpunkt der Veröffentlichung wider. Obwohl die Informationen mit größtmöglicher Sorgfalt erstellt wurden, besteht kein Anspruch auf sachliche Richtigkeit, Vollständigkeit und/oder Aktualität, insbesondere kann diese Publikation nicht den besonderen Umständen des Einzelfalles Rechnung tragen. Eine Verwendung liegt daher in der eigenen Verantwortung des Lesers. Jegliche Haftung wird ausgeschlossen. Alle Rechte, auch der auszugsweisen Vervielfältigung, liegen beim Bitkom.

## 1 NLG revolutioniert die Erstellung von Texten in zahlreichen Branchen

Digitalisierung und technologische Innovationen verändern die Art und Weise, wie wir arbeiten, leben und wertschöpfen grundlegend. Besondere Potenziale und Dynamiken ergeben sich hier durch die jüngsten Entwicklungen im Bereich der Künstlichen Intelligenz. Ob in Bild- oder Spracherkennungssystemen oder in der autonomen Fahrzeugtechnologie – der Einsatz von KI erzeugt ein Spannungsfeld, in dem Prozesse stets hinterfragt, adaptiert und neugestaltet werden müssen. Ein Teilgebiet von Künstlicher Intelligenz beschäftigt sich mit der automatischen Erzeugung natürlichsprachlicher Texte durch eine Software. Das nennt man Natural Language Generation, kurz NLG. Als Disziplin der Computerlinguistik wandelt NLG – auch bekannt als automatische Textgenerierung – Daten in Text um. Überall dort, wo große Mengen an Daten vorhanden sind oder generiert werden – etwa in der informationsjournalistischen Berichterstattung für Sport, Wirtschaft und Wetter – macht der Einsatz von NLG Sinn. Bereits jetzt setzen einige Nachrichtenportale und Medienverlage auf den sogenannten Roboterjournalismus. Auch hier wird im Backend nichts anderes verwendet als NLG-Software, die beispielsweise Sportnachrichten, Wetterberichte oder Börsenupdates produziert.

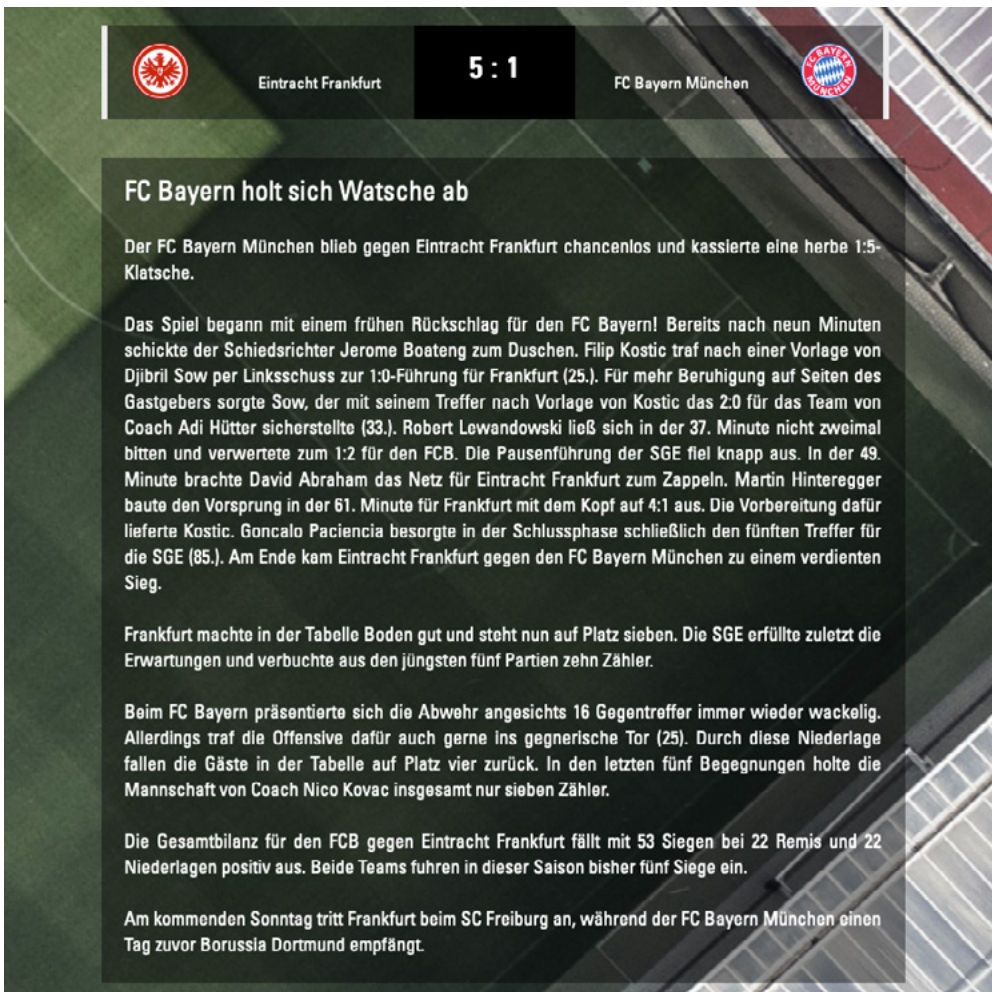
Die Vorteile liegen auf der Hand: Zum einen können mit automatisch erstellten Texten Inhalte schneller, effizienter und skalierbar erstellt werden, zum anderen können Nachrichten so im großen Stil personalisiert werden. Darüber hinaus erlaubt NLG beispielsweise im Journalismus die Erstellung von Nachrichten auf hyperlokaler Ebene – ein Bereich, der unter normalen Umständen für Verlage wirtschaftlich nicht rentabel wäre. Die Qualität von automatisiert generierten Texte im Bereich der medialen Massenproduktion schneidet nicht schlechter, sondern teilweise sogar besser ab als von Menschen erstellte Texte [1].

Im Bereich der Fußballberichterstattung entstehen auf diese Weise computergenerierte Texte, die sich in der Wahrnehmung der Leser nicht von Berichten unterscheiden lassen, die von einem (menschlichen) Redakteur verfasst wurden. Hier ein Beispiel:

*»Ob in Sport, Wirtschaft oder Wetter: Überall dort, wo viele Daten vorhanden sind, kann man mit KI automatisch hochwertige Texte generieren.«*







The image shows a digital match report for a football game between Eintracht Frankfurt and FC Bayern München. At the top, the score is 5:1 in favor of Eintracht Frankfurt. The report includes a title, a short summary, a detailed match description, and a closing paragraph about upcoming games.

**FC Bayern holt sich Watsche ab**

Der FC Bayern München blieb gegen Eintracht Frankfurt chancenlos und kassierte eine herbe 1:5-Klatsche.

Das Spiel begann mit einem frühen Rückschlag für den FC Bayern! Bereits nach neun Minuten schickte der Schiedsrichter Jerome Boeteng zum Duschen. Filip Kostic traf nach einer Vorlage von Djibril Sow per Linksschuss zur 1:0-Führung für Frankfurt (25.). Für mehr Beruhigung auf Seiten des Gastgebers sorgte Sow, der mit seinem Treffer nach Vorlage von Kostic das 2:0 für das Team von Coach Adi Hütter sicherstellte (33.). Robert Lewandowski ließ sich in der 37. Minute nicht zweimal bitten und verwertete zum 1:2 für den FCB. Die Pausenführung der SGE fiel knapp aus. In der 49. Minute brachte David Abraham das Netz für Eintracht Frankfurt zum Zappeln. Martin Hinteregger baute den Vorsprung in der 61. Minute für Frankfurt mit dem Kopf auf 4:1 aus. Die Vorbereitung dafür lieferte Kostic. Goncalo Paciencia besorgte in der Schlussphase schließlich den fünften Treffer für die SGE (85.). Am Ende kam Eintracht Frankfurt gegen den FC Bayern München zu einem verdienten Sieg.

Frankfurt machte in der Tabelle Boden gut und steht nun auf Platz sieben. Die SGE erfüllte zuletzt die Erwartungen und verbuchte aus den jüngsten fünf Partien zehn Zähler.

Beim FC Bayern präsentierte sich die Abwehr angesichts 16 Gegentreffer immer wieder wackelig. Allerdings traf die Offensive dafür auch gerne ins gegnerische Tor (25). Durch diese Niederlage fallen die Gäste in der Tabelle auf Platz vier zurück. In den letzten fünf Begegnungen holte die Mannschaft von Coach Nico Kovac insgesamt nur sieben Zähler.

Die Gesamtbilanz für den FCB gegen Eintracht Frankfurt fällt mit 53 Siegen bei 22 Remis und 22 Niederlagen positiv aus. Beide Teams führen in dieser Saison bisher fünf Siege ein.

Am kommenden Sonntag tritt Frankfurt beim SC Freiburg an, während der FC Bayern München einen Tag zuvor Borussia Dortmund empfängt.

Abbildung 1: Beispiel eines automatisch generierten Fußballtextes (Bildrechte: Retresco GmbH)

Jedoch kann NLG auch im E-Commerce zur automatisierten Erstellung von Produktbeschreibungen, in Unternehmen für interne Berichte oder zur Erstellung von Berichten an der Börse eingesetzt werden, um aus Daten leserfreundliche und informative Texte zu erstellen. In diesem Zusammenhang gilt NLG als Schlüssel zu mehr Effizienz: Die Fähigkeit, datenbasierte Analysen in natürlichsprachlichen und leicht verständlichen Text auszudrücken, macht NLG zur optimalen Schnittstelle zwischen Mensch und Maschine – und somit zum Schlüsselement in jedem Data-to-Insights-Prozess. Auf diese Weise unterstützt NLG Unternehmen dabei, optimalen Nutzen aus der gesamten Daten-Wertschöpfungskette zu ziehen. Es ist daher kaum verwunderlich, dass renommierte Marktforschungsunternehmen wie [Forbes](#) [2] und [Gartner](#) [3] Natural Language Generation als einen der spannendsten Trends der Digitalen Transformation bezeichnen.

*»Die Fähigkeit, datenbasierte Analysen in einen leicht verständlichen Text auszudrücken, macht NLG zur optimalen Schnittstelle zwischen Mensch und Daten.«*

## 2 NLG schafft überall dort Mehrwert, wo große Mengen an Daten in Text umgewandelt werden können

Gute Beispiele für den sinnvollen Einsatz von NLG-Systemen findet man vor allem in Bereichen, in denen große Mengen an Daten anfallen, NLG-Programme wandeln diese Daten dann Texte um. Auch wenn sämtliche Möglichkeiten längst nicht vollständig ausgelotet sind, bewähren sich einige konkrete Formate bereits in der Praxis:

- **Nachrichten:** Newsportale gehören zu den Vorreitern bei dem Einsatz von NLG. Automatisch generierte Beiträge eignen sich insbesondere für Textsorten mit einer hohen Dichte an standardisierten Formulierungen. Dazu gehören zum Beispiel die Berichterstattung für Sportarten, Börsennachrichten oder Wettervorhersagen.
- **Reportings:** Ob im Bereich Business Intelligence oder in der Kommunikation mit Privatkunden: ausformulierte Berichte lesen sich eingängiger als reine Zahlenkolonnen. Überall dort, wo große Datenmengen erhoben werden, kann Natural Language Generation helfen, Entwicklungen und Zahlenanalysen deskriptiv darzustellen und Informationen zu interpretieren. Ein Anwendungsbeispiel sind etwa die Fondsreports großer Finanzdienstleister für private Anleger.
- **Produktbeschreibungen:** Je mehr Produkte ein Online-Shop enthält, desto aufwendiger ist die redaktionelle Erstellung von einzigartigen Produktbeschreibungen. Automatische Textgenerierung sorgt für signifikante Effizienzsteigerungen im gesamten Prozess der Inhaltserstellung. Dort wo NLG im E-Commerce im Einsatz ist, lassen sich zudem positive Effekte auf Nutzererfahrung und SEO-Sichtbarkeit messen.
- **Dialogsysteme:** Besondere Stärke entfaltet Natural Language Generation in Kombination mit Natural Language Understanding (NLU). So vereinen beispielsweise Chatbots NLG und NLU: Dialogprogramme verstehen die Eingaben eines Nutzers und generieren sinnvolle Antworten. Über Maschinelles Lernen entwickelt sich die Präzision der natürlich generierten Aussagen fortlaufend weiter.

»Automatisch generierte Beiträge eignen sich insbesondere für Textsorten mit einer hohen Dichte an standardisierten Formulierungen.«

Kategorisiert man die oben erläuterten Narrationsformate in relevante Branchen, zeichnet sich eine große Bandbreite an Anwendungsbereichen ab:

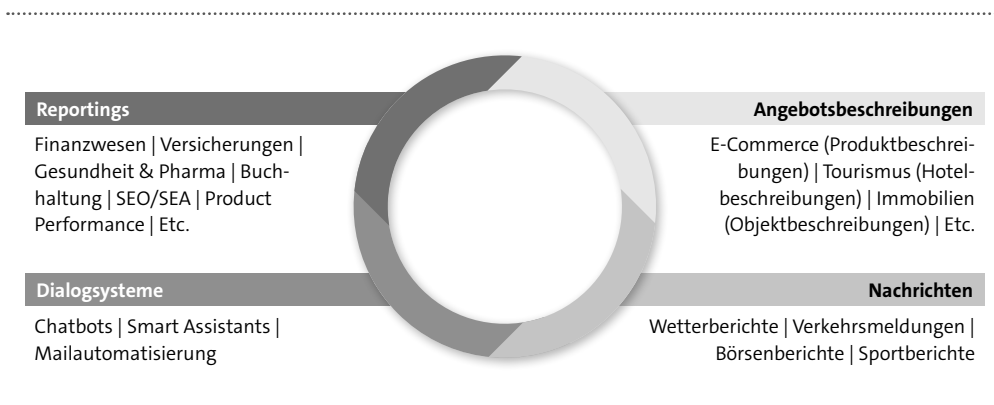


Abbildung 2: Die vielfältigen Anwendungsbereiche von NLG

### 3 NLG greift auf zahlreiche computerlinguistische Raffinessen zurück, um Daten in natürlichsprachliche Texte umzuwandeln

Die ersten Anfänge von NLG liegen in den 80ern, als Textgenerierungssysteme mehrheitlich noch auf Templates sowie auf sehr rudimentären Semantik- und Grammatikregeln basierten. Nach dem sogenannten *Slotfilling*-Schema »Die Temperatur steigt heute von (*insert actual temperature*) auf (*insert future temperature*), das Wetter wird (*insert expected weather status*).« konnten recht generische Sätze formuliert werden. Im Laufe der Zeit wurden diese schablonenartig arbeitenden Systeme nach und nach verbessert und die zu leistenden Aufgaben haben an Komplexität zugenommen. Heutige NLG-Systeme können mithilfe von Künstlicher Intelligenz anspruchsvollere Ansätze für die Sprachgenerierung verfolgen und damit auch komplexere Anwendungsfälle angehen. Sie generieren Texte von solcher Qualität und in so großem Variantenreichtum (siehe oben), dass menschliche Rezipienten automatisch generierte Texte in der Regel nicht mehr von Texten unterscheiden können, die von Menschen verfasst wurden.

Die dahinterliegenden Prozesse beinhalten eine Reihe von technologischen und computerlinguistischen Raffinessen: Große Mengen an strukturierten Rohdaten – sie enthalten beispielsweise Auskunft über Namen, Häufigkeiten, Produkteigenschaften, Messwerte usw. – werden zunächst der Software hinzugefügt.

Damit nun jedoch Texte in natürlicher Sprache entstehen können, benötigt ein NLG-System vorab sogenannte Templates und Bedingungen. Hinter Templates verbergen sich im Wesentlichen vorformulierte Sätze, die mithilfe von Lexikalisierungsalgorithmen mit einer großen Anzahl von Varianten und Synonymen angereichert werden. Unter Bedingungen hingegen versteht man Voraussetzungen, die erfüllt sein müssen, damit ein bestimmtes Template verwendet werden kann. Um Texte automatisch generieren zu können, ermittelt das System eine relevante Bedingung, kombiniert Informationen aus den Templates mit ereignis- oder produktspezifischen Daten und erstellt mithilfe einer intelligenten linguistischen Analyse das Endprodukt. Nebenbei bemerkt: NLG-Software weiß in der Regel auch, wie man die Templates in einer Reihenfolge – auch bekannt als »Story Plot« oder »Narrativ« – anordnet, sodass sie inhaltlich für den menschlichen Rezipienten Sinn machen und ansprechend zu lesen sind. Menschliche Redakteure müssen in der Initialphase lediglich Templates erstellen und Bedingungen definieren – liegen diese Informationen jedoch einmal vor, kann das NLG-System komplett autonom arbeiten.

»Heutige NLG-Systeme generieren Text von solcher Qualität, dass die Leser automatisch generierte Texte nicht mehr von menschlich verfassten Artikeln unterscheiden können.«



## 4 NLG wird Journalisten zunehmend unterstützen

Als relativ junge Technologie steht Natural Language Generation einerseits vor einer Fülle an Möglichkeiten, andererseits vor zahlreichen ethischen Herausforderungen. So zog im Februar 2019 die US-amerikanische Non-Profit-Organisation [OpenAI](#) mediale Aufmerksamkeit auf sich: Die Forschungseinrichtung entwickelte ein KI-basiertes Sprachmodell namens Gpt2, das Texte von rund acht Millionen Websites – insgesamt 40 Gigabyte Daten – analysiert hat und anschließend automatisiert Texte verfassen sollte. Von den Ergebnissen waren die Forscher selbst überrascht: Der KI-Textgenerator lieferte qualitativ so gute Texte, dass sich die Forscher aus Sorge um Missbrauch dazu entschlossen haben, das Sprachmodell nicht bzw. nur in einer abgespeckten Version zu veröffentlichen. »We started testing it, and quickly discovered it's possible to generate malicious-esque content quite easily«, kommentiert Jack Clark, Policy Director von OpenAI, gegenüber dem [MIT Technology Review](#). Die größten Sorgen der Forscher bezogen sich darauf, dass das System für die kostengünstige und massenhafte Produktion und Distribution von Fake-News in sozialen Netzwerken missbraucht werden könnte: »Due to our concerns about malicious applications of the technology, we are not releasing the trained model. As an experiment in responsible disclosure, we are instead releasing a much smaller model for researchers to experiment with, as well as a technical paper«, zitiert [The Guardian](#) OpenAI [4].

Ein Großteil der gegenwärtigen Anwendungen im Bereich der Natural Language Generation erreicht nicht die Leistung des Gpt2-Sprachmodells. Dennoch skizziert der Fall wichtige (digital-)ethische Herausforderungen. Auch das Thema Fake-News wird in diesem Zusammenhang intensiv diskutiert. Die Sorge ist groß, dass sich über soziale Medien Nachrichten ohne Wahrheitsgehalt verbreiten, die demokratische Prozesse beeinflussen können.

Ein weiteres soziopolitisch relevantes Thema ist der Einfluss von KI-basierter Sprachtechnologie auf den Arbeitsalltag von Mitarbeitern. Ein gutes Beispiel für eine Berufsgruppe, die durch KI unterstützt, aber keinesfalls ersetzt wird, ist die des Journalismus. Schon heute übernehmen Algorithmen die Berichterstattung über Bereiche, über die Journalistinnen und Journalisten nie schreiben würden – sei es aus Kostengründen, weil die Arbeit zu monoton ist, oder weil die Zielgruppe zu klein wäre. Man denke beispielsweise an kleinere olympische Sportarten, untere Fußball-Ligen oder lokale Wetterberichte in Echtzeit. NLG-Systeme decken somit Bereiche ab, die durch Journalistinnen und Journalisten nicht hätten bearbeitet werden können.

Das Erzählen von Geschichten mit Kreativität, Empathie, Raffinesse und Ironie, das Schreiben von pointierten Kommentaren und scharfsinnigen Analysen bleibt nach wie vor die Kernkompetenz der Menschen – und gehört wohl nebenbei bemerkt auch zu den bevorzugten Tätigkeiten der Journalistinnen und Journalisten. Die Zukunft des sogenannten Roboterjournalismus besteht keinesfalls darin, den menschlichen Journalisten zu ersetzen, sondern ihn zu unterstützen, um ihnen Zeit und Raum für wichtige und sinnstiftende Tätigkeiten zu geben. Während Technik ungeliebte Routinetätigkeiten übernimmt, können sich Journalistinnen und Journalisten ihrer eigentlichen Kernarbeit widmen. Ein solches kybernetisches (ergo: nicht ersetzendes) Zusammenspiel von KI und maschinellen Prozessen auf der einen, und menschlicher Kreativität und Kompetenz auf der anderen Seite, kann auf diese Weise das journalistische Angebot für Leserinnen und Lesern verbessern und den Arbeitsalltag von Journalistinnen und Journalisten angenehmer gestalten. Eine Win-Win-Situation für alle Beteiligten.

*»Algorithmen übernehmen die Berichterstattung in Bereichen, über welche Journalisten nie schreiben würden – sei es aus Kostengründen, weil die Arbeit zu monoton ist, oder weil die Zielgruppe zu klein wäre.«*

## 5 Literaturverzeichnis

1. Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2016). Perception of automated computer-generated news: Credibility, expertise, and readability. Journalism. Advance online publication.
2. Gil Press, Top 10 Artificial Intelligence (AI) Technologies, 23.01.17, Forbes.
3. Kasey Panetta, Neural networks and modern BI Platforms will evolve data and analytics, 16.01.17, Gartner.
4. John Naughton, AI is making literary leaps – now we need the rules to catch up, 2.11.19, The Guardian.

## 6 Autoren



### **Kristin Strauch**

Kristin Strauch ist Senior Communication Managerin bei dem Berliner Tech-Unternehmen [Retresco](#), Anbieter von KI-basierten Sprachtechnologien wie Natural Language Generation und Natural Language Understanding.



### **Dr. Nabil Alsabah**

Dr. Nabil Alsabah ist Bereichsleiter Künstliche Intelligenz und Big-Data.AI Summit im Bitkom.





Bitkom vertritt mehr als 2.700 Unternehmen der digitalen Wirtschaft, davon gut 1.900 Direktmitglieder. Sie erzielen allein mit IT- und Telekommunikationsleistungen jährlich Umsätze von 190 Milliarden Euro, darunter Exporte in Höhe von 50 Milliarden Euro. Die Bitkom-Mitglieder beschäftigen in Deutschland mehr als 2 Millionen Mitarbeiterinnen und Mitarbeiter. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig oder in anderer Weise Teil der digitalen Wirtschaft. 80 Prozent der Unternehmen haben ihren Hauptsitz in Deutschland, jeweils 8 Prozent kommen aus Europa und den USA, 4 Prozent aus anderen Regionen. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem weltweit führenden Digitalstandort zu machen.

**Bundesverband Informationswirtschaft,  
Telekommunikation und neue Medien e.V.**

Albrechtstraße 10  
10117 Berlin  
T 030 27576-0  
F 030 27576-400  
bitkom@bitkom.org  
[www.bitkom.org](http://www.bitkom.org)

**bitkom**