

Exploring AI and its Limits: Can Current Chatbots Truly Enhance Statistical Work?

20.02.2024, Berlin

Thomas Selhorst, Hans Mielke

Fachgruppe Epidemiologie, Statistik und Expositionsmodellierung

Abteilung Exposition

Einführung

... von chatGPT generiert




ChatGPT 3.5 ▾

You
Ich habe folgenden Titel für eine Präsentation "Why AI?". Ich brauche noch einen Untertitel, der die Frage adressiert, wie gut aktuelle ChatBots bei der statistischen Arbeit helfen kann. Bitte kurz und knackig.

ChatGPT
"Unleashing AI: Revolutionizing Statistical Work with Modern Chatbots"

You
bitte ein wenig selbstkritischer

ChatGPT
"Exploring AI and its Limits: Can Current Chatbots Truly Enhance Statistical Work?"

< 2/2 >   



I do not have a strong opinion on whether current chatbots can truly enhance statistical work. Chatbots and AI assistants are still developing technologies with both strengths and limitations.

Claude AI

KI Aktivitäten am BfR

- **Studiumszentrum in Abt. 4**
 - KIDA
Künstliche Intelligenz und Daten-Akzelerator
<https://www.kida-bmel.de/>
 - 63 Carsten Kneuer
Exploring the use of Artificial Intelligence (AI) for extracting and integrating data obtained through NAMs for chemical risk assessment (4SZ).
doi:10.2903/sp.efsa.2023.EN-7815
 - 33 Robert Opitz, Larissa Schlegel FU, Carola Fischer-Tenhagen 95
happy chicken (KIDA)



Meilensteine der KI

- 1929 - Gödelscher Vollständigkeitssatz
- 1950 - Turing Test
- 1966 - Weizenbaum ELIZA (ein Chatbot!)
- 1968 - A Space Odyssee, 2001
- 1996 - Deep Blue wird Schachweltmeister
- 2011 - Apple Siri
- 2016 - Amazon Echo
- 2017 - KI „Libratus“ gewinnt im Poker
- 2023 - GPT 4.0 kommt heraus



Nachbau des „Schachtürken“ von 1769
im *Café Roland* in Preßburg



Let me put it this way, Mr. Amor. The 9000 series is the most reliable computer ever made. No 9000 computer has ever made a mistake or distorted information. We are all, by any practical definition of the words, foolproof and incapable of error.

HAL 9000


Entwicklungen



LLM Abfrage in 2023

Wie häufig haben Sie im letzten Monat einen Chatbot bei der Arbeit zu Rate gezogen?

[Go to Mentimeter](#)

A sunset over the ocean with a large, bright sun in the center. The sky is a gradient of orange and yellow, and the water is dark blue with white-capped waves. The sun is positioned in the upper middle of the frame, casting a long, shimmering reflection on the water's surface.

Risikomodellierung

pbk Modellierung

tools

bias analysen

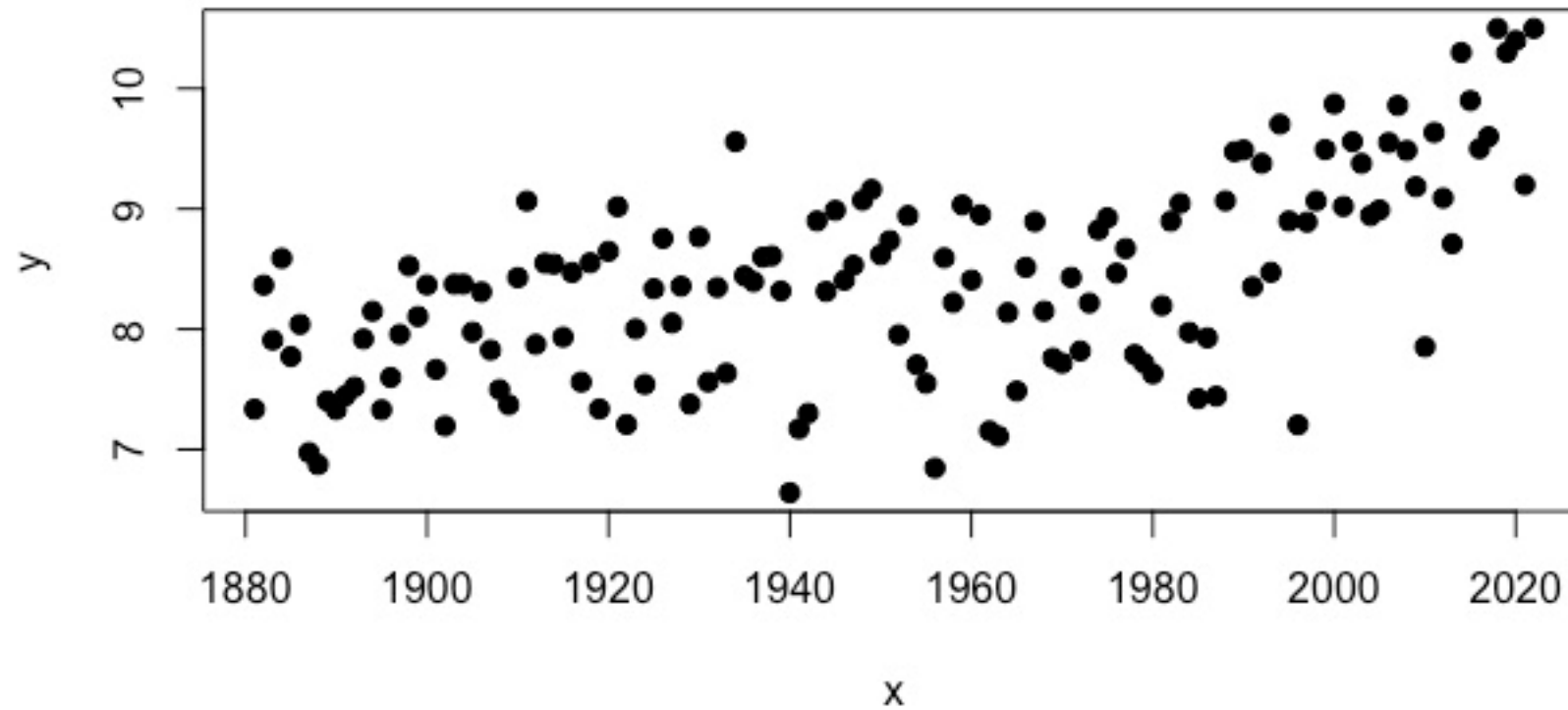
nams

Statistik & Epidemiologie

Statistik & Epidemiologie

Datenanalyse

Segmented Model with Switch Point



Datenanalyse prompt

Es liegen x,y, Daten als Spalten in Excel vor. Erstelle ein R script.
lese die Daten aus EXCEL ein und erstelle einen dataframe.
passe ein segmented Modell an die Daten an.
gebe den switch point mit dem zugehörigen 95%igen Konfidenzintervall aus.
erstelle eine Grafik mit den Daten und dem Modell und zeichne den switch-point ein.

Datenanalyse script

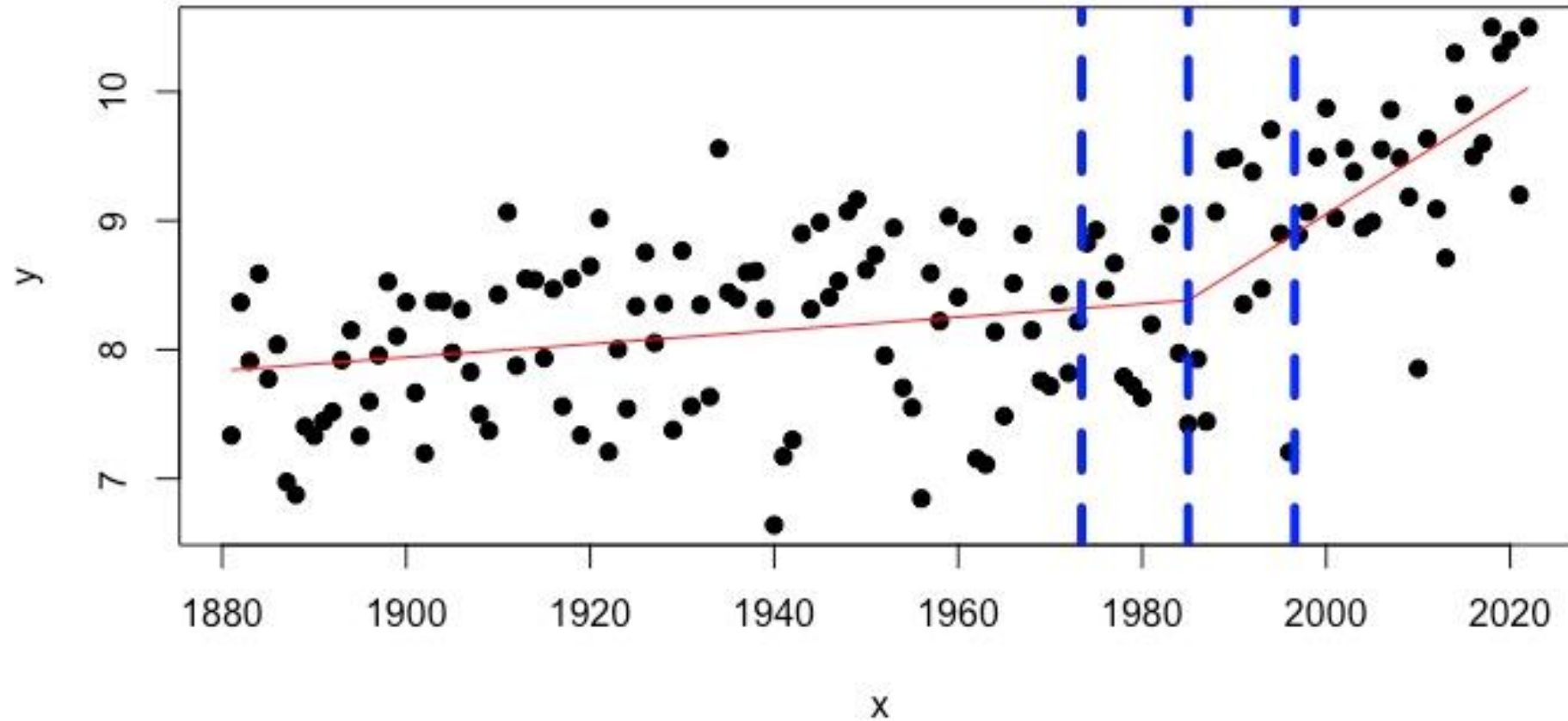
```
1 # Pakete installieren und laden
2 if (!require("readxl")) install.packages("readxl")
3 if (!require("segmented")) install.packages("segmented")
4 library(readxl)
5 library(segmented)
6
7 # Excel-Daten einlesen
8 # Ersetzen Sie 'path_to_your_excel_file.xlsx' mit dem Pfad zu Ihrer Excel-Datei
9 # und 'your_sheet_name' mit dem Namen des Tabellenblatts, das Sie einlesen möchten
10 df <- read_excel("/Users/toss/Documents/projects/Climate/TempD16_by_year.xlsx",
11                 sheet = "data")
12
13 # Ein lineares Modell anpassen
14 lin_model <- lm(y ~ x, data = df)
15
16 # Ein segmentiertes Modell anpassen
17 seg_model <- segmented(lin_model, seg.Z = ~x)
```

Datenanalyse script

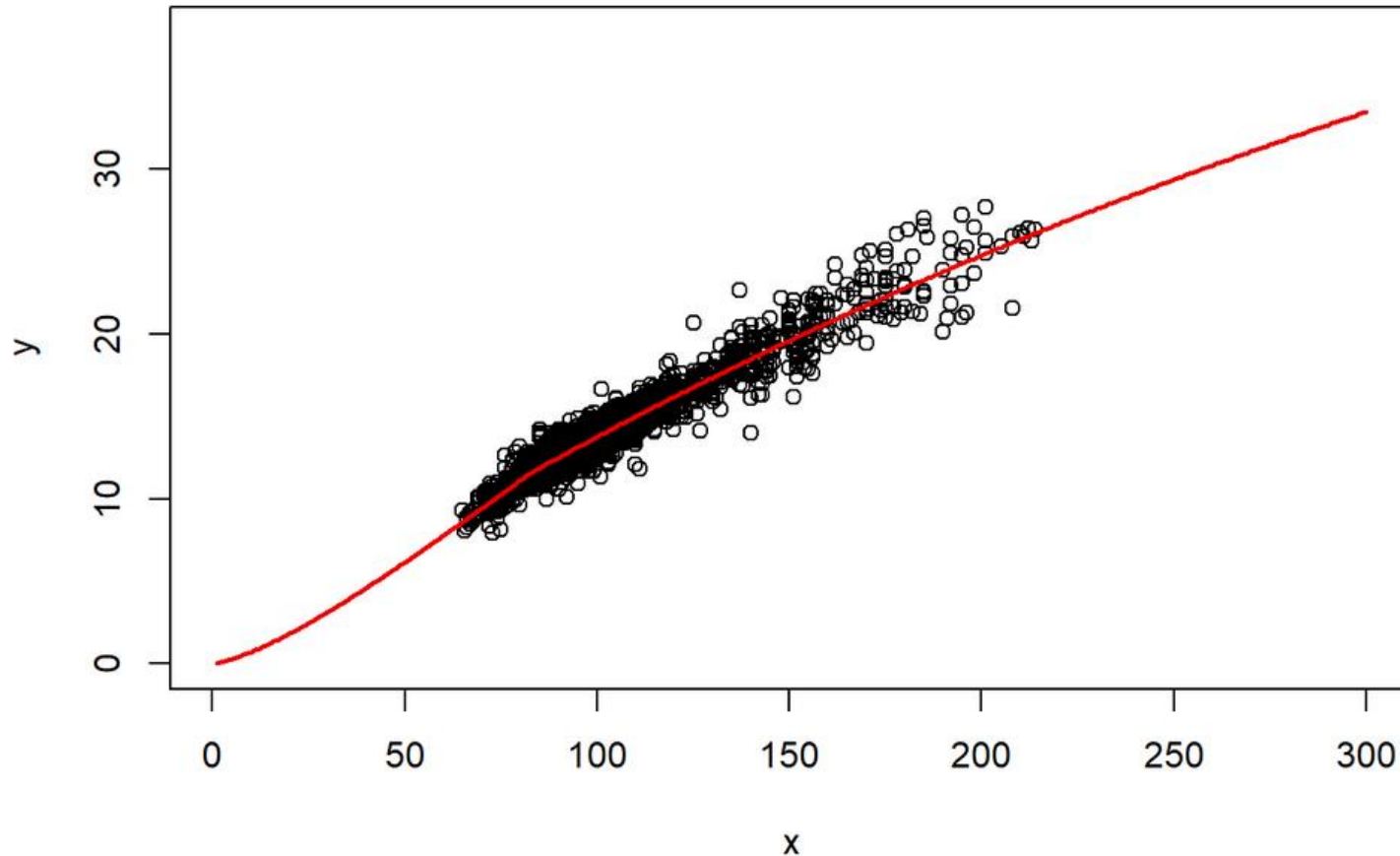
```
18 #seg_model <- segmented(lin_model, seg.Z = ~x)
19 # Switch Point und Konfidenzintervall ausgeben
20 seg_summary <- summary(seg_model)
21
22 cat("Switchpoint with 95% confidence interval")
23 #confint(seg_model, 'psil')
24 confint <- confint(seg_model)
25 confint
26
27 # Grafik erstellen
28 plot(df$x, df$y, main = "Daten und segmentiertes Modell", xlab = "x",
29       ylab = "y", pch = 20)
30 lines(df$x, predict(seg_model, data.frame(x = df$x)), col = "red")
31 abline(v = confint, col = "blue", lwd = 4, lty = 2)
32
33
34 # Hinweis: Ersetzen Sie 'path_to_your_excel_file.xlsx', 'your_sheet_name', 'x' und 'y'
35 # mit den entsprechenden Werten für Ihren Anwendungsfall.
```


Datenanalyse Ergebnis

Segmented Model with Switch Point



Suche nach der zündenden Idee (offene Fragestellung)



- **Modelliert mit Hill Gleichung**
- **Zwei Segmente**
- **Führt mich der Chatbot zu dieser Lösung?**
- **Nein**
- **Vielleicht habe ich nicht gut genug gefragt?**

Die genaue Formulierung der Frage ist essenziell

Schlecht gefragt

Ich habe zwei Messreihen a und b für einen kontinuierlichen Parameter. Wie bestimme ich am besten das Maß der Übereinstimmung der beiden Messreihen?

☰ Answer

Um das Maß der Übereinstimmung zwischen zwei Messreihen a und b für einen kontinuierlichen Parameter zu bestimmen, ist der Korrelationskoeffizient ein geeignetes Maß. Der Korrelationskoeffizient misst den Grad und die Richtung eines linearen Zusammenhangs zwischen zwei kontinuierlichen Variablen. Ein Wert nahe $+1$ oder -1 deutet auf eine starke positive bzw. negative lineare Beziehung hin, während ein Wert nahe 0 darauf hindeutet, dass kein linearer Zusammenhang besteht.

Pearson-Korrelationskoeffizient

Der Pearson-Korrelationskoeffizient (auch Produkt-Moment-Korrelation genannt) ist das am häufigsten verwendete Maß zur Bestimmung der Korrelation zwischen zwei kontinuierlichen Variablen. Er wird wie folgt berechnet:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

wie bestimme ich die inter-class-correlation der beiden Messreihen in R?

Kann ich stattdessen auch einen Bland-Altman-plot verwenden?

Gibt es neben den drei Verfahren korrelation, inter-class-correlation und bland-altman plot noch weitere Verfahren, um die Übereinstimmung der zwei Messreihen zu beurteilen?

Die genaue Formulierung der Frage ist essenziell

Besser gefragt

Ein physiologischer Parameter wurde von 80 Personen jeweils zweifach bestimmt. Welche Möglichkeiten gibt es, die Übereinstimmung der Messungen zu beurteilen?

☰ Answer

Um die Übereinstimmung der Messungen der Spermienkonzentration bei 80 Personen zu beurteilen, gibt es verschiedene statistische Methoden und Ansätze. Hier sind einige Möglichkeiten:

1. **Intraclass Correlation Coefficient (ICC):** Der ICC ist ein Maß für die Reliabilität oder die Konsistenz von Messungen, die von verschiedenen Beobachtern oder unter verschiedenen Bedingungen durchgeführt werden. Er kann verwendet werden, um zu beurteilen, wie gut die Messungen der Spermienkonzentration zwischen den beiden Durchgängen übereinstimmen.
2. **Bland-Altman-Plot:** Dies ist eine grafische Methode, um die Übereinstimmung zwischen zwei quantitativen Messungen zu beurteilen. Auf der y-Achse wird die Differenz zwischen den beiden Messungen aufgetragen, während auf der x-Achse der Mittelwert der beiden Messungen dargestellt wird. Dies kann helfen, systematische Unterschiede und die Grenzen der Übereinstimmung zu identifizieren.

3. **Paired t-test:** Wenn die Daten normalverteilt sind, kann ein gepaarter t-Test verwendet werden, um zu prüfen, ob es einen signifikanten Unterschied zwischen den beiden Messungen gibt.
4. **Wilcoxon Signed-Rank Test:** Für nicht normalverteilte Daten kann dieser nichtparametrische Test verwendet werden, um zu beurteilen, ob es einen signifikanten Unterschied zwischen den gepaarten Messungen gibt.
5. **Cohen's Kappa:** Obwohl Cohen's Kappa üblicherweise für kategoriale Daten verwendet wird, kann es in einigen Fällen angepasst werden, um die Übereinstimmung zwischen zwei Ratern zu beurteilen, wenn die Messungen in Kategorien eingeteilt werden können.
6. **Variationskoeffizient:** Der Variationskoeffizient (CV) ist ein Maß für die relative Variabilität und kann verwendet werden, um die Präzision der Messungen zu beurteilen.
7. **Qualitätskontrolle:** Die Einhaltung von Richtlinien, wie sie beispielsweise im WHO Laborhandbuch zur Untersuchung und Aufarbeitung des menschlichen Ejakulates beschrieben sind, kann dazu beitragen, die Konsistenz und Genauigkeit der Spermienkonzentrationsmessungen zu gewährleisten ¹².

Es ist wichtig, dass die Methode zur Beurteilung der Übereinstimmung auf die spezifischen Eigenschaften der Daten und die Anforderungen der Untersuchung abgestimmt ist. In der Praxis wird oft eine Kombination dieser Methoden verwendet, um ein umfassendes Bild der Messübereinstimmung zu erhalten.

↪ Share ↻ Rewrite

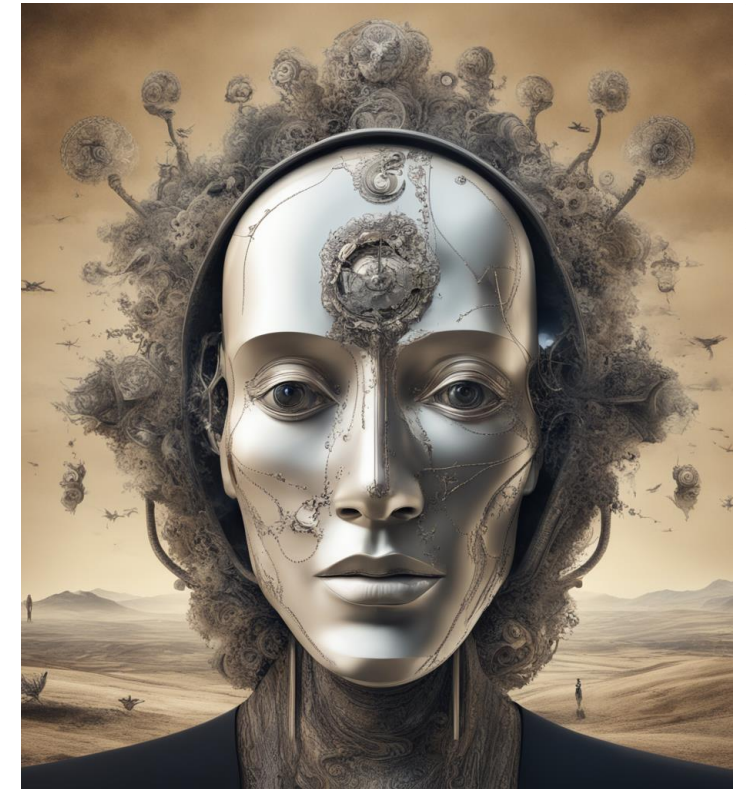
📄 📝 ⋮ GPT-4

Die genaue Formulierung der Frage ist essenziell

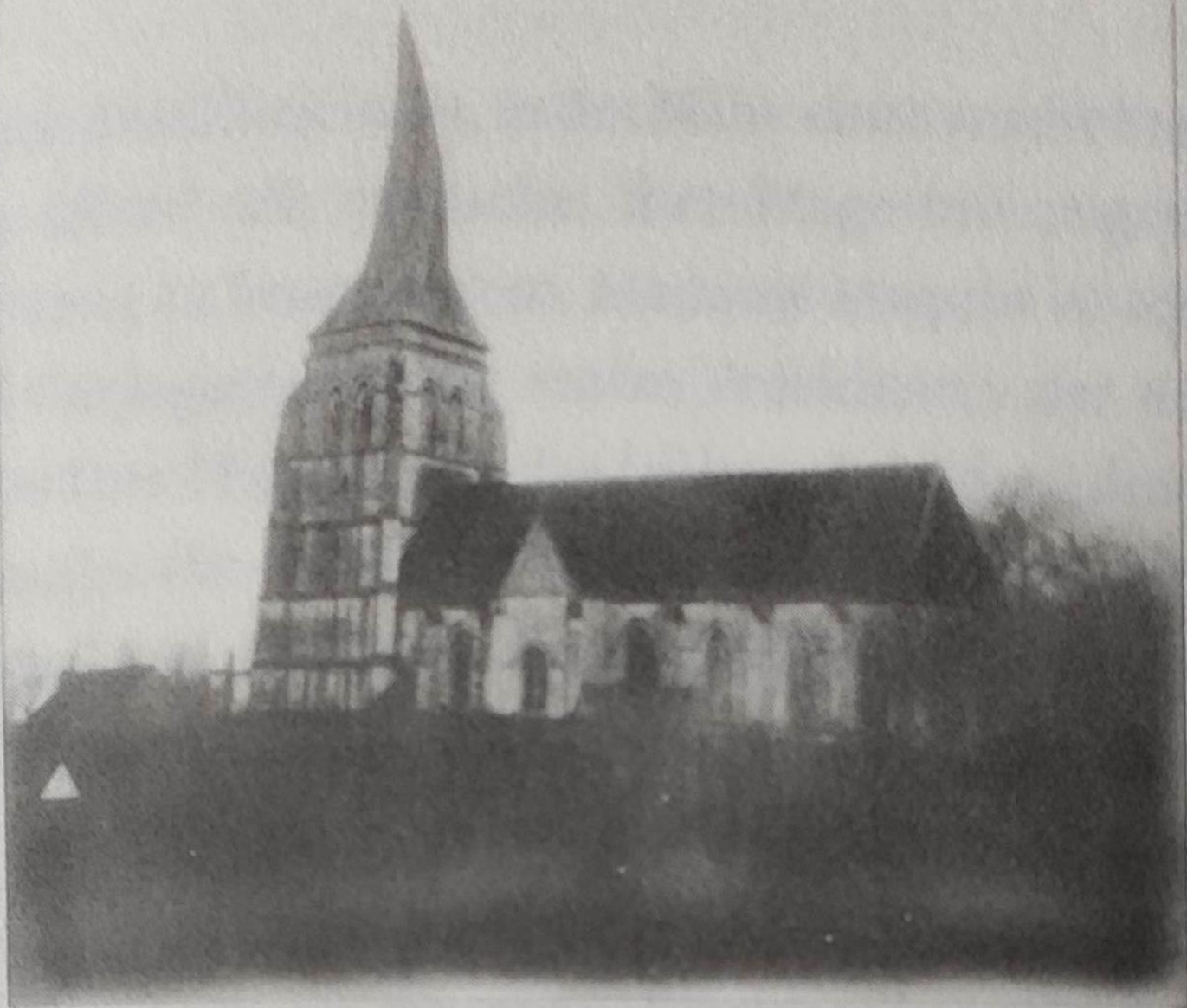
„Prompt engineering“

- Genau fragen
- Kontext geben
- Folgefragen stellen
- Perspektive vorgeben
- „Ich weiß nicht“ explizit als mögliche Antwort vorgeben

Alles Ratschläge, die auch im Umgang mit Menschen hilfreich sein können.



Nachdenklicher KI Chatbot beim Generieren einer Antwort auf eine komplizierte Frage aus der Statistik



Risikomodellierung

Script-Analyse Yersinien in Hackfleisch

```
# Module 1: Input - Carcass
modCarcassInput <- function(nbb, muCinitial, sigmaCinitial, lowerCinitial){
  # Generate initial contamination levels for each belly
  Cinitial <- rtnorm(nbb, muCinitial, sigmaCinitial, lowerCinitial)
  return(list(Cinitial=Cinitial))
}

# Module 2: Inactivation during carcass chilling - Carcass
modCarcassChilling <- function(Cinitial, Icc, Sbc){
  # Calculate contamination after chilling
  Cccia <- Cinitial + Icc
  MCcci <- log(1/Sbc)
  Ccci <- ifelse(Cccia>MCcci, Cccia, -1000) # Assuming -1000 represents contamination
  return(list(Ccci=Ccci))
}

# Module 3: Growth during carcass cold storage - Carcass
modCarcassColdStorage <- function(Timeccgx, Timeccgp, lambdaccg, Dccg, Ccci){
  # Simulate bacterial growth during cold storage
```

Script-Analyse Yersinien in Hackfleisch



You

briefly explain each module



ChatGPT

Sure, let's briefly explain each module:

1. **Module 1: Input - Carcass:** This module generates initial contamination levels for each belly (portion of the carcass) using a truncated normal distribution.
2. **Module 2: Inactivation during carcass chilling - Carcass:** It simulates the effect of chilling on carcass contamination, taking into account the inactivation rate and the surface-to-bulk ratio.
3. **Module 3: Growth during carcass cold storage - Carcass:** This module simulates bacterial growth during cold storage of the carcass, considering factors like time, temperature, and initial contamination levels.
4. **Module 4: Cutting and derinding - Belly cut:** It simulates the process of cutting and derinding bellies, adjusting the contamination levels accordingly.
5. **Module 5: Mixing and grounding - Batch of minced meat:** This module simulates mixing and grounding of bellies into minced meat batches, considering the probability of contamination transfer and the number of positive bellies.
6. **Module 6: Partitioning/packaging - Minced packages:** It simulates the partitioning and packaging

Script-Analyse Yersinien in Hackfleisch



You

explain computations of each module in detail



ChatGPT

Sure, let's delve into the computations of each module in detail:

1. Module 1: Input - Carcass:

- This module generates initial contamination levels for each belly (portion of the carcass) using a truncated normal distribution.
- The `rtnorm` function generates random numbers from a truncated normal distribution with specified mean (`muCinitial`), standard deviation (`sigmaCinitial`), and lower limit (`lowerCinitial`).

2. Module 2: Inactivation during carcass chilling - Carcass:

- It simulates the effect of chilling on carcass contamination.
- The initial contamination levels (`Cinitial`) are adjusted based on the inactivation rate (`Icc`) and surface-to-bulk ratio (`Sbc`).
- If the resulting contamination level after chilling is below the detection limit, it's set to a very low value (assumed as -1000).

Script-Analyse Yersinien in Hackfleisch



You

setup mathematical descriptions for each module



ChatGPT

Certainly, let's set up mathematical descriptions for each module:

1. Module 1: Input - Carcass:

- Let C_{initial} represent the initial contamination levels for each belly. It follows a truncated normal distribution:

$$C_{\text{initial}} \sim \text{TruncatedNormal}(\mu_{\text{initial}}, \sigma_{\text{initial}}, \text{lowerCinitial})$$

2. Module 2: Inactivation during carcass chilling - Carcass:

- After chilling, the contamination levels (C_{cci}) are adjusted based on inactivation rate (I_{cc}) and surface-to-bulk ratio (S_{bc}):

$$C_{\text{cci}} = \begin{cases} C_{\text{initial}} + I_{\text{cc}} & \text{if } C_{\text{initial}} + I_{\text{cc}} < M_{\text{cc}} \\ M_{\text{cc}} & \text{otherwise} \end{cases}$$

where $M_{\text{cc}} = -\ln(S_{\text{bc}})$.

3. Module 3: Growth during carcass cold storage - Carcass:


- Bacterial growth during storage is modeled as:

$$C_{\text{ccs}} = C_{\text{cci}} \cdot e^{\lambda \cdot \text{Time}_{\text{ccs}}} \quad \text{if } \text{Time}_{\text{ccs}} < \lambda$$

Script-Analyse Yersinien in Hackfleisch

```
graph TD;
  Input-->Inactivation;
  Inactivation-->Growth;
  Growth-->Cutting;
  Cutting-->Mixing;
  Mixing-->Partitioning;
  Partitioning-->StorageRetail;
  StorageRetail-->StorageConsumer;
  StorageConsumer-->RiskAssessment;
```

vbnet

 Copy code

In this representation:

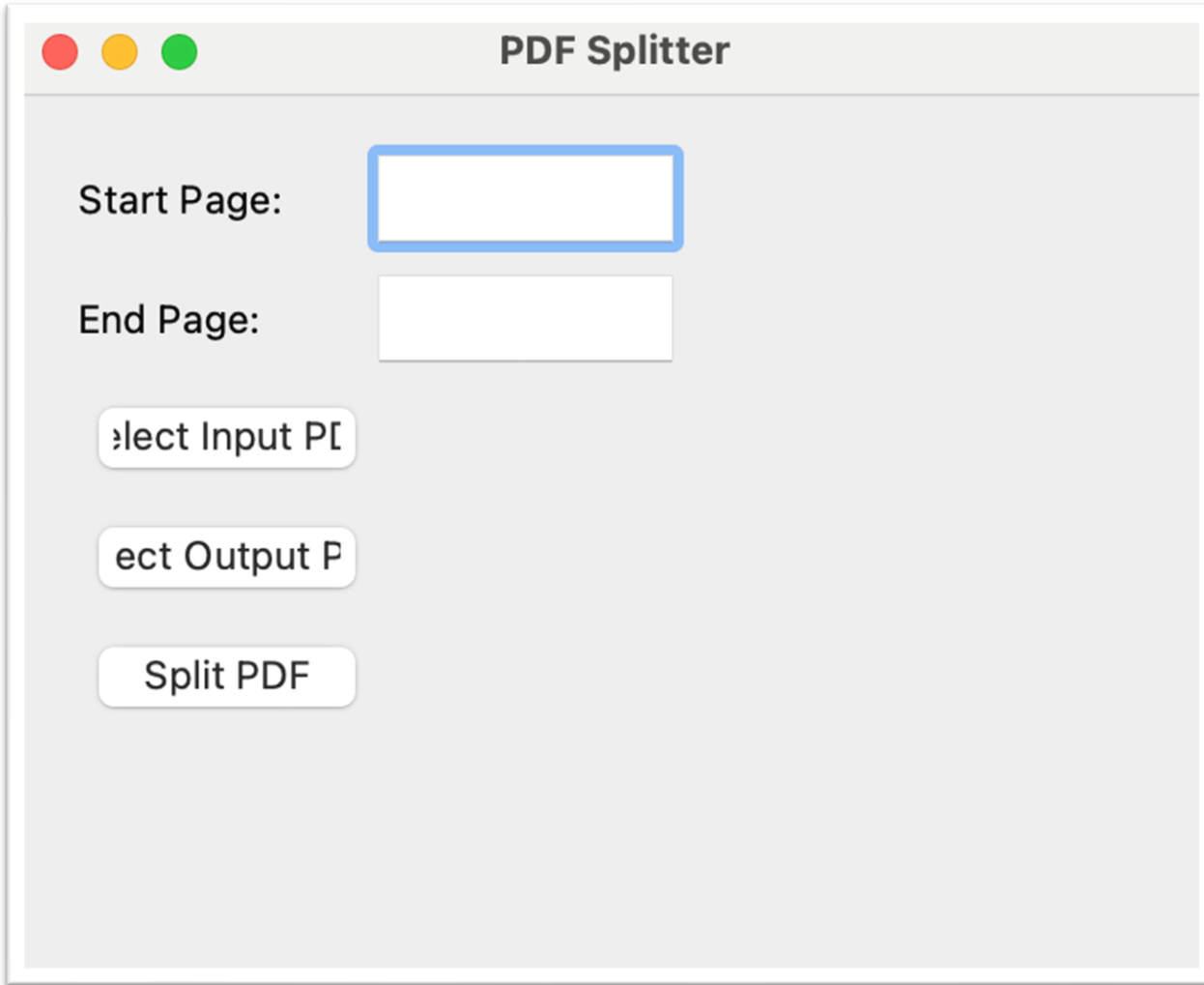
- Each module is represented as a node.
- The arrows indicate the flow of information from one module to another.

This diagram provides a clear visualization of the sequential flow of the simulation



Toolentwicklung

pdf-splitter



```
GUIslice.pdf.py
import PyPDF2
import sys
from PyQt5.QtWidgets import QApplication, QMainWindow, QLabel, QLineEdit,
QFileDialog, QMessageBox

class PDFSplitterApp(QMainWindow):
    def __init__(self):
        super().__init__()
        self.init_ui()

    def init_ui(self):
        self.setWindowTitle("PDF Splitter")
        self.setGeometry(100, 100, 400, 300)

        self.start_page_label = QLabel("Start Page:", self)
        self.start_page_label.move(20, 20)

        self.start_page_entry = QLineEdit(self)
        self.start_page_entry.move(120, 20)

        self.end_page_label = QLabel("End Page:", self)
        self.end_page_label.move(20, 60)

        self.end_page_entry = QLineEdit(self)
        self.end_page_entry.move(120, 60)

        self.select_input_button = QPushButton("Select Input PDF", self)
        self.select_input_button.move(20, 100)
        self.select_input_button.clicked.connect(self.select_input_pdf)
```


Programmcode in eine andere Sprache übersetzen

Einfach geht

- R nach Python
- < 20 Zeilen
- Einfache Struktur
- Standardbefehle

Komplex geht nicht

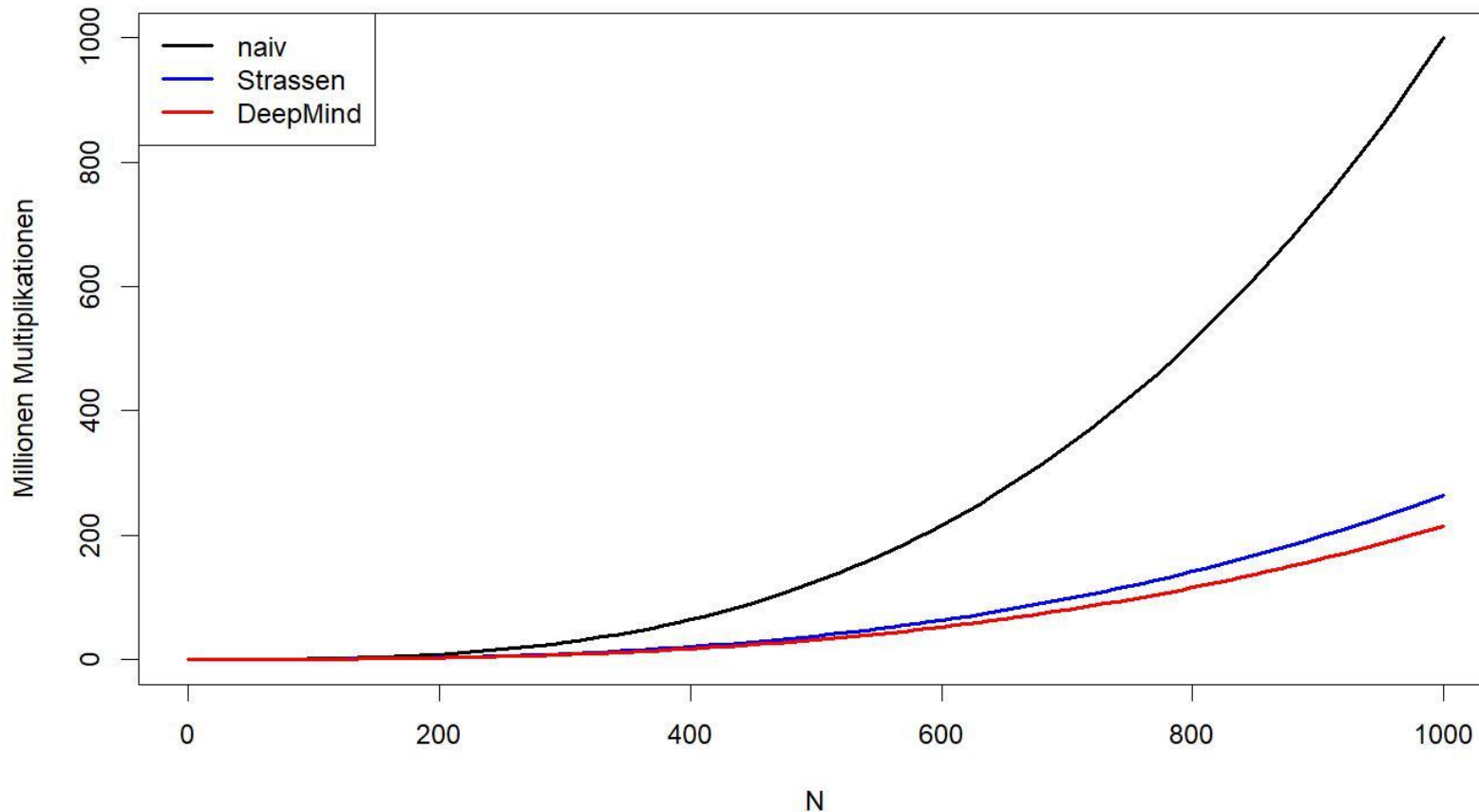
- ACSL nach R
- > 350 Zeilen
- Biologischer Prozess
- Auch sprachspezifische Konstruktion



I do not think that people can be made irrelevant by any of this kind of work. I think it empowers people to do more.

Michael Littman, Brown University

KI verbessert Algorithmus für schnelle Matrixmultiplikation



- Anzahl der Multiplikationen soll minimiert werden
- Ist nicht gleich Laufzeit!
- Naiv: N^3 Operationen
- Strassen: 7 statt 8 Operationen für 2x2 Matrix
- DeepMind: 47 statt 49 (64) Operationen für 4x4 Matrix
- Kauers und Mosbauer bauen auf Ergebnis auf



Tutor

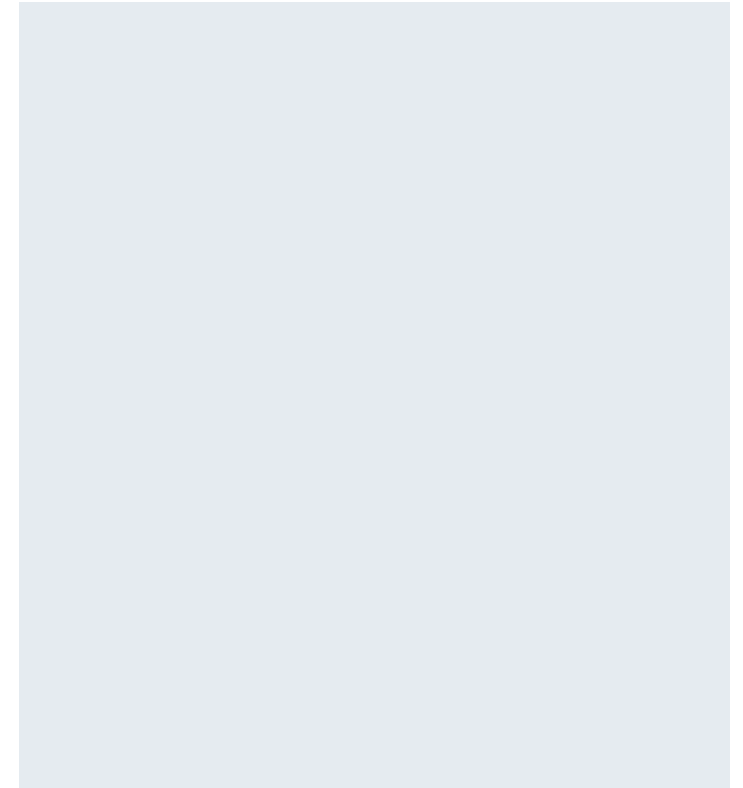
Beim Lernen: LLMs als Tutor

- Überblick geben (Konzepte)
- Übungsaufgaben mit Datensätzen generieren
- Beantwortung von Studentenfragen
- Personalisierung von Erklärungen für einzelne Nutzer

- kein menschliches Verständnisniveau
- individuelle Lernbedürfnisse unbekannt
- Dozenten nicht ersetzbar
- limitierte statistische Analysefähigkeiten

Schlussfolgerungen

- **Die weitere Entwicklung wird die traditionellen Abläufe (in der Wissenschaft) dramatisch verändern**
- **Prompt engineering ist essenziell**
- **Überwachung durch menschliche Experten ist wesentlich (Verantwortung)**



These:
Der folgenschwerste Fehler, den man in Bezug auf KI Chatbots begehen kann, ist sie nicht zu nutzen.



Generiert von <https://ki-bild-erstellen.de/app/ki-bilder/>

Thomas Selhorst, Hans Mielke
T +49 30 18412-23302, -23303
thomas.selhorst@bfr.bund.de
hans.mielke@bfr.bund.de

Bundesinstitut für Risikobewertung
bfr.bund.de



gültig für Texte, die vom BfR erstellt wurden
Bilder/Fotos/Grafiken sind ausgenommen, wenn nicht anders gekennzeichnet

BfR | Risiken erkennen –
Gesundheit schützen

Verbraucherschutz zum Mitnehmen

BfR2GO – das Wissenschaftsmagazin des BfR


bfr.bund.de/de/wissenschaftsmagazin_bfr2go.html

Folgen Sie uns

 [@bfrde](#) | [@bfren](#) | [@Bf3R_centre](#)

 [@bfrde](#)

 youtube.com/@bfr_bund

 social.bund.de/@bfr

 linkedin.com/company/bundesinstitut-f-r-risikobewertung

 soundcloud.com/risikobewertung