SAS Club 2023

Der Business Analytics Club für SAS User

Wien, SAS Office Trabrennstraße
19. Oktober 2023

Gerhard Svolba, Phillip Manschek, Jens-Ole Harden,
Michael Weberberger (Premedia), Florian Stammer

Agenda

14:15 - 14:20 Uhr	Begrüßung / Intro / News Gerhard Svolba, SAS
14:20 - 14:50 Uhr	Es geht auch anders! - Erstellung analytischer Modelle mit SAS Viya Gerhard Svolba, SAS
14:50 - 15:20 Uhr	SAS und Generative AI - Überblick, Entwicklungen und Anwendungsbeispiele aus dem Marketing Michael Weberberger, Premedia // Florian Stammer & Gerhard Svolba, SAS
15:20 - 15:35 Uhr	Die SAS Explore Konferenz in Las Vegas - Ein Vor-Ort Bericht Gerhard Svolba, SAS
15:35 - 15:55 Uhr	PAUSE
15:55 - 16:25 Uhr	Fuzzy Matching von Steuernummern in externen Datenquellen mit SAS Mihai Paunescu, Bundesministerium für Finanzen
16:25 - 16:50 Uhr	SAS Studio Analyst und die Erweiterungsmöglichkeiten mit Custom Steps Phillip Manschek, SAS
16:50 - 17:15 Uhr	SAS Tipps und Tricks Session Jens Ole Harden, SAS
ab 17:15 Uhr	Gemütliches Get-Together mit Buffet

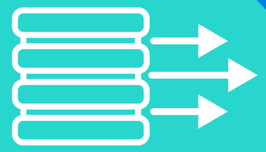
SAS und Generative AI - Überblick, Entwicklungen und Anwendungsbeispiele aus dem Marketing

Michael Weberberger, Premedia // Florian Stammer & Gerhard Svolba, SAS

Künstliche Intelligenz

ist die Disziplin, Systeme zu trainieren, um Abläufe der menschlichen Tätigkeit durch LERNEN und AUTOMATISIERUNG zu emulieren.

Lernen aus Erfahrungen/ Daten



Anpassen an neue Fakten

Automatisierung des Prozesses

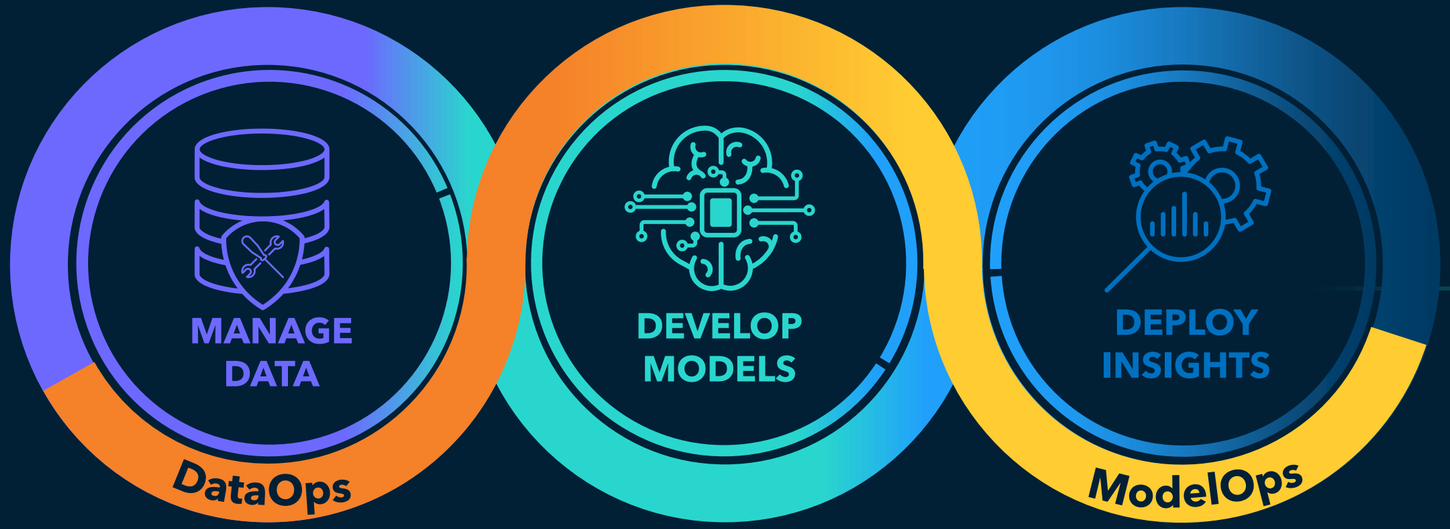
sas

Statistik, Machine & Deep Learning

Natural Language Processing

Daten Management

Produktiv-stellung



Visualisierung

Entscheidungs Management

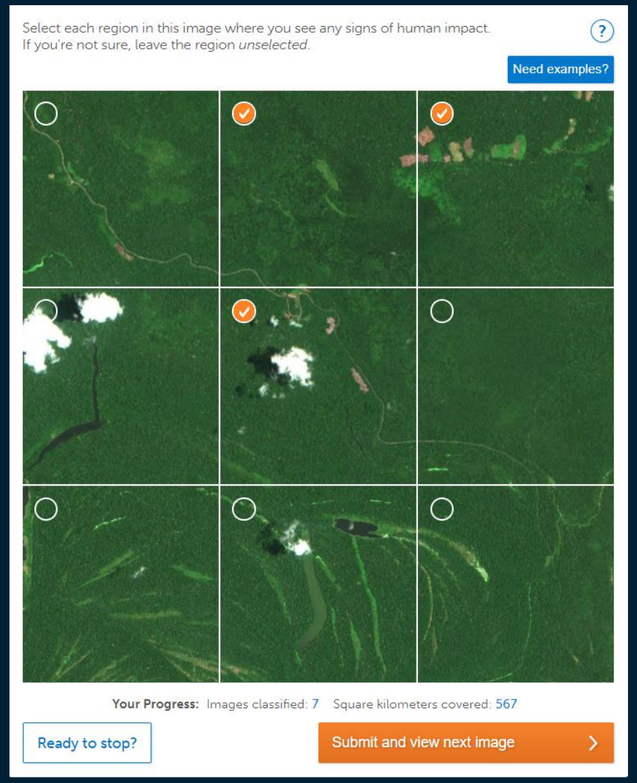
Computer & Machine Vision

Forecasting, Optimierung

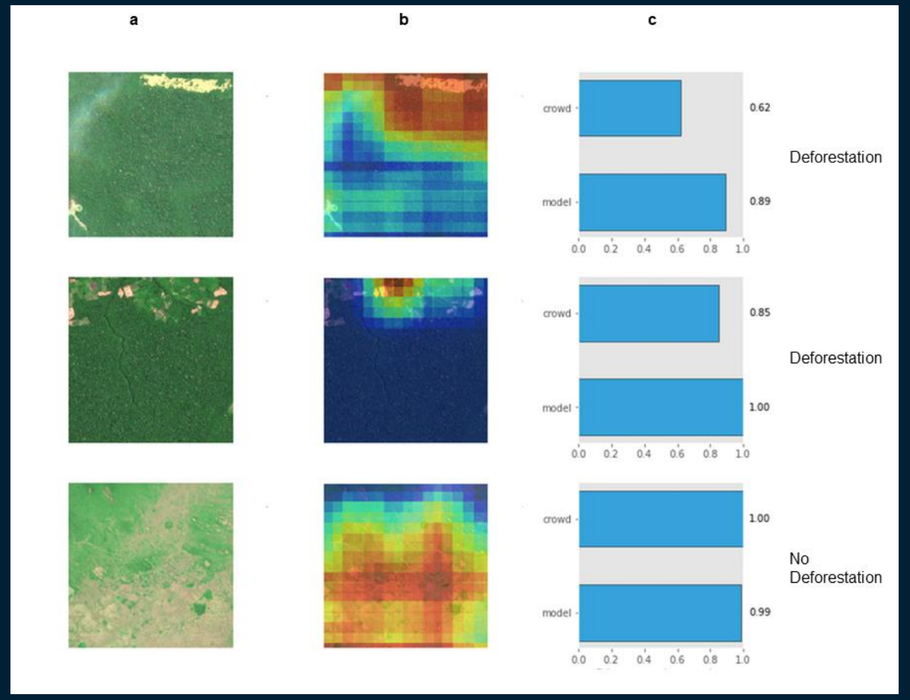
Analyse der Abholzung des Amazonas Regenwalds auf Basis der automatischen Klassifikation von Satellitenbildern

(Kooperation mit der IIASA, International Institute for Applied System Analysis)

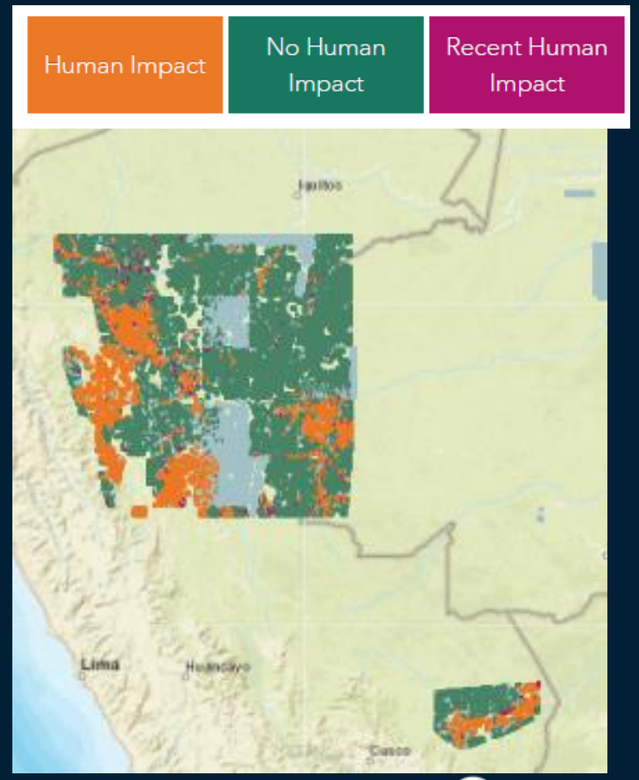
Crowd Sourcing (“Labelled Data”)



Modellierung mit SAS DLPY



Ziel: laufendes Scoring und Monitoring



Text-Analytik (Natural Language Processing)

Einsatz von Technologie, um die menschlichen Handlungen des Lesens, Organisierens und Quantifizierens von Freiform-Texten auf sinnvolle Weise zu skalieren.

Defining Generative AI (GAI)

Broadly speaking, whereas traditional AI/ML systems recognize patterns and make predictions, GAI systems learn real-world data to generate data – *like text, images, audio, tabular data, simulated data, code* – with similar probabilistic distributions and characteristics of the real-world data.

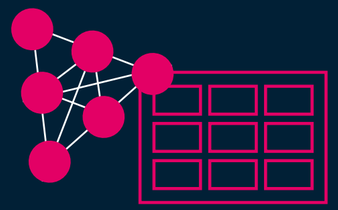
Given the transformative nature of GAI, business value, risk, and ethical considerations coexist.

These three considerations have yet to be reconciled.

GAI regulatory controls and governance are fluid. Controls are critical.

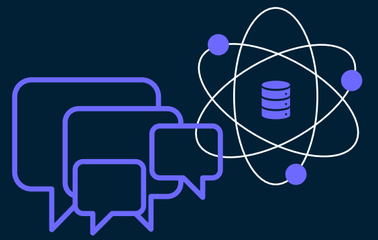
While there is no official or generally agreed upon definition of Generative AI (GAI) in the market, SAS considers these three AI/ML categories to be generative in nature

Synthetic Data Generation



Digital Twin

Large Language Models



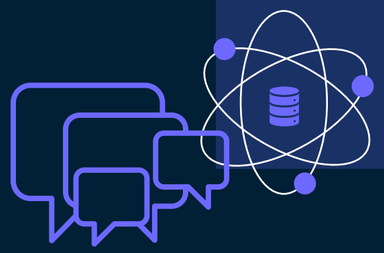
What are Large Language Models (LLMs)?

A large language model (LLM) is a type of natural language processing (NLP) artificial intelligence (AI) model that is designed to process and generate natural language text.

These models are typically trained on massive amounts of text data, using techniques like deep learning and neural networks to identify complex relationships in language.

SAS considers Large Language Models to be generative in nature given…

Algorithms like GPT, BERT, and recurrent neural networks (RNN) are used to generate text that is similar in style and content to the data the algorithms are trained on.



- Generate embeddings or representations of text that can be used as inputs to downstream models
- Generate/scale new insights by analyzing large amounts of text data and uncovering patterns and relationships that may not be immediately apparent
- Generate/combine with other NLP techniques like clustering and topic modeling to help balance and improve unstructured data diversity
- Generate text-based simulations of various scenarios like customer interactions or medical consultations for training purposes
- Generate/augment data labels or annotations for real-world data in areas such as sentiment, topics, or entity resolution
- Generate/augment text data to clean and normalize through spelling correction, format standardization, or the removal of redundant or irrelevant information

Which methods are associated with Large Language Models?

Sequence models

- Recurrent neural network (RNN) models like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) were the gold standard in language modeling before transformers.
- They are called sequence models because sentences must be processed word by word.
- They're still used in speech recognition, music generation, sentiment analysis, machine translation, and DNA sequence analysis.

Transformer models

- Transformer models are the current gold standard machine learning models in natural language processing. The first of these models, simply called Transformer was released in a paper called Attention is All You Need by Vaswani, et al., in 2017.
- They have the benefit of being **non-sequential** which means sentences are processed as a whole rather than word by word.
- **Self-attention** enables the model to look at the whole context of a sequence while encoding each of the input elements. It means the models will not "forget" parts of the input sequence because the window of retaining information is exactly as large as needed.
- Transformers feature positional embeddings, another innovation introduced to help replace recurrence. The idea is to use fixed or learned weights which encode information related to a specific position of a token in a sentence.

NLP & LLM capabilities at SAS

Current Capability

BERT-based Text Classifier action set

Capture the context and meaning of words in a text to improve accuracy compared with traditional models. In addition to general classification, the BERT-based classification can be used to do sentiment analysis.

RNN-based speech-to-text capabilities

Build n-gram language models and apply them to generate transcripts for speech-to-text applications. Evaluate the application's performance by calculating error rates of the transcription results.

Future Development

As SAS continues its large language model research and development efforts, teams are maintaining a key focus on innovating responsibly and safely in this emergent technology area.

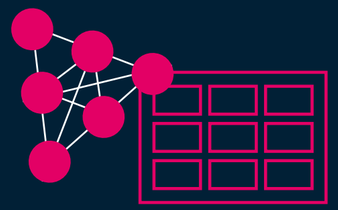
SAS currently *isn't* prescribing LLM solutions to customers given GAI regulation is fluid. However, SAS is collecting customer feedback on potential LLM use cases.

Text Classifier Action Set: Details

Provides actions for classifying textual data

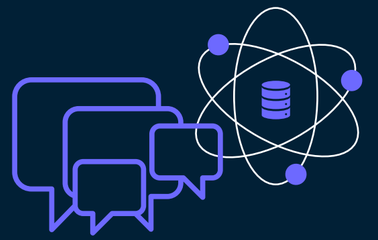
While there is no official or generally agreed upon definition of Generative AI (GAI) in the market, SAS considers these three AI/ML categories to be generative in nature

Synthetic Data Generation



Digital Twin

Large Language Models



Synthetic data is…

Synthetic data is on demand, self-service, or automated data generated by algorithms or rules, vs gathered in the real world, to meet conditions that may be lacking in real-world data.

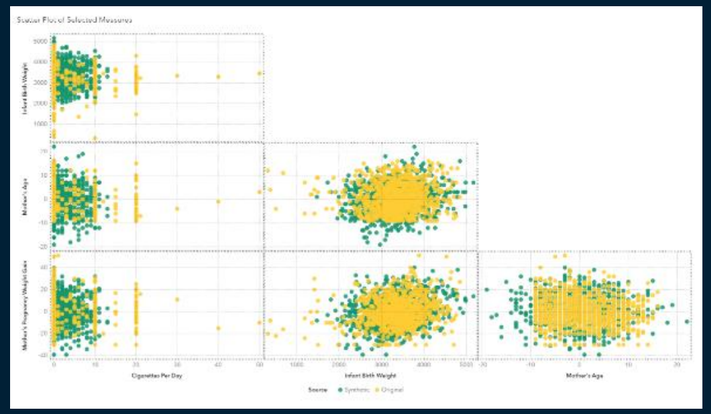
Synthetic data reproduces the same statistical properties, probability, patterns, and characteristics of the real-world dataset from which the synthetic data is trained.

Synthetic data is considered to be a privacy preservation enabler.

Many regulated industries are interested in using synthetic data.

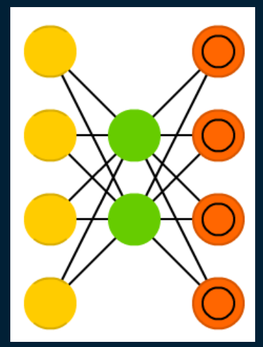
Source: Leigh Cullen, SAS & MIT Sloan Management Review

Gängige Methoden zur Datensynthetisierung



Multivariate Distributions

Auto Encoders



kNN k-nearest-neighbour

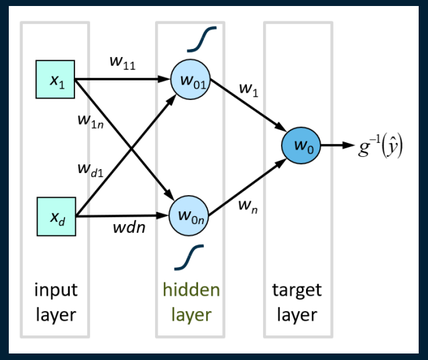
Generative Adversarial Network Action Set Action Set

Provides actions for training generative adversarial network models.

Syntax ▾ Details ▾ Examples ▾ References

Table of Actions

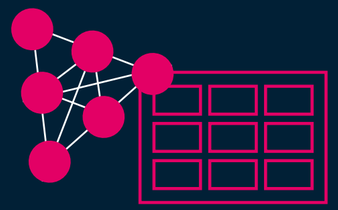
Action Name	Description
styleGanTrain	Trains a styleGAN model
tabularGanTrain	Trains a tabular GAN model



Generative Adversarial Networks

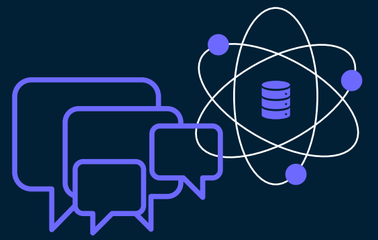
While there is no official or generally agreed upon definition of Generative AI (GAI) in the market, SAS considers these three AI/ML categories to be generative in nature

Synthetic Data Generation



Digital Twin

Large Language Models



A digital twin is…

A digital, animate, dynamic ecosystem – comprised of an interconnected network of software, generative & non-generative models, & (historical, real-time, & synthetic) data – that both mirrors & synchronizes with a physical system

Digital twins simulate "what-if" scenarios & stress test systems in the digital world to prescribe actions that optimize the physical world – to improve the lives of individuals, populations, cities, organizations, the environment, systems, products, & more

SAS success stories for digital twins

Digital Twins

MANUFACTURING

Georgia-Pacific: Optimizing the supply chain with analytics & IoT

"Because SAS makes it easy for people to do their own modeling, it's translated into millions of dollars in value for us." - GP

150
manufacturing facilities

85K
vibration sensors

1TB
data generated / day

15K+
models to calculate optimal production settings based on current business needs

10%
improvement in overall equipment efficiency

30%
reduction in unplanned downtime

Click here to read more

MANUFACTURING

USG Corporation: Optimizing time to value, production & modern manufacturing with analytics

Testing materials to ensure products meet quality standards:

Before SAS
24+
hours to test

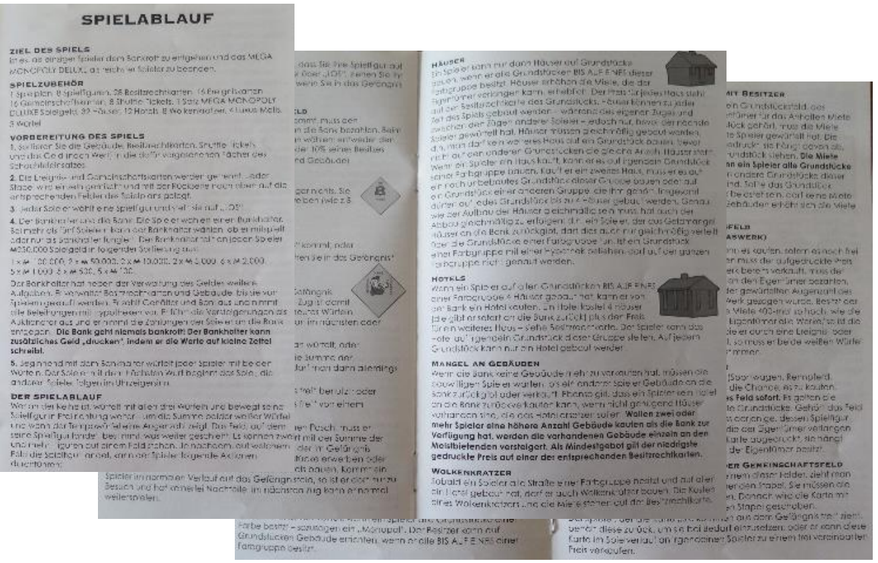
With SAS
Real-time
testing

With SAS:

- *analyze plant inputs, such as flow rates & raw material additives, to predict quality outcome before production even starts*
- *improve the quality, efficiency, safety, & cost of products*

Click here to read more

Das Monopoly Spiel ist ein komplexes System


SPIELABLAUF

ZIEL DES SPIELS

VORBEREITUNG DES SPIELS

DER SPIELABLAUF

HÄUSER

HOTELS

MANGEL AN GEBÄUDEN

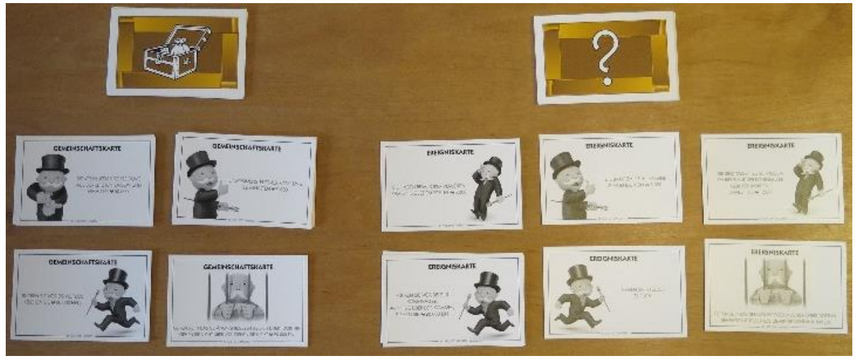
WOLKENKRATZER

MIT BESITZEN

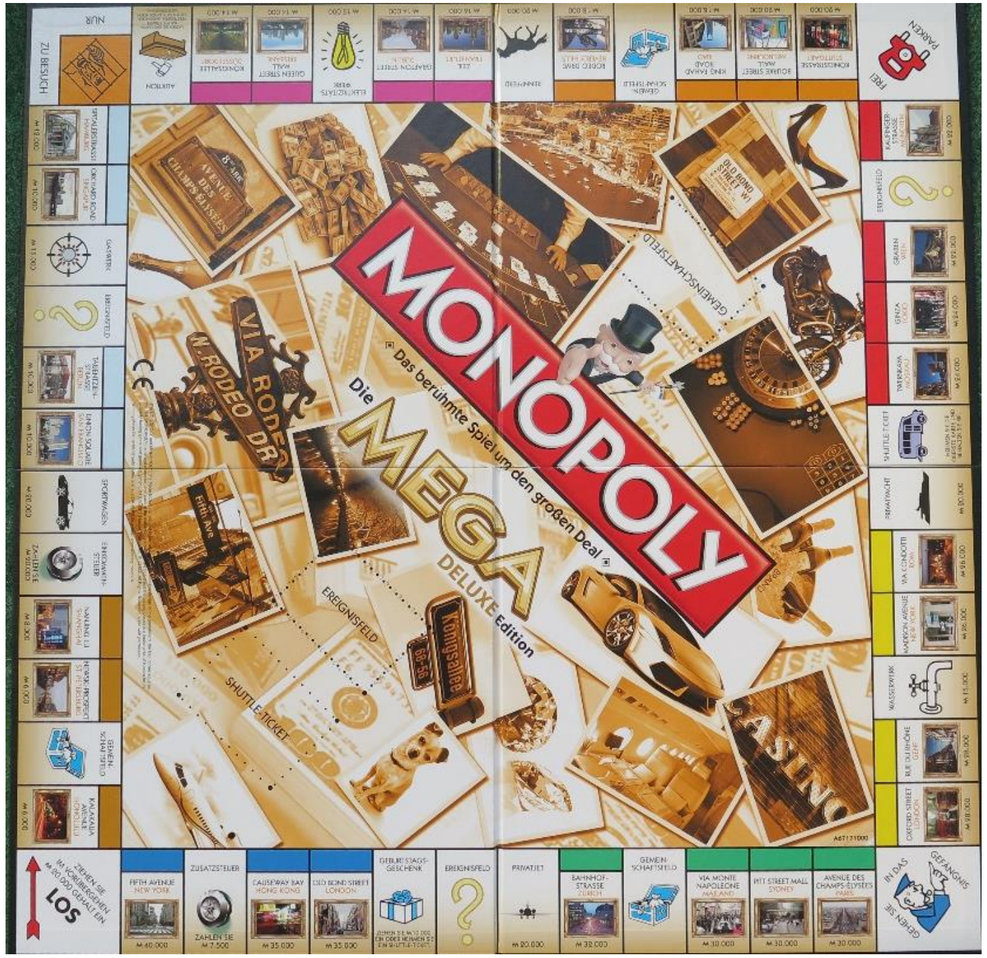
FELD ABWERB)

ER GEMEINSCHAFTSFELD

Komplexe Regeln



Zusätzliche Anweisungen



Rahmenwerk von Möglichkeiten und Ereignissen

Monetäre Dimension

Dynamische Komponenten

Zufällige Komponenten

Applying Generative AI in SAS Software

File Edit Selection View Go ··· [Extension Development Host] explore2023

kidneys2min.sas ● | kidneys2.sas | Settings | setup.sas

demos > kidneys2min.sas

Generate code | Add comments

```
1  /* Modify the kidneyTest table to convert the character variable potassiumchar to a numeric variable called potas
2
3  /* Create the month and year variables using the date variable, DATE, in the kidneyTest table  */
4
5  /* Use SQL to create the kidneyData in the public library by joining the kidneyPatientinfo and kidneyTest using i
6
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

SAS Log

0 ⚠ 0 SAS

Ln 1, Col 1 Spaces: 4 UTF-8 CRLF SAS