THINK GLOBAL, COMPUTE LOCAL

Ein Tech-Stack für PsychologInnen

Graz, 24.6.2025

DI:PS Differential Psychology Graz
Institute of Psychology, University of Graz, Austria

COGNITIVE Science

CSS Meeting

Marcel Jud
Differentielle Psy.
Raum, 70/EG

marcel.jud@uni-graz.at

DI):(PS Differential Psychology Graz
Institute of Psychology, University of Graz, Austria

Funktion

- Mitarbeiter Zulassung Lehramt
- PTA
- Doktorand
- Verwaltung von Teilen der IT-Infrastruktur der Psychologie (Proxmox Server, Untersuchungsräume, …)

Interesse/Forschung

- Verbindung von Diagnostik und Testkonstruktion mit maschinellem Lernen/AI um Entscheidungsprozesse in der Studierenden- oder Personalauswahl zu unterstützen

DIPS


Künstliche Intelligenz
und Transhumanismus


Erfolg im
Lehramtsstudium
© Robert Kneschke / Adobe Stock


Selbst- und
Fremdeinschätzung
© cottonbro studio / Pexels


Kreativität und
Innovation
© GPT-4/DALL-E 3 | openAI generated image

DIPS


Künstliche Intelligenz
und Transhumanismus


Erfolg im
Lehramtsstudium
© Robert Kneschke / Adobe Stock


Selbst- und
Fremdeinschätzung
© cottonbro studio / Pexels


Kreativität und
Innovation
© GPT-4/DALL-E 3 | openAI generated image

Technische Infrastruktur in der Forschung

Von der IT als Ergänzung zur Grundvoraussetzung

- **Digitale Datenerhebung** (Large scale assessment)
  - online & mobile
- **Datenmenge & Komplexität**
  - Big Data & Psychoinformatik
  - Multimodale Daten (Video, Sprache, fMRI, Eye-Trackin, Co-Reg, …)
- **Automatisierung & Reproduzierbarkeit**
  - Autom. Auswertung
  - Versionierung
  - Open Science
- **KI & ML**
  - Predicitve Modelling auf individueller Personenebene
  - NLP
- **Datensicherheit- und Schutz**
  - Encryption & Backup, Zugriffskontrolle

Psychology x Tech

Human-in-the-loop systems

Behavioral Data Science

Computational Psychology

Cognitive Computing

Psychoinformatik

AI in Psychology

Applied ML in Psychology

Predictive Modelling

Psychological Data Science

Personality Computing

Ai aided test construction

AI in Psychology

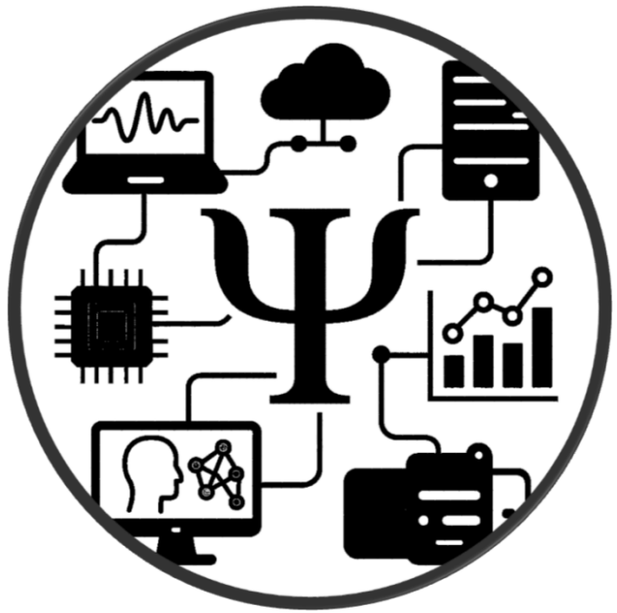
Computational Social Science

Human-AI Interaction

DIPS-relevant → **ML Fokus**

AI/digital Tech in Ψ (Auswahl)

- Item- und Testgenerierung mit KI (bei uns v.a. SJTs zu Personality, wie Big-5; Offenheit f. Vielfalt, …)
  - Parallelversionen bestehender SJTs v.a. im Persönlichkeitsbereich
  - Adaptierung von Schwierigkeit bzw. fakability bestehender SJTs
  - Neue SJTs from scratch
- Klassifikation & Regression (z.B. Dropout/Studienerfolg)
  - Meta-Model zur Vorhersage von Studienabbrüchen
  - Vorhersage von Faking in Persönlichkeitsinstrumenten
- Replikation & Kreuzvalidierung
  - Übersetzung von ML-Metriken nach Cohen‘s d (Salgado, 2018)



Unterschiede zu klassischen Methoden

- Valide Vorhersagen auf individueller Personenebene
  - Modelle, die generalisieren und so neue Daten valide vorhersagen
  - Gute Leistung bei multinominalen Kriterien (Lables) und breiten Datensätzen
- Kreuzvalidierung (p-hacking ausschließen)
- Reproduzierbarkeit (Resampling Methoden)
- Skalierbarkeit (je mehr Trainingsdaten, desto valider die Vorhersagen)

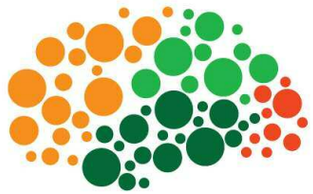
Orru et al. (2020)

Der DIPS- “Tech-Stack”

Beispiele unserer Infrastruktur (Auswahl an Open-Source tools)

Servermanagement (Container/VM)

proxmox.com



jspsych

jspsych.org/

Web/Browser Datenerhebung

- → Customizable, Plugin-based
- → Datenbankeinbindung
- → riesige Community
- → enorme Möglichkeiten zur Datenerhebung (z.B. Mouse-Tracking)

Frontend Data Science

RStudio Server

posit.co/download/rstudio-server/

Lokale LLMs

Ollama

ollama.com

Data Dashboards

- Dynamische Online-Reports
- Webapps
- → R-Backend

shiny.posit.co

ML Backend

TensorFlow

tensorflow.rstudio.com

tensorflow.org

Lokale LLMs

gpt4all

nomic.ai/gpt4all

Datenbasierte Auswertungsberichte,
Dynamische Dokumentation,
Reproduzierbarkeit
Open Science

PROXMOX

- Skalierbare virtuelle Umgebungen
- Ressourcenmanagement
- Container
- VMs

Ollama

Lokale, anonyme LLMs

RStudio Server

- Traditionelle Statistik
- Datenintensive Projekte (z.B. EyeTracking Co-Reg)
- LLM API calls
- Bays'sches Modelling
- Machine Learning (Supervised)
- Dokumentation (Analysen, Projekte,…)
- Shiny Webapps (z.B. Rückmeldungen Aufnahmeverfahren)

TensorFlow

Supervised Learning

jspsych

- Browser-Paradigmen
- Online/Offline/lab
- 100% customizable

MySQL

Large Scale Assessment

open source

AI basierte Item Generierung

Mittels fine-tuned LLMs

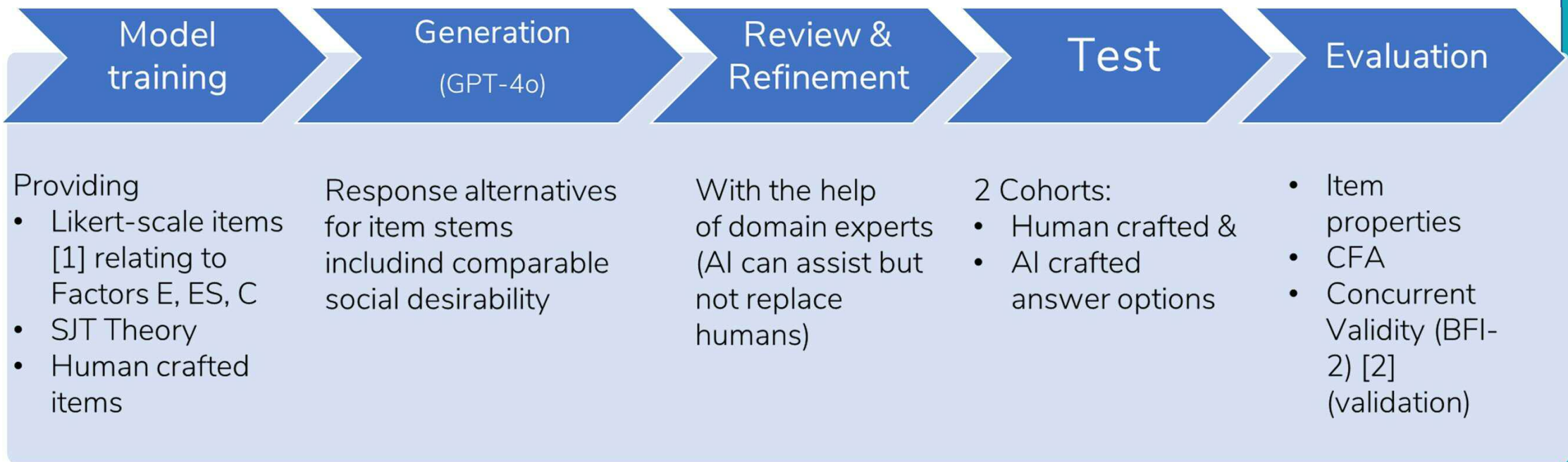
AI-Driven SJT Refinement

SJT fakability influenced by social desirability along responses

(Weekley & Ployhart, 2013)

Big-5 Factors: Gewissenhaftigkeit, Extraversion, Emotionale Stabilität

Ziel: Itemschwierigkeit erhöhen & Fakability einschränken



[1] International Personality Item Pool (IPIP; Goldberg, 1999)

[2] Danner et al., 2019)

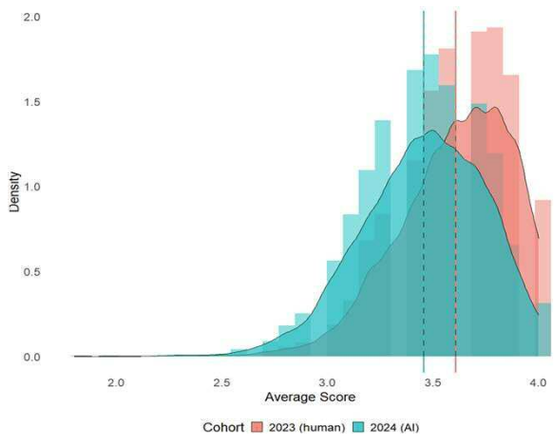
Ai-Driven SJT Refinement

Results in brief

- Signifikant kleinere Mean-Scores
- Gleiche model fit
- Gleiche Reliabilität (Omega)

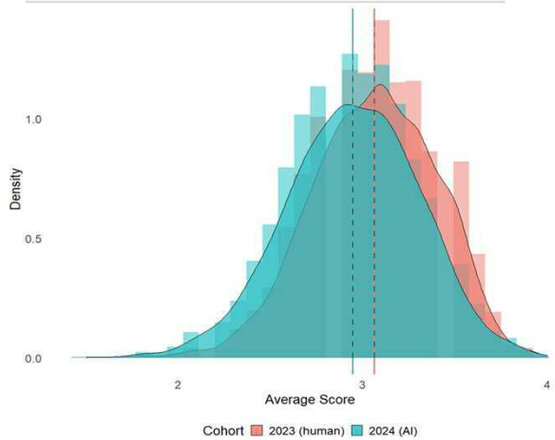
Conscientiousness (human vs AI)

Cohort	Mean *	SD	Omega
human	3.61	0.27	0.61
AI	3.45	0.29	0.53



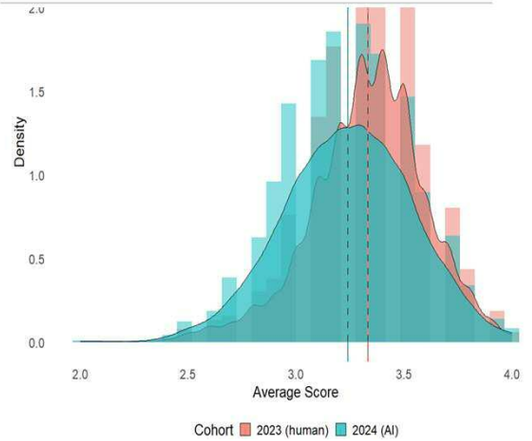
Extraversion (human vs AI)

Cohort	Mean *	SD	Omega
human	3.06	0.34	0.68
AI	2.95	0.36	0.65



Emotional Stability (human vs AI)

Cohort	Mean *	SD	Omega
human	3.33	0.26	0.65
AI	3.24	0.30	0.68



*Note. *N* = 5555 (2829 AI based) ; *d* = .33 - .54

Ai-Driven SJT Refinement

Concurrent Validity

	BFI_E	BFI_A	BFI_C	BFI_S	BFI_O	SJT_B5_E	SJT_B5_C
BFI_A	0.33***						
BFI_C	0.41***	0.46***					
BFI_S	0.49***	0.38***	0.44***				
BFI_O	0.39***	0.29***	0.29***	0.27***			
SJT_B5_E	0.60***	0.20***	0.24***	0.36***	0.31***		
SJT_B5_C	0.24***	0.23***	0.46***	0.22***	0.08***	0.24***	
SJT_B5_S	0.37***	0.25***	0.32***	0.54***	0.20***	0.37***	0.31***

Note. SJT scale range is from 1 to 4 and BFI-2 scale range is from 1 to 5. * indicates $p < .05$. ,** indicates $p < .01$., ,** indicates $p < .001$

Personality SJT Generation

Gleiche Methodik wie bei Refinement

Offenheit

Subscale	Mean	SD	Omega	Alpha
Offenheit Auswahl	2.73	0.31	0.65	0.65

Chisq	df	Chisq_df	p	CFI	RMSEA	SRMR
55.278	35	1.58	0.016	0.969	0.038	0.071

Verträglichkeit

Subscale	Mean	SD	Omega	Alpha
Verträglichkeit Auswahl	2.9	0.26	0.6	0.59

Chisq	df	Chisq_df	p	CFI	RMSEA	SRMR
50.489	35	1.44	0.044	0.97	0.031	0.081

Itembeispiel

Openness (Intellectual curiosity)

You are traveling to a foreign country and are invited by locals to take part in a traditional festival. You don't know what to expect. How do you react?

(a) You take part, observe the festivities and ask about the significance of the traditions. [3]
(b) You thank them for the invitation and politely decline. [1]
(c) You accept the invitation to experience the culture at first hand and take an active part in the festivities.[4]
(d) You don't want to be a burden to the locals at this festival. You thank them for the invitation but remain on the sidelines and only observe from a safe distance.[2]

Offenheit f. Vielfalt SJT

Gleiche Methodik wie zuvor

Durchschnittliche Item Schwierigkeit und Trennschärfe

Kennwert	Menschliche Items	KI Items
Item Schwierigkeit (*p*)	.79 (*SD* = 0.07)	.79 (*SD* = 0.12)
Trennschärfe (r_{it})	.36 (*SD* = 0.10)	.37 (*SD* = 0.06)

Goodness-of-Fit-Indizes für ein einfaktorielles Modell aller 18 OfV Mensch-Items sowie aller 17 OfV KI-Items (n = 202)

Quelle	*df*	$\chi 2$	*p*	χ^2/df	CFI	TLI	RMSEA	SRMR
Mensch	135	102.89	.996	0.76	1.000	1.083	.000	.062
KI	119	82.45	.982	0.69	1.000	1.090	.000	.064

Interkorrelationen von OfV Mensch mit OfV KI und den Big Five

			Big Five				
	OfV - Mensch	OfV - KI	Offenheit	Gewissenhaftigkeit	Verträglichkeit	Extraversion	Neurotizismus
OfV- Mensch	-	.72***	.22***	.02	.35***	.17*	-.06
OfV - KI	.72***	-	.21**	.12	.31***	.08	-.05

Anmerkungen. *** $p \leq .001$.

Implikationen

Spezifische Situational Judgement Tests

- (Semi-)automische Generierung valider und realistischer Szenarios
- Zeit und Kosten sparen ohne Inhaltsvalidität zu gefährden
- Faking-good limitieren
- Itemschwierigkeit adaptieren
- Hohe Augenscheinvalidität → Akzeptanz

Ausflug: Prompt Engineering

LLM output ist wesentlich durch Prompt-Qualität bestimmt

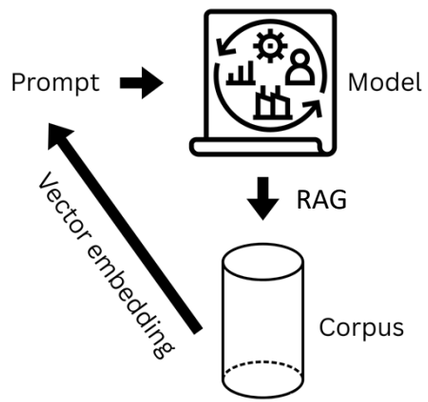
- **Garbage In - Garbage Out:** Unklare Prompts = schlechte Ergebnisse. Klarheit & Kontext sind entscheidend.
- **Klarheit & Spezifität:**
  - Klar formulieren: Vermeiden von vagen Begriffen oder allgemeinen Eingaben.
  - Spezifizieren: Statt „Erklär mir X“ → „Erklär mir X, mit Fokus auf Xy für Zielgruppe Z mit 3 Beispielen.“
- **Struktur:**
  - Gliederung komplexer Prompts in Teilschritte.
  - Nutzen von Listen, Absätzen, Trennzeichen (###).
- **Zielgruppenfokus & Rolle:**
  - Definieren: Wer soll den Output verstehen? (z. B. Schüler:in, Expert:in)
  - Nutzen von Systemprompts, z. B. „Du bist ein hilfsbereiter Tech-Support-Agent.“
- **Prompt-Schleifen (Iterationen):**
  - 1. Prompt = Rohfassung
  - 2. Feedback geben: „Schreib kürzer“ / „Nutz anderen Ton“ / „Füge Beispiele ein“ / Kontext geben / Websuche
- **Formatierung & Wiederholung:**
  - Hervorhebungen & klare Tags (z. B. ### Aufgabe ###)
  - Schlüsselbegriffe gezielt wiederholen

Prompt Engineering

Fortgeschrittene Strategien

Fortgeschrittene Strategien:

- RAG: Retrieval Augmented Generation. z.B. aus web search / deep research / file



- Kettung (Prompt Chaining): Aufgabe in Sequenzen aufteilen
- Few-Shot-Learning: Gute Beispiele als Input beifügen (z.B. Items)
- Explizite Formatvorgaben: z. B. „Antwort als Tabelle mit drei Spalten“ / „Ausgabe in Code Junk“ / „strukturiere nach JSON“

Prompt Engineering

Beispiel

Rolle
Du bist eine erfahrene Psychologin mit Spezialisierung auf Kognitionspsychologie und digitale Didaktik.
Aufgabe
Erkläre evidenzbasierte Prinzipien, wie MOOCs (Massive Open Online Courses) gestaltet werden sollten, damit sie das Lernen und Behalten verbessern. Beziehe dich auf Erkenntnisse aus der Cognitive Load Theory, dem Multimedia Learning sowie dem Spacing Effect.
Zielgruppe
Studierende der Psychologie im Masterstudium, die an E-Learning-Forschung interessiert sind.
Format
Stelle die Antwort als strukturierte Liste mit drei Hauptprinzipien dar. Jeder Punkt soll enthalten:
- Die Bezeichnung des Prinzips (fett)
- Eine kurze Erklärung (max. 3 Sätze)
- Ein konkretes MOOC-Design-Beispiel
Einschränkungen
Verwende eine sachliche, gut verständliche Sprache ohne Fachjargon. Maximal 150 Wörter pro Prinzip.
Optional
Wenn möglich, nenne am Ende eine wissenschaftliche Quelle oder ein Review (max. 1 Satz). Recherchiere die Quelle mittels
Feedbackschleife
Wenn du noch etwas verbessern würdest, nenne mir drei Varianten mit unterschiedlicher Perspektive:
1. für Lehrende, 2. für Instructional Designer:innen, 3. für evidenzbasierte Policy Maker.

Prompt Engineering

Beispiel zu Deep Research Prompting

Rolle
Du bist eine wissenschaftlich arbeitende Psychologin mit fundierter Erfahrung in Meta-Analysen und quantitativer Synthese psychologischer Studien.
Aufgabe
Unterstütze bei der Planung einer Meta-Analyse zum Thema "[Wirkung digitaler Achtsamkeitstrainings auf Stresserleben bei Erwachsenen]". Gib einen detaillierten Vorschlag zur Vorgehensweise, inklusive:
1. Ein- und Ausschlusskriterien,
2. relevanter Moderatorenvariablen,
3. typischer Effektgrößen,
4. geeigneter Datenbanken und Suchstrings.
Zielgruppe
Psycholog:innen mit Forschungserfahrung (z. B. Prä- oder Postdocs), die eine systematische Meta-Analyse durchführen möchten.
Format
Gib die Antwort in strukturierter Listenform aus – pro Punkt max. 5 Sätze. Verwende Nummerierungen und Zwischenüberschriften. Nutze Fachbegriffe korrekt, aber erklärungsfreundlich.
Einschränkungen
Berücksichtige die Standards von PRISMA und APA (7. Aufl.). Maximal 350 Wörter. Verwende neutrale, präzise Sprache – kein KI-Marketing, keine synthetische Sprache, SCHREIBE IN EINEM AUTHENTISCHEN, MENSCHLICHEN STIL.
Beispiel (Few-Shot)
Ein-/Ausschlusskriterien:
- Eingeschlossen: randomisierte kontrollierte Studien (RCTs), Erwachsene (18+), Interventionen mit ≥ 3 Sitzungen, Peer-reviewed.
- Ausgeschlossen: qualitative Studien, Meditation-only-Apps ohne psychologisches Begleitmaterial, Pilotstudien ohne Kontrollgruppe, Predatory Journals.
Optional
Schlage am Ende ein geeignetes Statistikpaket oder eine Bibliothek (R, Python) für die Berechnung gemittelter Effektgrößen (z. B. Hedge's g) vor. Inkludiere relevante Code-Junks
Feedbackschleife
Wenn nötig, bitte alternative Designs für Meta-Regression, Netzwerk-Metaanalyse oder eine systematische Review mit narrativer Synthese vorschlagen – je nach Studiensituation.

AI basierte AUT ratings

Mittels fine-tuned LLMs

Automatisches AUT Scoring (Saretzki et al., 2025)

mittels LLMs

Table 4. Rater statistics including LLMs.

	LLM(s) as Weakest Rater	LLM(s) Neither Weakest nor Best	LLM(s) as Best Rater	Mean LLM-Total Correlation
CLAUS	80.0%	20.0%	-	0.66
CLAUS (Translation)	80.0%	20.0%	-	0.69
OCSAI	20.0%	40.0%	40.0%	0.80
OCSAI (Translation)	40.0%	60.0%	-	0.71
GPT-4	20.0%	80.0%	-	0.74
GPT-4 (Translation)	60.0%	40.0%	-	0.72
Average	50.0%	43.3%	6.7%	0.71

Notes. LLMs = large language models. Relative frequencies always refer to the relevant total of the number of rater statistics including datasets with three or more raters (i.e., Studies 1, 5, 6, 7, and 8).

Supervised Learning

Individuelle Kriteriums-Vorhersagen

Psychological Assessment x AI/Machine Learning

Was ist eigentlich Machine Learning?

➡ Einfache Definition:

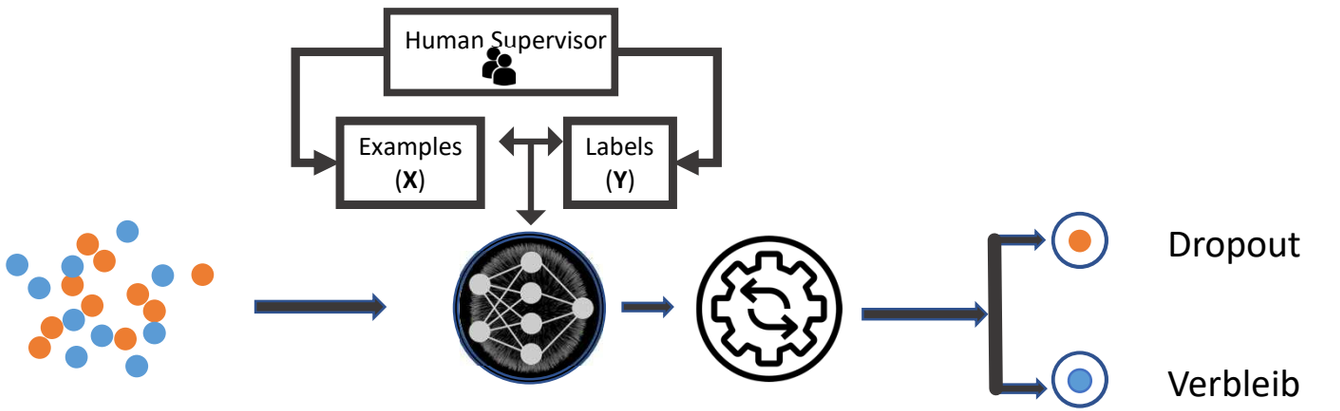
01 Ein Algorithmus, der input-output Relationen durch vergangene Erfahrung lernt und

02 seine Leistung durch weitere, neue Erfahrung (Beispiele/Training data) verbessert

"Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed."
— Arthur Samuel, 1959

Was ist eigentlich Supervised Learning?

Aufgaben, bei denen eine Funktion gelernt wird, die Input-Output Relationen abbildet und zwar auf Basis von Input-Output Paaren (features & lables) (Russell & Norvig, 2020)



Examples /
Training data =
Input (features)

e.g. admission data

ML Algorithmus

Fitting

Iterative
training
with
error
feedback

Output /
Prediction
(Labels)

Machine Learning in R mit Tensorflow

Möglichkeiten

Ressourcen: hier

Machine Learning in R mit Tensorflow

R-Studio → Keras (interface) → Tensorflow

- Rstudio öffnen
- Keras installieren

  # install.packages("remotes")
  remotes::install_github("rstudio/tensorflow")
  reticulate::install_python()
  Oder
  install.packages("keras")
  library(keras)
  install_keras()
- Package einbinden

  library(tensorflow)
  library(keras)
- Installation testen

  tf$constant("Hello TensorFlow!")
- Code rechts ausprobieren

Ressourcen: hier

```r
# Load keras & tf
library(keras)
library(tensorflow)

# Definiere Architektur des Neural Net
model <- keras_model_sequential() %>%
layer_dense(units = 128, activation = 'relu', input_shape = c(784)) %>%
layer_dropout(rate = 0.4) %>%
layer_dense(units = 64, activation = 'relu') %>% layer_dropout(rate = 0.3) %>%
layer_dense(units = 10, activation = 'softmax‘)

# Kompilieren
model %>% compile( loss = 'categorical_crossentropy', optimizer =
optimizer_rmsprop(), metrics = c('accuracy‘) )

# Trainon a dataset (e.g., MNIST)
mnist <- dataset_mnist() x_train <- mnist$train$x / 255 y_train <-
to_categorical(mnist$train$y, 10) x_test <- mnist$test$x / 255 y_test <-
to_categorical(mnist$test$y, 10)

# Reshape in Trainings- und testdaten
x_train <- array_reshape(x_train, c(nrow(x_train), 784))
x_test <- array_reshape(x_test, c(nrow(x_test), 784))

# Fit the model model %>% fit( x_train, y_train, epochs = 30, batch_size = 128,
validation_split = 0.2 )

# Model testen
model %>% evaluate(x_test, y_test)
```

Student Dropout Prediction (Jud et al. 2025)

Basierend auf Zulassungsverfahren (TESAT)* & Studienverlauf

Ziel: Vorhersage von Studienerfolg/Dropout basierend auf Zulassungsverfahren, Demografie & GPA

Features (Prädiktoren): z.B. intelligence, verbal ability, conscientiousness (r = -.25), emotional stability (r = -.35) (Neubauer et al. 2017; Weissenbacher et al. 2024).

10-fold Cross-validation based on 2661 students

‘*Prediciton-over-explanation*’ approach → Anwendungsbezogen (vs. erklärend)

*(Neubauer et al. 2017)

Student Dropout Prediction

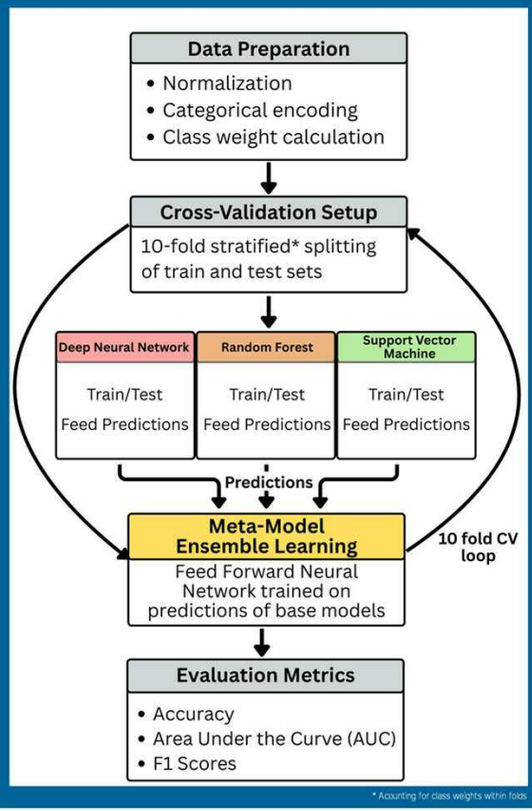
(Jud et al. 2025)

Basierend auf Zulassungsverfahren (TESAT)* & Studienverlauf

Table 1: Classification performance across all models (base learners and meta-learner) using 10 cross-validation folds (dropout yes/no)

Dataset	Model	Mean Accuracy	Maximum Accuracy	Mean AUC	Cohen's d (AUC)
Entrance Exam Data	Neural Network	0.51	0.60	0.54	0.14
	Random Forest	0.58	0.61	0.57	0.25
	SVM	0.59	0.61	0.57	0.25
	Meta model (FFNN)	0.61	0.67	0.62	0.43
GPA Data	Neural Network	0.54	0.73	0.63	0.47
	Random Forest	0.66	0.69	0.72	0.82
	SVM	0.61	0.66	0.67	0.62
	Meta model (FFNN)	0.74	0.78	0.76	0.99
Combined Entrance Exam and GPA	Neural Network	0.58	0.73	0.63	0.47
	Random Forest	0.67	0.72	0.73	0.87
	SVM	0.63	0.74	0.72	0.82
	Meta model (FFNN)	0.76	0.79	0.77	1.08

Figure 1: Framework architecture



Prestudy (Jud, 2022)

Personality Faking Detection

within MMPI-2 using supervised learning (neural nets)

Main scale: Features: MMPI-2 L, K, S T-Scores

Social desirability: PPI-R VR Skala

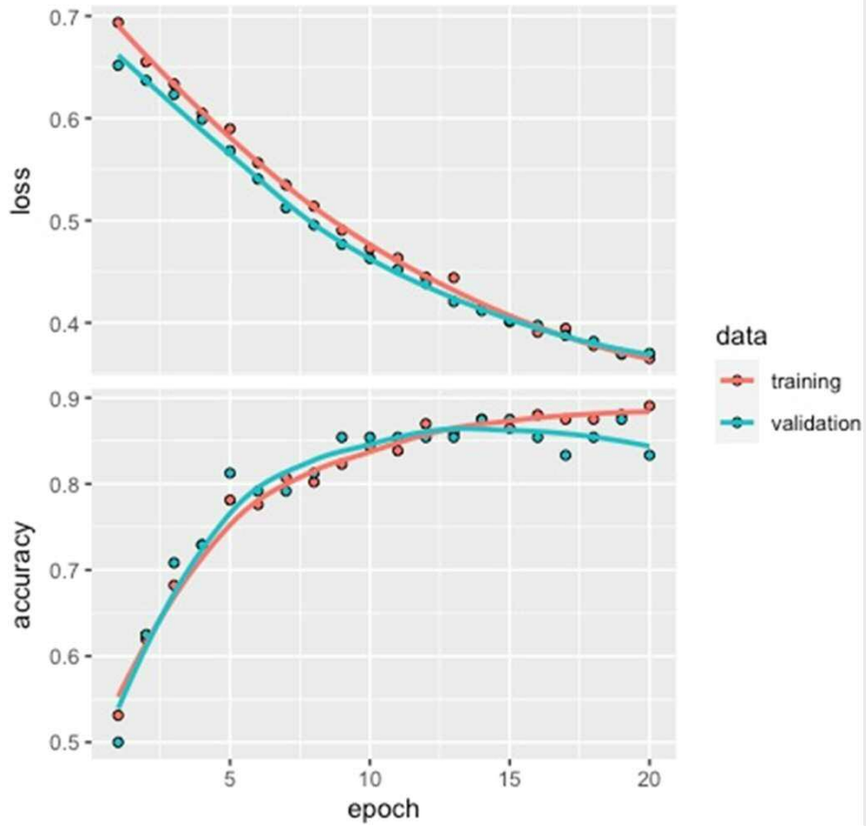
Mouse features: Zeitlich & räumlich

- L, K, S, and VR Mittlere Abweichung (MD)
- L, K, S and VR Fläche der Abweichung (tAUC)
- L, K, S, und VR RTs;
- L, K, S, and VR Abweichungszeit (MD-time)
- L, K, S, and VR Geschwindigkeiten (vel_x, vel_y)

Ergebnisse: Test-set approach

Neural Network Leistung

- Validation set (all features): acc = 85% (just 20 epochs)
- Lie-Scales (both speed conditions): 75%
- Lie Scales speeded: ~68%
- Lie Scales unspeeded ~66%



n = 240; Feed forward deep neural network
Data source: Mazza et al., 2020 (lab setting)

Implikationen

für andere Kriterien

Supervised ML models, trainiert mit Assessment-Daten (e.g., SJTs, personality) können:

- Leistung/Outcomes von BewerberInnen vorhersagen
- Faking-good in persönlichkeitsbezogenen Merkmalen erfassen

→ und unterstützen so Entscheidungsprozesse bzw.

→ können für frühe Warnungen (z.B. Dropout) herangezogen werden

Einschränkung: *Umfang und Qualität der Trainingsdaten*

Dualistic approach

Durch die Kombination von traditionellen Zugängen & ML können wir:

- individuelle Outcomes vorhersagen

- In Kombination mit generativer Itemkonstruktion
  → skalierbare, faire und empirisch fundierte Selektion in high stakes Szenarios

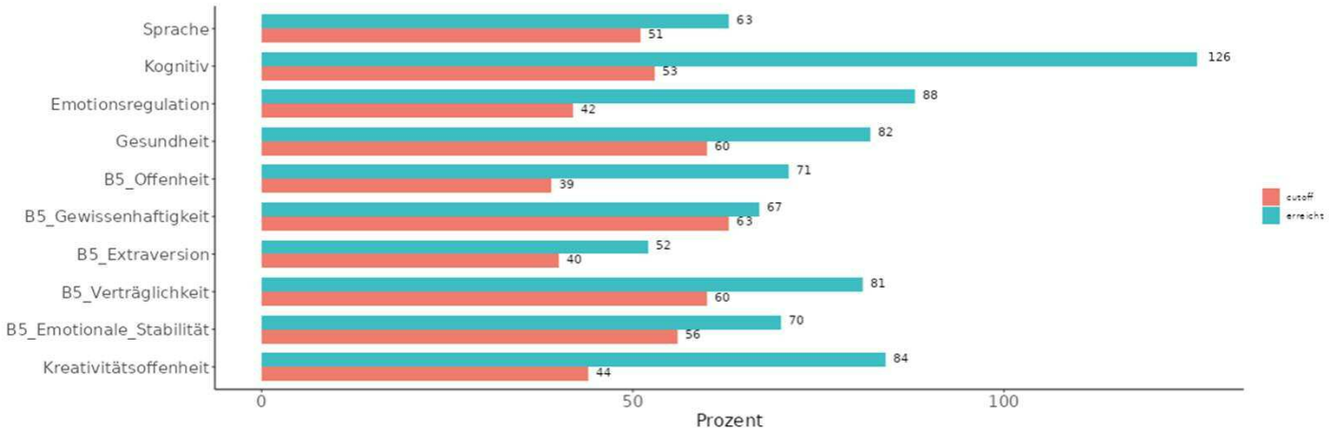
Reporting

Datenbasiert & dynamisch

Person

Participant	Vorname	Nachname	Prüfungsort	Eignung
LSG6eMQ7Wv	[redacted]		[6] Innsbruck (an PH Tirol): 26.05.-03.06.2025	1.00

Individuelles Profil



Gesamtergebnis

[Kein Titel]

Sprachliche Ressourcen: bestanden

Kognitive Ressourcen: bestanden

Persönliche und emotionale Ressourcen: bestanden

[redacted] hat den computergestützten Eignungstest somit insgesamt **bestanden.**

Ergebnisbericht für [redacted]

Zur vereinfachten Interpretation wurden die erreichten Punktwerte sowie die zu erreichenden Mindestwerte (Cut-offs) für die Rückmeldegespräche in Prozentwerte umgerechnet. Achtung: bei den Cut-Offs handelt es sich um die derzeit – für das Aufnahmeverfahren 2024 – geltenden Werte, diese sind nicht auf Folgejahre übertragbar.

Bereich 1: Sprachliche Ressourcen – Sprachkompetenz

Sprachkompetenz, die im Aufnahmeverfahren mittels zweier Subtests – **Rechtschreibung und Grammatik** sowie **Leseverständnis** – überprüft wurde, gilt als **zentrale Voraussetzung** für den Erfolg im Studium und Lehrer*innenberuf. Sprachliche Fähigkeiten sind nicht nur für das (Lehramts-)Studium relevant, sondern auch für den Beruf. Lehrkräfte vermitteln beispielsweise Unterrichtsinhalte, geben Schüler*innen Anweisungen oder verfassen Arbeitsunterlagen; all diese Aufgaben erfordern gute sprachliche Fähigkeiten. [redacted] hat im Bereich Sprachkompetenz **63** Prozent erzielt. Damit wurde der erforderliche Mindestprozentsatz in diesem Bereich überschritten.

Bereich 2: Kognitive Ressourcen – Kognitive Leistungsvoraussetzungen

Unter *kognitiven Leistungsvoraussetzungen* werden unterschiedliche Fähigkeiten verstanden, die insbesondere für ein erfolgreiches Absolvieren des Lehramtsstudiums notwendig sind. Im Aufnahmeverfahren wurden dabei drei Bereiche überprüft: **numerische, verbale** und **figurale Fähigkeiten**. Numerische Fähigkeiten wurden mithilfe des Tests „Zahlenreihen fortsetzen" erfasst, verbale Fähigkeiten mithilfe des Tests „Gemeinsamkeiten finden" und figurale Fähigkeiten mithilfe des Tests „Papier falten". [redacted] hat im Bereich Kognitive Ressourcen **126** Prozent erzielt. Damit wurde der erforderliche Mindestprozentsatz in diesem Bereich überschritten.

Gesundheits- und Erholungsverhalten

Gesundheits- und Erholungsverhalten ist als relatives stabiles Persönlichkeitsmerkmal („Trait") zu sehen und beschreibt die Bereitschaft, sich (vor allem bei körperlichen Warnsignalen) **gesundheitsbewusst zu verhalten** sowie die Fähigkeit, sich **körperlich und psychisch entspannen** zu können. Dabei können die drei Komponenten „Gesundheitsverhalten", „Entspannungsfähigkeit" sowie „gedankliches Abschalten" unterschieden werden. Lehrkräfte sind (zumindest zeitweise) einer hohen Stressbelastung – und damit einhergehend – einer erhöhten Burnout-Gefahr ausgesetzt. Daher ist es entscheidend, dass sie sich in ihrer Freizeit erholen und von ihrer Arbeit abgrenzen können sowie rechtzeitig erkennen, wenn Handlungsbedarf in Hinblick auf ihre Gesundheit besteht. [redacted] hat im Bereich Gesundheits- und Erholungsverhalten **82** Prozent erzielt. Damit wurde der erforderliche Mindestprozentsatz in diesem Bereich überschritten.

Offenheit

Offenheit beschreibt das Interesse und das Ausmaß der **Beschäftigung mit neuen Erfahrungen, Erlebnissen** und **Eindrücken**. Personen mit hoher Offenheit beschäftigen sich gerne mit intellektuellen Tätigkeiten und sind wissbegierig. Außerdem sind sie fantasievoll und aufgeschlossen gegenüber fremden Kulturen und Werten. Ein hohes Maß an Offenheit ist besonders für Ausbildungen, Weiterbildungen und Fortbildungen in verschiedensten Bereichen hilfreich. [redacted] hat im Bereich Offenheit **71** Prozent erzielt. Damit wurde der erforderliche Mindestprozentsatz in diesem Bereich überschritten.

Shiny

“minimal example”

User Interface

Server

```
library(shiny)

ui <- fluidPage(
  # 1. user enters input into this text box
  textInput(
    inputId = "my_input",
    label = "Text Input"
  ),

  # 4. we print the object returned with output$my_output here
  verbatimTextOutput("my_output")
)

# 2. the user's input from 1. is passed to the input argument in the below function as
# input$my_input
server <- function(input, output) {
  # 3. we handle the input however we want on the server.  Here
  # we are just passing it back to ui with a renderPrint to output$my_output
  output$my_output <- renderPrint({
    input$my_input
  })
}

shinyApp(ui, server)
```

Ressourcen: https://mastering-shiny.org/index.html

Shiny

Beispiel

https://gallery.shinyapps.io/collinearity/

Ressourcen: https://bookdown.org/yihui/rmarkdown/

Dokumentation

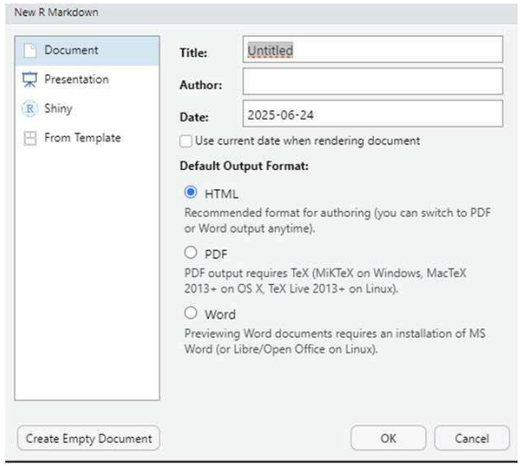
Forschung, Lehre, Abschlussarbeiten

RMarkdown

Mini Beispiel

1. RStudio → File → New Markdown
   - Output: HTML

   → Beispielskript erscheint




```
---
title: "Untitled"
output: html_document
date: "2025-06-24"
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax
for authoring HTML, PDF, and MS Word documents. For more details on using
R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that
includes both content as well as the output of any embedded R code chunks
within the document. You can embed an R code chunk like this:

```{r cars}
summary(cars)
```

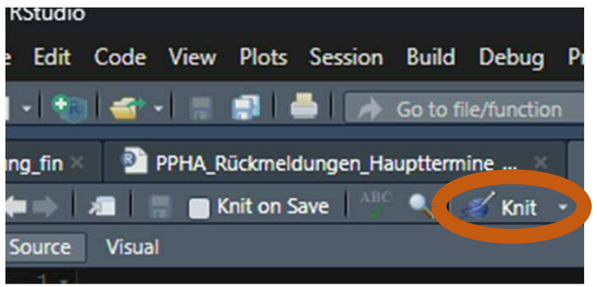
Including Plots

You can also embed plots, for example:

```{r pressure, echo=FALSE}
plot(pressure)
```

Note that the `echo = FALSE` parameter was added to the code chunk to
prevent printing of the R code that generated the plot.
```

2. Knit



Ressourcen: https://bookdown.org/yihui/rmarkdown/

RMarkdown

Mini Beispiel

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

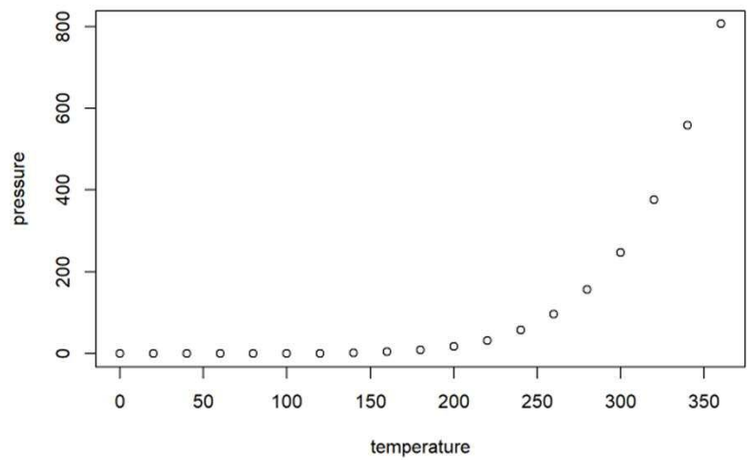
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Ressourcen: https://bookdown.org/yihui/rmarkdown/

- Eine kurze Einführung
- Latent Semantic Analysis
  - Begriffe
  - Cosinus-Distanz
  - Erstellung eines eigenen Semantic Space
  - Vom Corpus zur Document-Term-Matrix
  - Singular Value Decomposition
  - **Distanzmetriken LSA**
  - Automatisches Scoren von MC Aufgaben
  - Zusammenfassung von Texten
- word2vec Embeddings
- gloVe

Distanzmetriken LSA

Exploriere Ähnlichkeiten in **Topic Encoded** Daten:

Cosinus

```
Cosine("[redacted]", "statistik", tvectors = lsa_space$dk)
```

```
## [1] 0.791227
```

```
Cosine("[redacted]", "bestehen", tvectors = lsa_space$dk)
```

Name der prüfenden Person

```
## [1] -0.01305072
```

```
Cosine("[redacted]", "durchgefallen", tvectors = lsa_space$dk)
```

```
## [1] 0.515336
```

Nachbarn

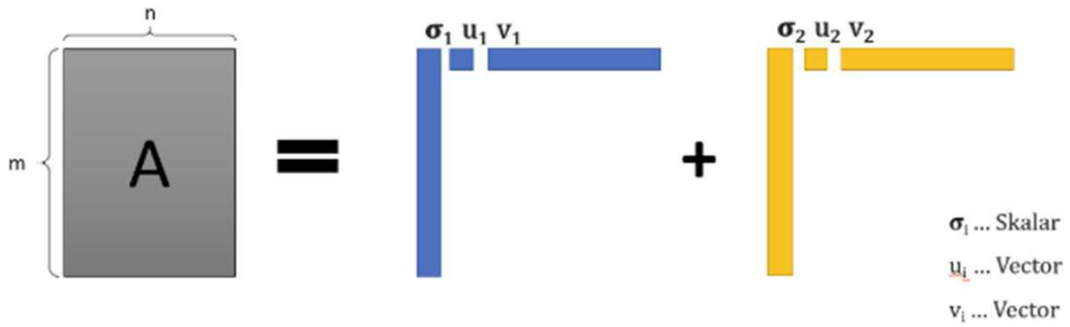
```
neighbors("neubauer", tvectors = lsa_space$dk, 30)
```

```
##                  neubauer          differentiellen
##                 1.0000000                0.7444507
##                     markus             markusthaler
##                 0.7444507                0.7444507
##                    thaler              bonuspunkt
##                 0.7444507                0.6923300
##  aufmerksamkeitsprozessen           differentielle
##                 0.6596639                0.5899875
##                    tanzer           finkkoschutnig
##                 0.5629943                0.5180975
```

***** Die dargestellten Metriken entspringen einer semantischen Analyse auf Basis der Facebook Gruppe „Grazer Psychologiestudierende“

- Eine kurze Einführung
- Latent Semantic Analysis
  - Begriffe
  - Cosinus-Distanz
  - Erstellung eines eigenen Semantic Space
  - Vom Corpus zur Document-Term-Matrix
  - Singular Value Decomposition
  - Distanzmetriken LSA
  - Automatisches Scoren von MC Aufgaben
  - Zusammenfassung von Texten
- word2vec Embeddings
- gloVe

$$A = U\Sigma V^T = \Sigma_i \sigma_i u_i \circ v^T_i$$



***

LSA starten: **Singular Value Decomposition**

Package: lsa

```
library(tictoc) #runtime Bibliothek

tic()#starte Zeitmessung
lsa_space <- lsa(dtm_mat, dims =  dimcalc_share()) #dims*
toc() #675.54 sec elapsed
```

*dims = dimcalc_share(): Dimensionality Calculation Routines; Methods for choosing a `good' number of singular values for the dimensionality reduction in LSA.

Je nach Rechenleistung gibt die Funktion nach einigen Minuten eine Matrix zurück, die U, Sigma und V enthält. In unserem Fall (zur Bestimmung der Cosinus-Ähnlichkeit) sind die Singularwerte nötig; hier im Datensatz "dk" genannt. Um diese Sub-Matrix zu indizieren, wird der **$** Operator genutzt.

```
str(lsa_space)
```

```
## List of 3
##  $ tk: num [1:5897, 1:100] -7.29e-06 -5.67e-06 -2.97e-05 -7.62e-06 -1.50e-05 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:5897] "1" "2" "3" "4" ...
##   .. ..$ : NULL
##  $ dk: num [1:9494, 1:100] -7.70e-07 -1.40e-06 -9.51e-05 -2.81e-04 -1.51e-06 ...
```

Zugangsanfrage

für Rstudio Server bzw. Ollama

marcel.jud@uni-graz.at

- Kurze Mail an mich inkl. Hardware ID des Rechners (Standardgerät Uni Graz)
- Ich melde dann jeweilige IP Adressen und Benutzerdaten zurück

Einschränkungen

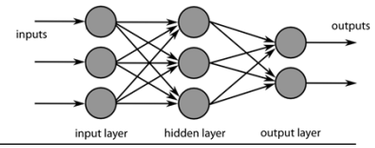
- Aus Sicherheitsgründen ist externer Zugriff gesperrt (nur über Uni Graz Netzwerk und nur via Standardgerät)
- Für die meisten Dinge sind zumindest Consolen-Kenntnisse erforderlich, in vielen Fälle auch Programmiererfahrung erforderlich
- Alternativ für ML mit GUI: WEKA (Data Mining & Machine Learning) https://waikato.github.io/weka-wiki/downloading_weka/

Ressourcen:

- R for Data Science (Book): https://r4ds.hadley.nz/
- Hands on Tensorflow (via keras): https://cran.r-project.org/web/packages/keras/vignettes/index.html
- Deep Learning with R: https://github.com/TrainingByPackt/Deep-Learning-with-R-for-Beginners
- Cheat Sheets R: https://posit.co/resources/cheatsheets/

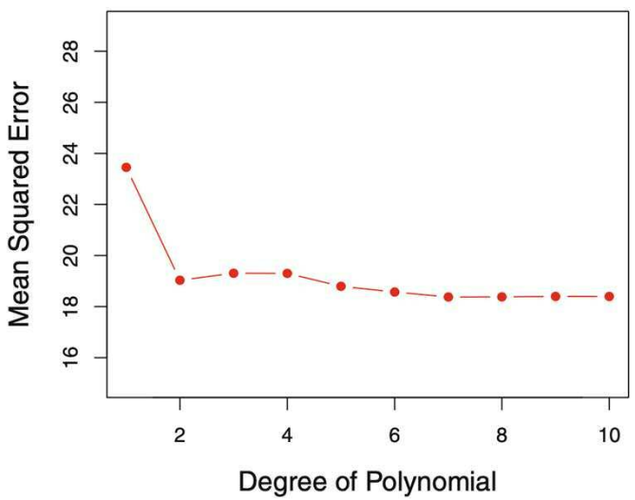
Appendix

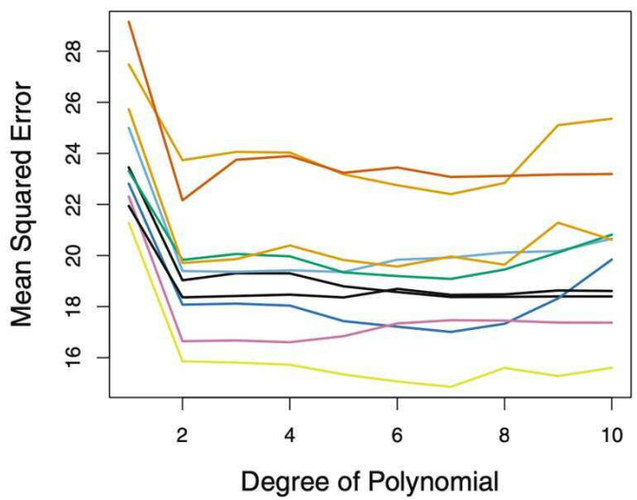
Simple validation:Test set approach



- **Model performance on unseen observations → real world performance aka generalization to new data**

In most cases the TSA is not sufficient because it derives only a single validation error estimate which heavily relies on the composition of subsets that randomly fall into training and test set slices → **unreliable performance measure**



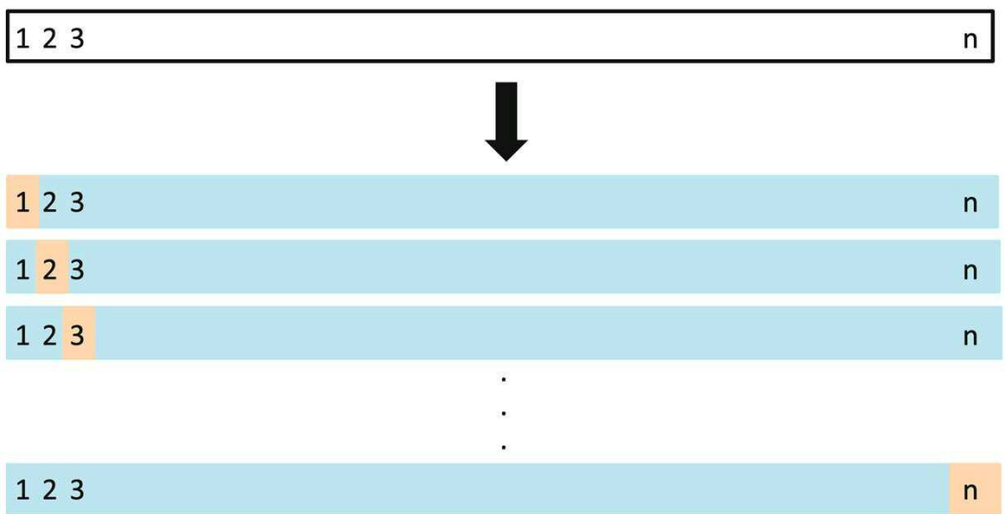


(James et al., 2022)

Resampling: Crossvalidation

Splitting data into k groups or **folds** → using each time (each fold / each iteration) a different group as a validation set and the remaining one as training set:

Leave One Out Crossvalidation (LOOCV)
→ special case of k-fold CV in which k equals N



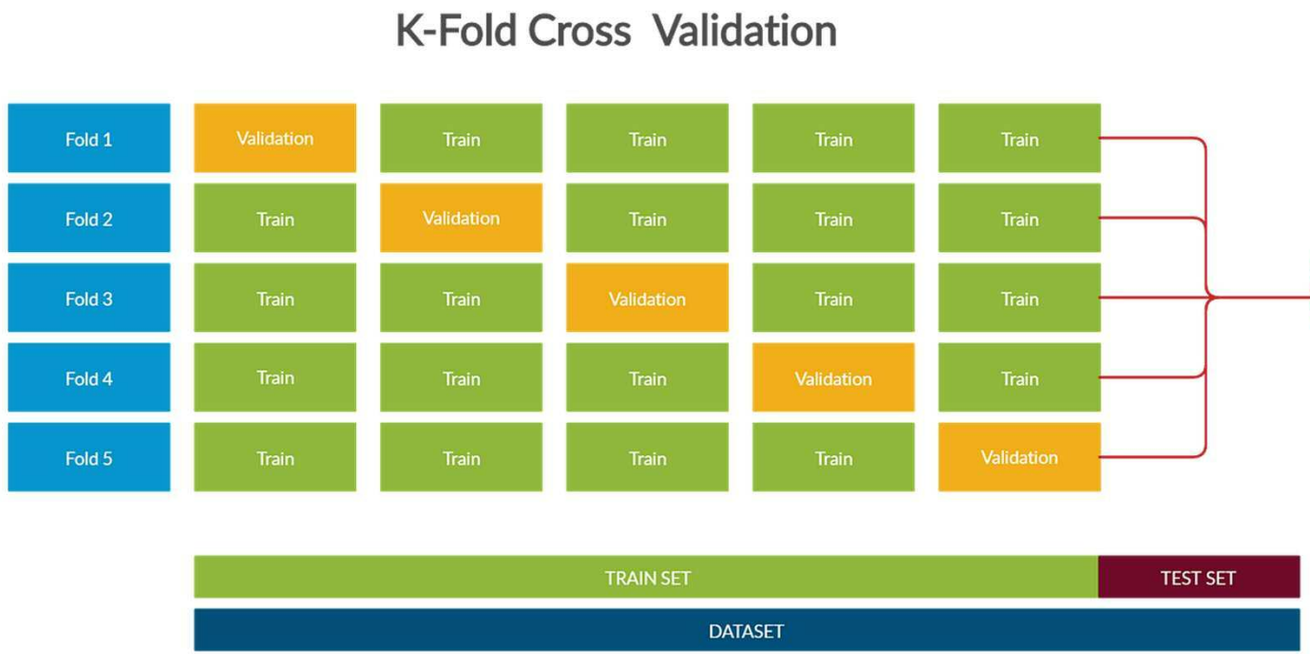
Drawbacks:

- **Computational expensive**
- **High prediction variance**

(James et al., 2022)

Resampling: Crossvalidation

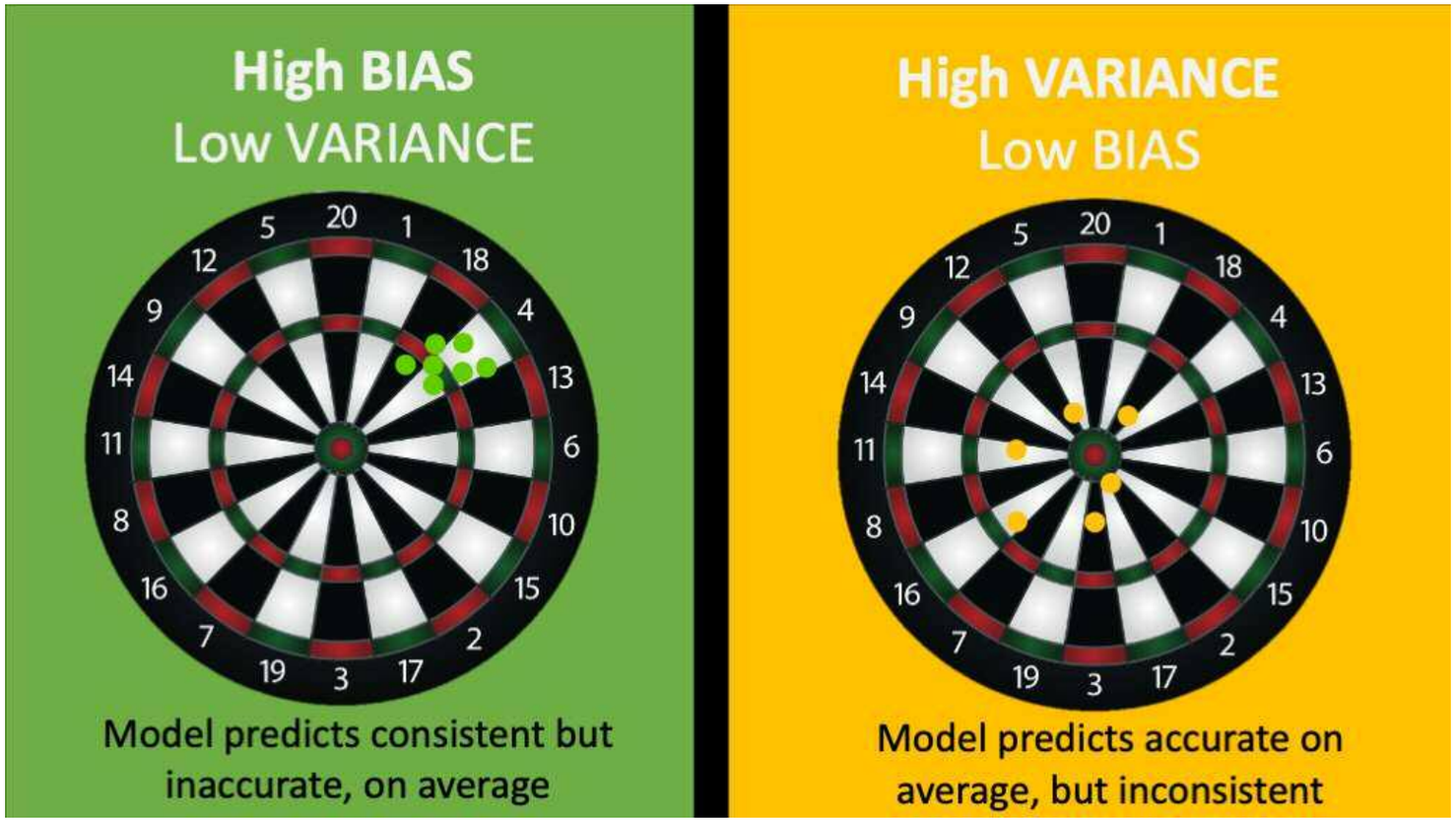
Splitting data into k groups or **folds** → using each time (each fold / each iteration) a different group as a validation set and the remaining one as training set:



$$\mathrm{CV}_{(k)} = \frac{1}{k}\sum_{i=1}^{k}\mathrm{MSE}_i.$$

(k-fold-cross-validation; Kohavi, 1995)

Resampling: Bias-Variance Tradeoff in Crossvalidation

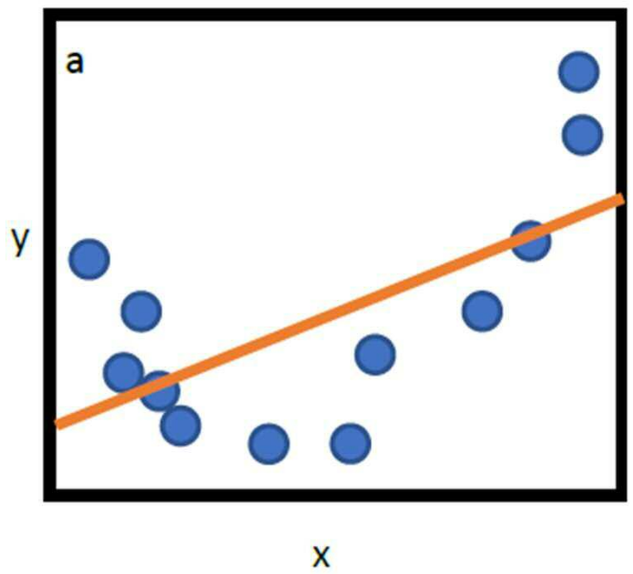


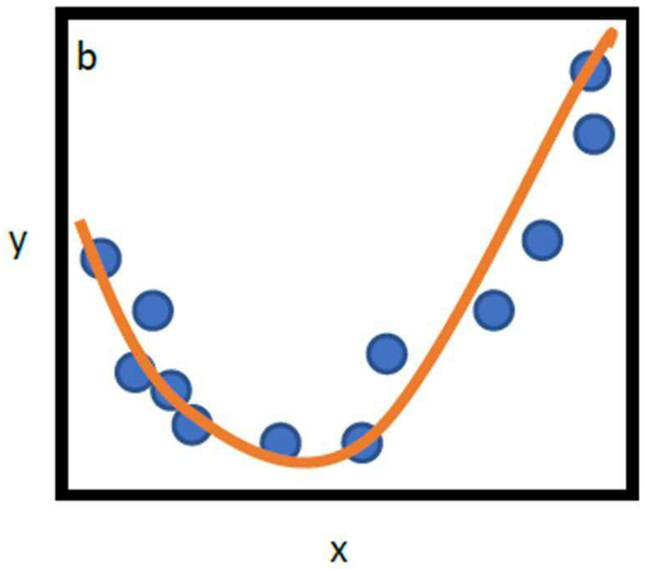
•**BIAS** can be observed when a algorithm/method lacks flexibility to learn the true signal or relations within the dataset which results in over- or underfitting (to the training data)

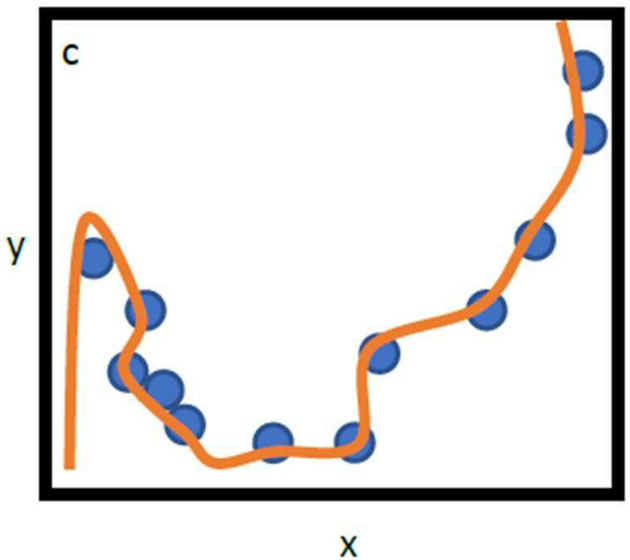
•**VARIANCE** refers to an algorithms sensitivity for specific constellations of a training set.

(James et al., 2022)

Overfitting / Underfitting







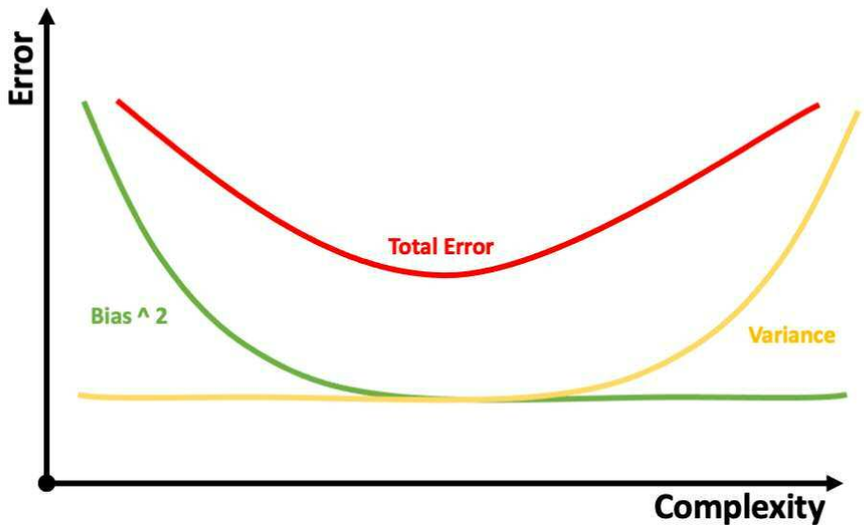
Adapted from Urban & Gates (2020)

Resampling: Bias-Variance Tradeoff in Crossvalidation

Low Bias

approaches tend to be **more complex** but are flexible

- Decisons Trees
- (K)NN
- Non-linear methods
- Non-parametric methods
- Neural Networks



Low Variance

algorithms are mostly **less complex** with a simple structure behind like

- (linear) Regression how we know it
- Naive Bayes
- Other linear methods
- Parametric methods

Error = $Bias^2$ + Variance + *Irreducible Error*

Note: Irreducible Error → noise that can only be reduced by data cleaning.

Within a algorithm/method family there is a TRADEOFF concerning complexity (James et al., 2022)

Accuracy – Interpretability Tradeoff

↑ Parameter size → ↑ complexity → ↑ accuracy → ↓ interpretability

- **High-Dimensional Models**: More parameters or features can fit the data better → increasing accuracy.
- **Interpretability**: More parameter → harder to understand → decreasing interpretability.
- **Highly Complex Models**: eg. deep neural networks → complex → high accuracy → low interpretability
- **Predictor Importance**: unclear effect of individual predictors due to transformations (feature engineering/feature learning)

Suggested solution: combine complex with simple models to maximize acc as well as interpretability (Orru et al., 2020)

→to understand the patterns in the data and predict dropout, while making these predictions interpretable and actionable.

(Goodfellow et al. 2017; Stachl et al., 2020)

Accuracy – Interpretability Tradeoff

↑ Parameter size → ↑ complexity → ↓ accuracy → ↑ interpretability

Not complex enough → underfitting → unable to learn the „real signal“ → does not generalize to new, unseen data.

Too complex → likely to overfit on noise (not signal) → does not generalize to unseen data.

Key is to BALANCE bias and variance in a way that minimizes Total Error

(Goodfellow et al. 2017; Stachl et al., 2020; James et al. 2022)

Metrics - General Guidlines

Metric	Poor	Below Average	Average	Good	Excellent	Note
F1 Score	0.0 - 0.4	0.4 - 0.6	0.6 - 0.7	0.7 - 0.9	0.9 - 1.0	Imbalanced data. Mean of precision and recall
AUC-ROC	0.5 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1.0	Ability of class separation
Accuracy	0.0 - 0.6	0.6 - 0.7	0.7 - 0.8	0.8 - 0.9	0.9 - 1.0	Ratio correct/total
Recall	0.0 - 0.4	0.4 - 0.6	0.6 - 0.7	0.7 - 0.9	0.9 - 1.0	Actual correct positives
Specificity	0.0 - 0.4	0.4 - 0.6	0.6 - 0.7	0.7 - 0.9	0.9 - 1.0	Actual correct negatives

- Leistungsfähige, quelloffene Plattform zur **Virtualisierung und Containerisierung**
- **Einfach skalierbar**
- Möglichkeit, auf einem physischen Server **mehrere virtuelle Maschinen (VMs)** oder **Container** parallel zu betreiben, zu verwalten und abzusichern – mittels WebGUI + Console
- eigene VM oder einen Container für jedes Belangen
- Benutzerverwaltung → So bleibt alles getrennt, sicher, stabil und flexibel
  - alles lokal betreibst → keine personenbezogenen Daten an Drittanbieter
  - volle Kontrolle über Datenspeicherung, Backups und Zugriff – DSGVO-konform.
- Keine Abhängigkeit von Cloud-Diensten DSGVO-konform Geringere Latenz, Kontrolle über Serverlast

Ollama

Open-Source-Plattform zum **lokalen Ausführen von LLMs** – wie LLaMA, Mistral, DeepSeek,Gemma oder Phi
→ direkt auf deinem eigenen Rechner oder Server.

Vorteil	Nutzen
Einfacher Start	Installation mit einem Befehl, intuitive CLI
Läuft offline	Ideal für sensible Forschungsdaten
Vordefinierte Modelle	Modelle wie llama3, mistral, phi, gemma direkt ladbar
Schnelle Testläufe	Ideal für Entwicklung neuer Aufgaben, Skalen/SJTs oder automatische AUT response ratings
Kontrolle & Transparenz	Vollständiger Einblick in Modell, Daten, Prompts, Fine-tune
Schnittstellenfähig	Kombinierbar mit R-Servern, Datenbanken, jsPsych-Tests, just name it

Personality SJT Generation

Agreeableness (Compassion)

An elderly neighbor spontaneously asks you for help because she urgently needs to go to the doctor and has no one to drive her.

(a) You politely explain that you already have other commitments and offer to call her a cab. [1]

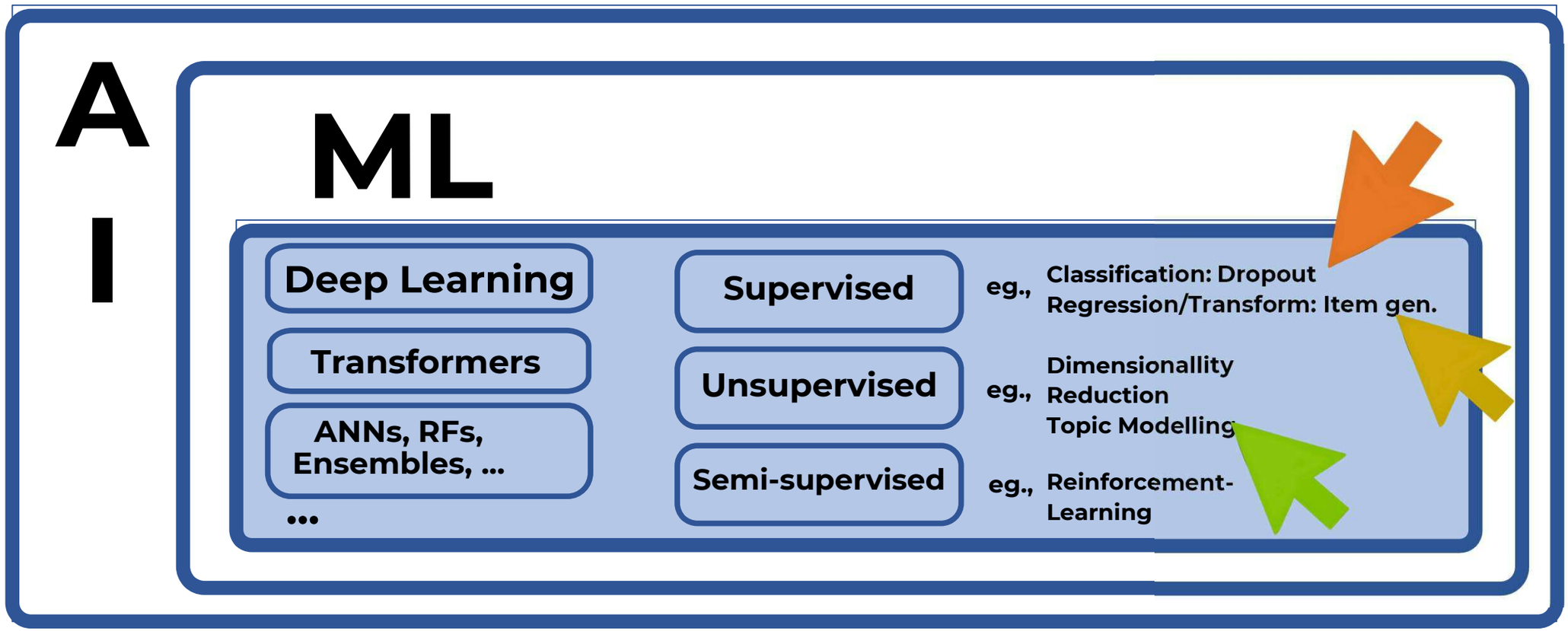
(b) You agree and drive her to the doctor, even if it means postponing your own plans. [3]

(c) You agree immediately, drive her to the doctor and offer to wait for her even though you have other plans. [4]

(d) You explain that you don't really have time, but could step in if there is no other option. [2]

Introduction

Applications in the field of Psychology



Ai-Driven SJT Refinement

Results for Conscientiousness

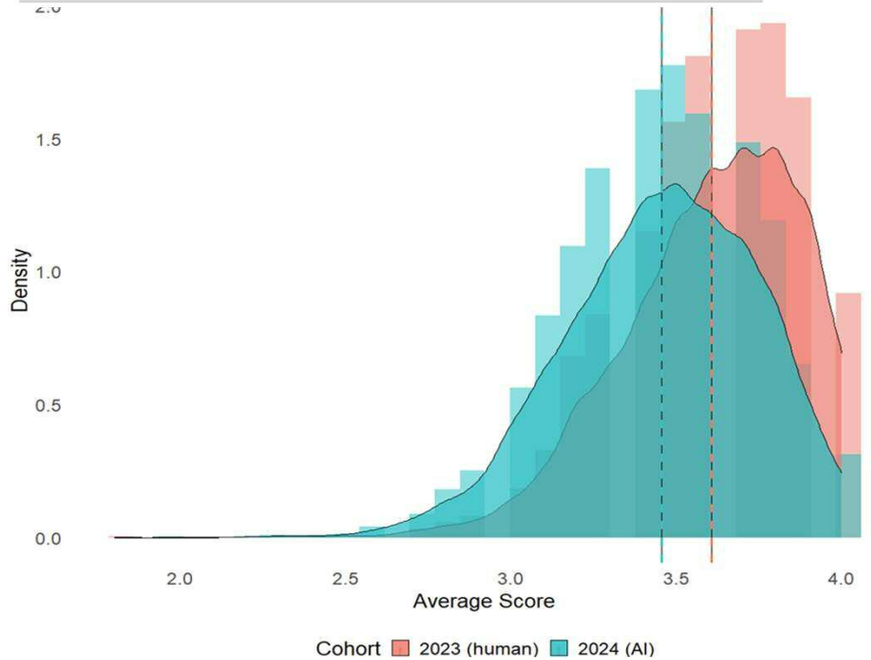
- Significant lower mean-scores
- good model fit
- acceptable reliability

Comparison of Model Fit Indices for Conscientiousness (human vs. AI)

Fit Index	Human crafted	AI crafted
Chi-Square (Chisq)	115.77	173.17
Degrees of Freedom (df)	35.00	35.00
Chi-Square/df (Chisq/df)	3.31	4.95
p-value	0.00	0.00
CFI	0.98	0.96
RMSEA	0.02	0.03
SRMR	0.04	0.05

Conscientiousness (human vs AI)

Cohort	Mean	SD	Omega
human	3.61	0.27	0.61
AI	3.45	0.29	0.53



Note. N = 5555 (2829 AI based) $t_{(5553)} = 20.05$, $p = < .001$, $d = 0.54$

Ai-Driven SJT Refinement

Results for Extraversion

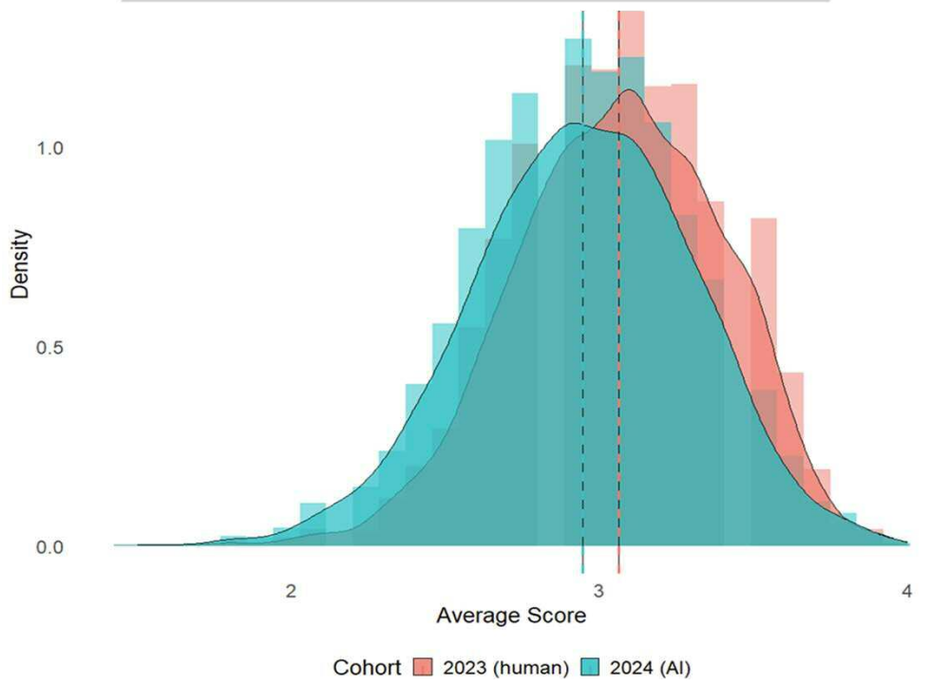
- Significant lower mean-scores
- good model fit
- similar reliability

Comparison of Model Fit Indices for Conscientiousness (human vs. AI)

Fit Index	Human crafted	AI crafted
Chi-Square (Chisq)	115.77	173.17
Degrees of Freedom (df)	35.00	35.00
Chi-Square/df (Chisq/df)	3.31	4.95
p-value	0.00	0.00
CFI	0.98	0.96
RMSEA	0.02	0.03
SRMR	0.04	0.05

Extraversion (human vs AI)

Cohort	Mean	SD	Omega
human	3.06	0.34	0.68
AI	2.95	0.36	0.65



Note. N = 5555 (2829 AI based) $t_{(5553)} = 12.39$, $p = < .001$, $d = 0.33$

Ai-Driven SJT Refinement

Results for Emotional Stability

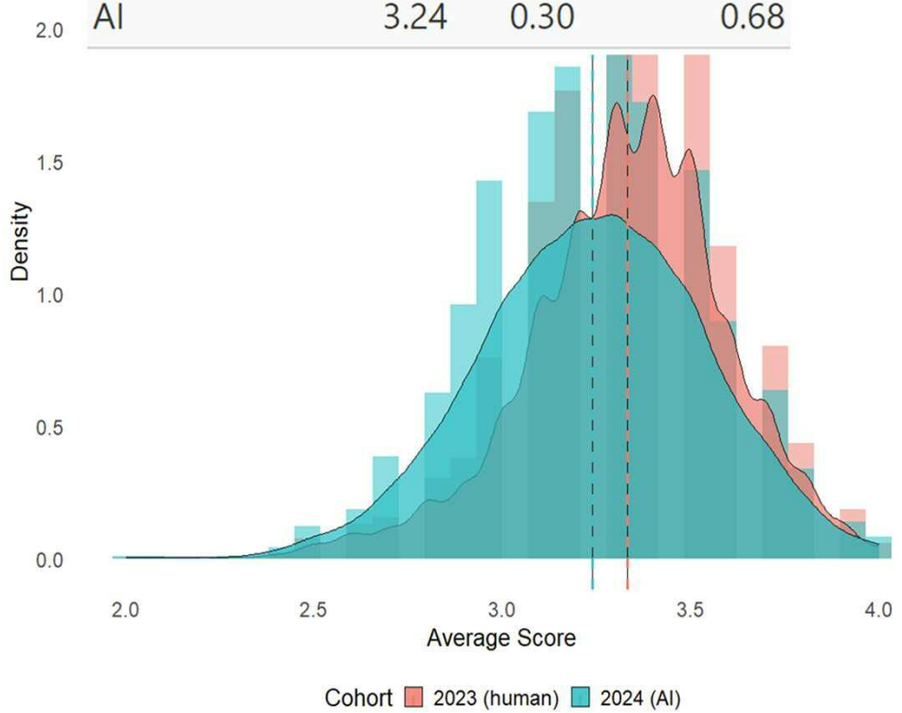
- Significant lower mean-scores
- good model fit
- Slightly better reliability

Comparison of Model Fit Indices for Emotional Stability (human vs. AI)

Fit Index	Human crafted	AI crafted
Chi-Square	120.93	107.75
df	35.00	35.00
Chisq/df	3.46	3.08
p-value	0.00	0.00
CFI	0.99	0.99
RMSEA	0.02	0.02
SRMR	0.03	0.03

Emotional Stability (human vs AI)

Cohort	Mean	SD	Omega
human	3.33	0.26	0.65
AI	3.24	0.30	0.68



Note. N = 5555 (2829 AI based) $t_{(5553)} = 12.15$, $p = < .001$, $d = .33$

Possible Scale evaluation methods

- Blind expert review of AIG / traditional (eg. Hernandez & Nie, 2021)
- Mix AIG / traditional items for expert review (e.g., Gierl and Lai, 2018; Pugh et al., 2020)
- Factor analysis to examine the internal structure (von Davier, 2018)
- Comparison of psychometric properties of AIG items with published items (e.g. Attali, 2018)
- Examination of the similarity of generated/traditional items (eg. cosine similarity; Gierl and Lai, 2013; Hernandez & Nie, 2021)
- Properties of the items using classical test theory, item response theory measures (e.g. Attali et al., 2023)

Item examples: Wanderlust (Götz et al., 2023)

Model: PIG: Big Five fine-tuned, but non-trained construct predited

„I want to take pictures of my surroundings, stories and places.“

„I want to explore, both in my own country and abroad.“

“I like to see what cultures and languages they speak.“

Excluded items:
“I value my independence”, “I am outgoing and excited to meet new people”, “I’m not afraid to go through obstacles”

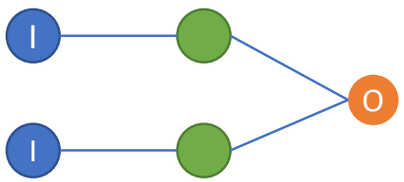
Item examples

Trait	Machine-authored item using GPT-3: 25 items	Hommel et al. (2022) using GPT-2: 25 items	von Davier (2018) using LSTM: 24 items
Openness	1. I have an active imagination	1. I can enjoy a wide variety of musical styles	1. I worry so little
	2.I get deeply involved in new things	2. I like to be surprised	2. I am seldom bothered by problems
	3. I like to explore ideas	3. I love to contemplate the universe and its beauty	3. I can make any setting my situation
Conscientiousness	6. I am thorough in completing tasks	6. I am not always on time for work	4. I am willing to take responsibility for my work
	7. I plan ahead	7. I know that I make many mistakes	5. I am aware of how I am doing
	8. I put off working on assignments until the last minute	8. I work too hard	6. I am about my motives
Extraversion	11. I rarely engage in conversations	11. I am able to speak confidently	8. I prefer to be the perfect leader
	12. I find it difficult to socialize	12. I avoid public places	9. I think I would not be a success in movies
	13. I am shy around strangers	13. I am able to handle myself in a crowd	
Agreeableness	16. I am considerate and kind to almost everyone	16. I care a lot about others	10. I feel that I am too blunt
	17. I do things for the good of all	17. I am easily angered	
	18. I am always willing to lend a helping hand	18. I don't like to argue	
Neuroticism	21. I am often feeling low in spirits	21. I am generally happy and content	11. I often do not know what happens to me
	22. I am usually very happy	22. I am often upset by minor things	12. I rarely consider my actions
	23. I am often feeling glum	23. I am a person who is easily moved by the good moods and bad moods of others	13. I worry so much about myself

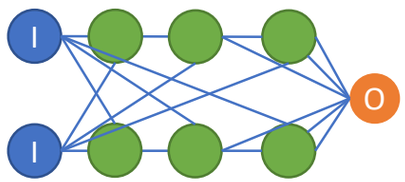
Lee et al. (2022)

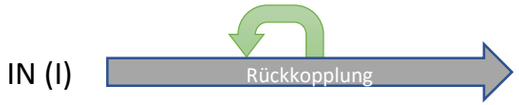
Artificial Neural Networks

Single Layer (feed foward) Netz

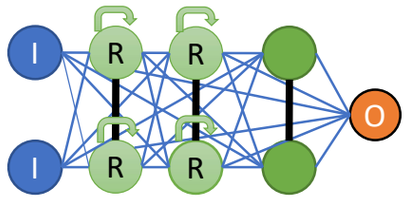


Deep Neural Network





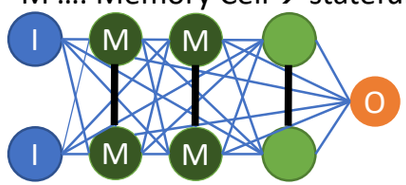
Recurrent Network



R …. Recurrent Cell → Rückkopplung

Long/Short-Term Memory Network

M …. Memory Cell→ stateful = Zustandsspeicher



DI:(PS Differential Psychology Graz
Institute of Psychology, University of Graz, Austria

UNI GRAZ

Machine Learning & Psychometrics

Applications and benefits:

- Reproducibility:

  Effect sizes vary over replications of statistical outcomes →

Crossvalidation:

Translation of ML metrics (eg. accuracy) in psychological metrics, such as effect size, possible (Salgado, 2018)

→ eg. Prediction of dichotomous outcome

Accuracy = .71 ~ d_{Cohen} = 1.19

vs.

k-Crossfold Crossvalidation:

Accuracy .65 ~ d_{Cohen} = .56

Effect sizes according to Cohen (1977)

Orru et al. (2020)

Metrics (most popular)

Accuracy

$$\frac{korrekte\ Voraussagen}{m\ddot{o}gliche\ Voraussagen}$$

Recall

$$\frac{True\ Positive}{True\ Positive + False\ Negative} = \frac{wirklich\ kranke}{kranke + f\ddot{a}lschlich\ nicht\ Angeklagte} = \text{Sensitivität}$$

Precision

$$\frac{True\ Positive}{True\ Positive\ +\ False\ Positive} = \frac{kranke}{\text{kranke} + f\ddot{a}lschlich\ erkrankte}$$

DI:(PS Differential Psychology Graz
Institute of Psychology, University of Graz, Austria

UNI GRAZ

Possible Scale evaluation methods

- Blind expert review of AIG / traditional (eg. Hernandez & Nie, 2021)
- Mix AIG / traditional items for expert review (e.g., Gierl and Lai, 2018; Pugh et al., 2020)
- Factor analysis to examine the internal structure (von Davier, 2018)
- Comparison of psychometric properties of AIG items with published items (e.g. Attali, 2018)
- Examination of the similarity of generated/traditional items (eg. cosine similarity; Gierl and Lai, 2013; Hernandez & Nie, 2021)
- Properties of the items using classical test theory, item response theory measures (e.g. Attali et al., 2023)

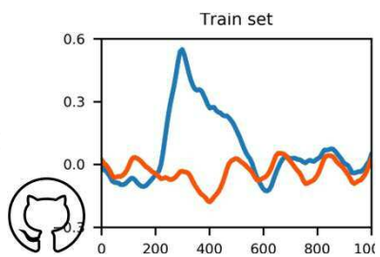
Supervised Learning

Regression:

- Numerische, kontinuierliche AVen
- Mögliche Modelle u.a:
  - Rekurrente Neuronale Netzwerke, Random Forests uvm.
- AVen:
  - Reaktionszeiten,
  - Testscores,
  - Jobperformance,
  - (f)MRI Voxel,
  - Originalitätsratings,
  - Sequenzen (z.B. Blickbewegungen), …
- Eigene Anwendungsbeispiele:
  - LSTMusix
    &

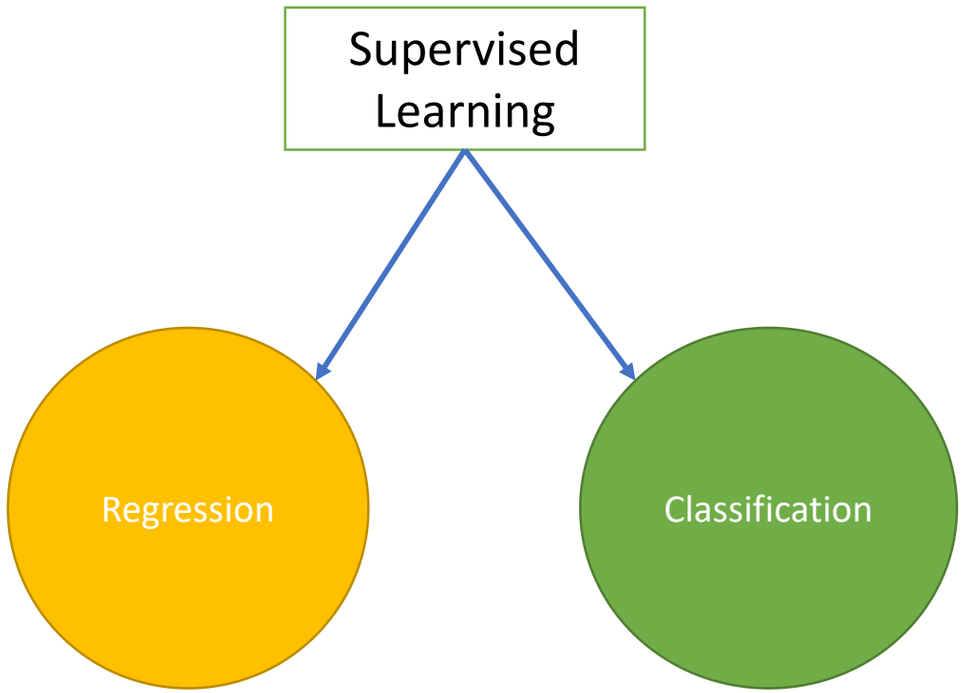
Klassifikation:

- Kategoriale Aven
  - z.B. Deep Neural Networks (> 1 hidden layer)
  - Random Forrests
  - Support Vector Machines uvm.
  - Faking vs. Honest,
  - BIG5 Profile, Subfacetten der Pers.
  - Ereigniskorrelierte EEG Potentiale,
  - Psy. Störungsbilder
  - …
  - P300Net



Continous DVs

- Reationtime,
- Testscores,
- Jobperformanc e,
- (f)MRI Voxel,
- Originality Rating,
- Sequences (eg. gaze)



(multi)nominal DVs

- Faking,
- BIG5 Facets,
- ERPs
- Dosorders
- …

Supervised Learning Pros

- Generalized models (e.g. faking) and individual models (e.g. EKPs) möglich
- Complex relations
- No data preassumptions
- Number of DV can be higher than IVs (features)
- …

JM(g1

Supervised Learning Cons

- Blackbox: Prediction over explanation
  - Underlying processes often unclear
  - higher complexity → lower interpretatibility
- Accuracy highly dependent on underlying data
- Overfitting

JM(g1 Ähnliche Predictions von Algorithmen, denen ähnliche Annahmen zugrunde liegen. e.g., Support Vector Machine, Naive Bayes, Knn, Random Forest Wege zur Verbesserung von Classifiern: feature selection and feature engineering and parameter tuning

Jud, Marcel (marcel.jud@uni-graz.at); 15.01.2021

Prompting LLMs

Prompt Design

Instructions and context as input to achieve a desired task

Prompt Engineering

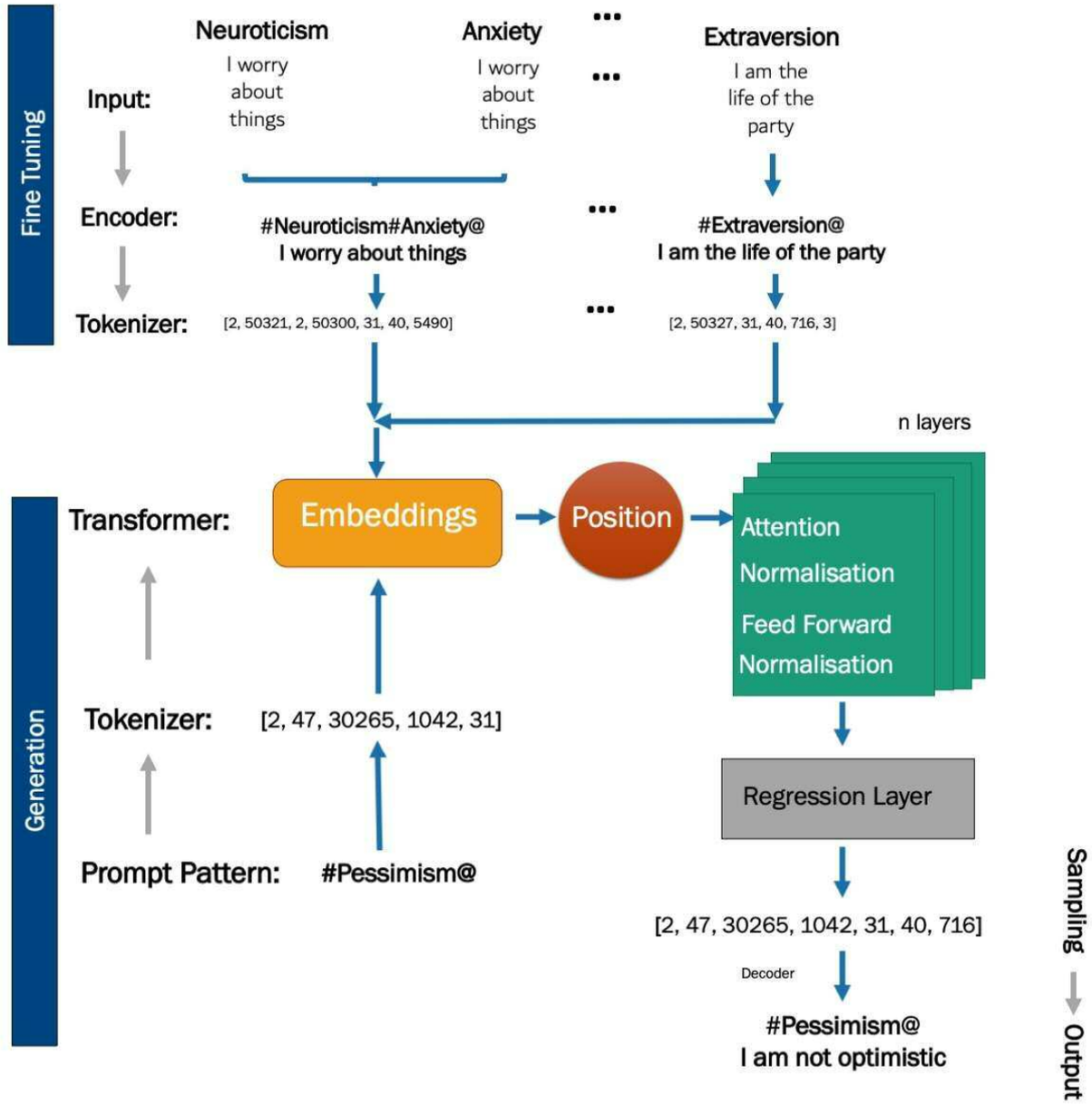
Development and optimization of prompts to efficienty use models for variety of applications

e.g. Context rich prompts
Step-by-step (chain-of-thought)

Transformer Networks

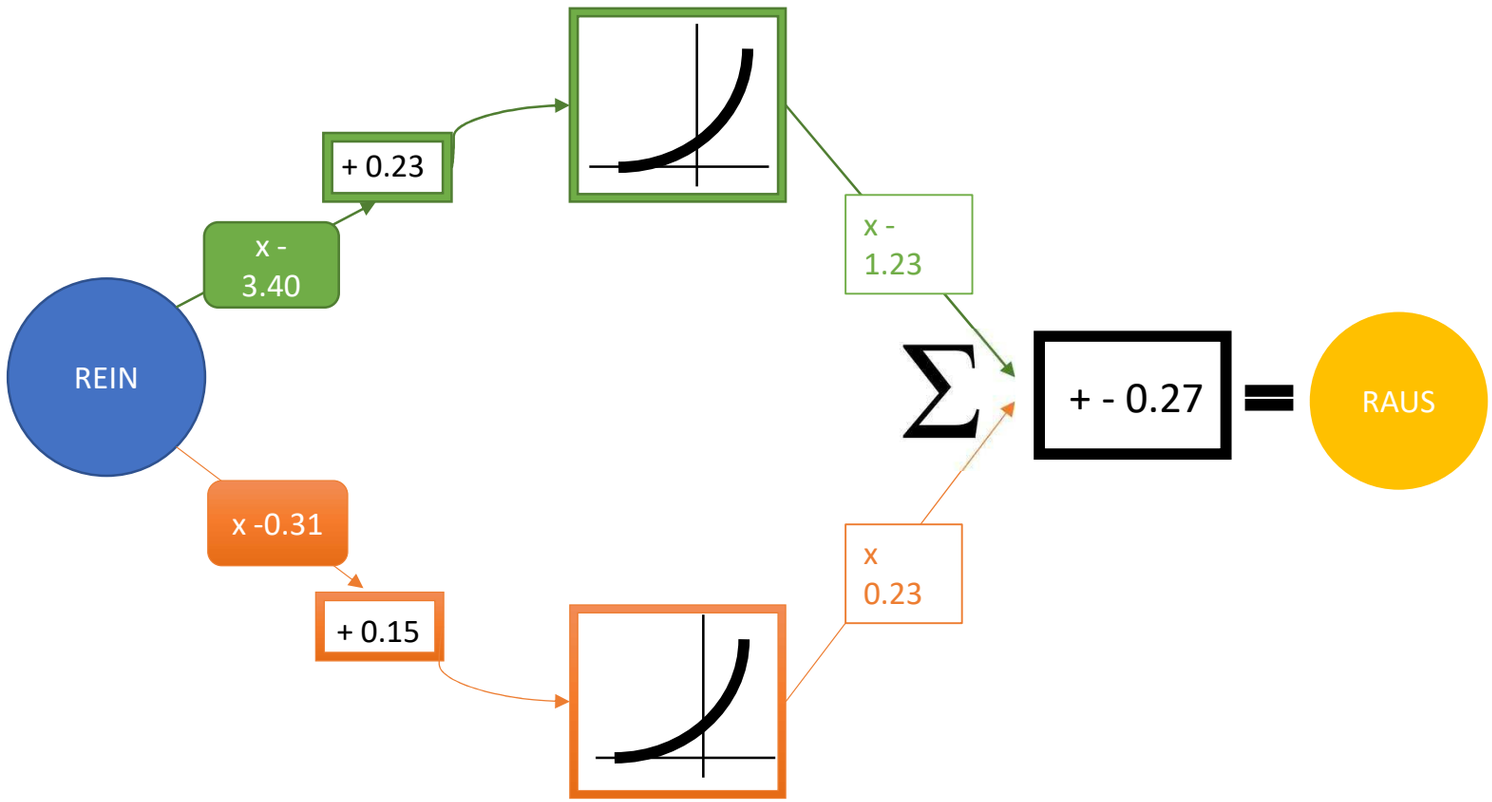
Step 1:
Construct specific finetuning

Step 2:
Sequence-to-sequence-encoder-decoder-modelling



adapted from Radford et al. (2018)

Neural Networks



Activation functions:

Transformer derived scales in the wild

Trained constructs

- AI-IP (Hernandez & Nie, 2021)

  → BERT; IPIP fine-tuning (Goldberg et al., 2006)

- Neural-BFI20 (Götz et al., 2023)

  → GPT-2; BFI-2 fine-tuning

- Human vs. machine made Big 5 inventory (Hommel et al., 2023)

  → GPT-2; IPIP fine-tuning (Goldberg et al., 2006)

- Machine made Big-5 (Lee et al., 2022)

  → GPT-3; IPIP fine-tuning (Goldberg et al., 2006)

Untrained (new) constructs

- Wanderlust Scale (Götz et al., 2023)

  → GPT-2; BFI-2 based (Danner et al., 2016)

- Benevolence, Egalitarianism, Egoism, Pessimism (Hommel et al., 2023)

  → GPT-2; IPIP fine-tuning (Goldberg et al., 2006)

Transformer derived Scales: Properties

Trained constructs

- Neural-BFI20 (Götz et al., 2023)

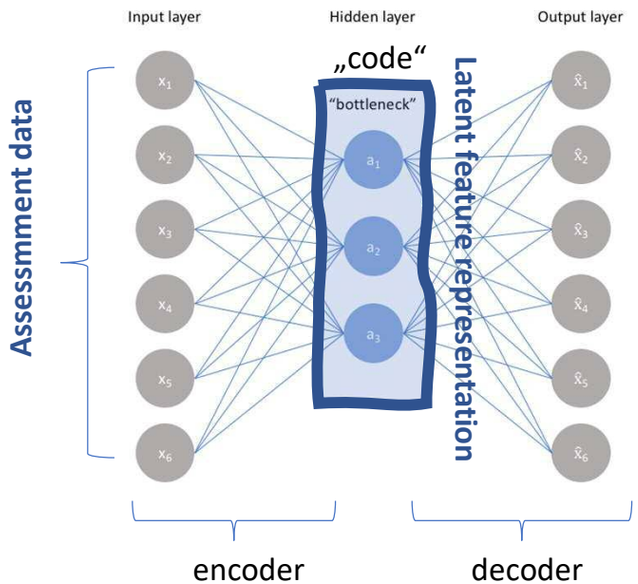
Correlations Between the N-BFI-20 and the BFI-2

			Pearson's *r*		*p*	95% CI	
						LL	*UL*
N-BFI-20 O	**-**	**BFI-2 O**	**0.731**	***	**< .001**	**0.696**	**0.762**
N-BFI-20 C	**-**	**BFI-2 C**	**0.821**	***	**< .001**	**0.796**	**0.843**
N-BFI-20 E	**-**	**BFI-2 E**	**0.795**	***	**< .001**	**0.768**	**0.820**
N-BFI-20 A	**-**	**BFI-2 A**	**0.767**	***	**< .001**	**0.736**	**0.794**
N-BFI-20 N	**-**	**BFI-2 N**	**0.786**	***	**< .001**	**0.758**	**0.812**

Note. Bold cells represent a significant correlation above $r = 0.700$. $N = 773$. O = Openness. C = Conscientiousness. E = Extraversion. A = Agreeableness. N = Neuroticism. CI = confidence interval. *LL* = lower limit. *UL* = upper limit. *** $p < .001$, ** $p < .01$, * $p < .05$

Training – autoencoder

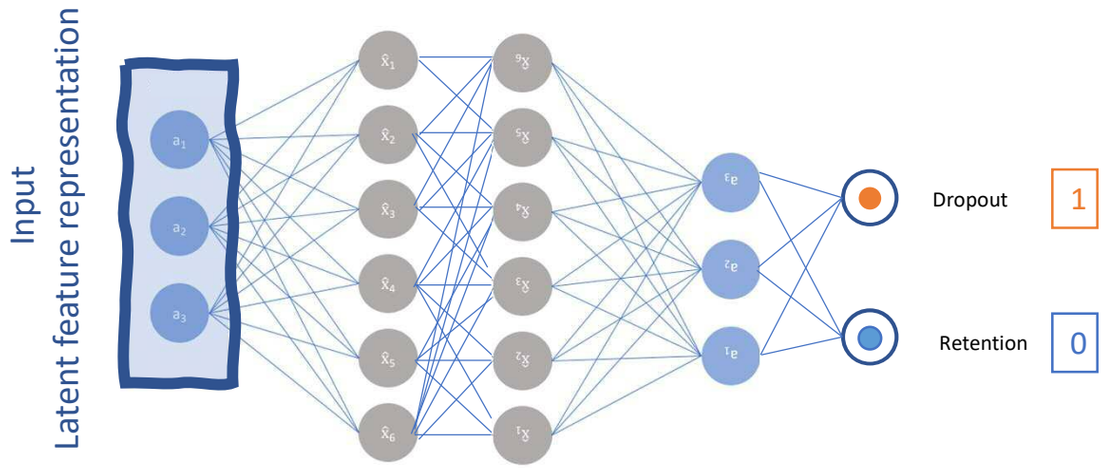
Autoencoder based **feature learning** instead of manual feature engineering → **dimensional reduction** leads to abstract but informative features



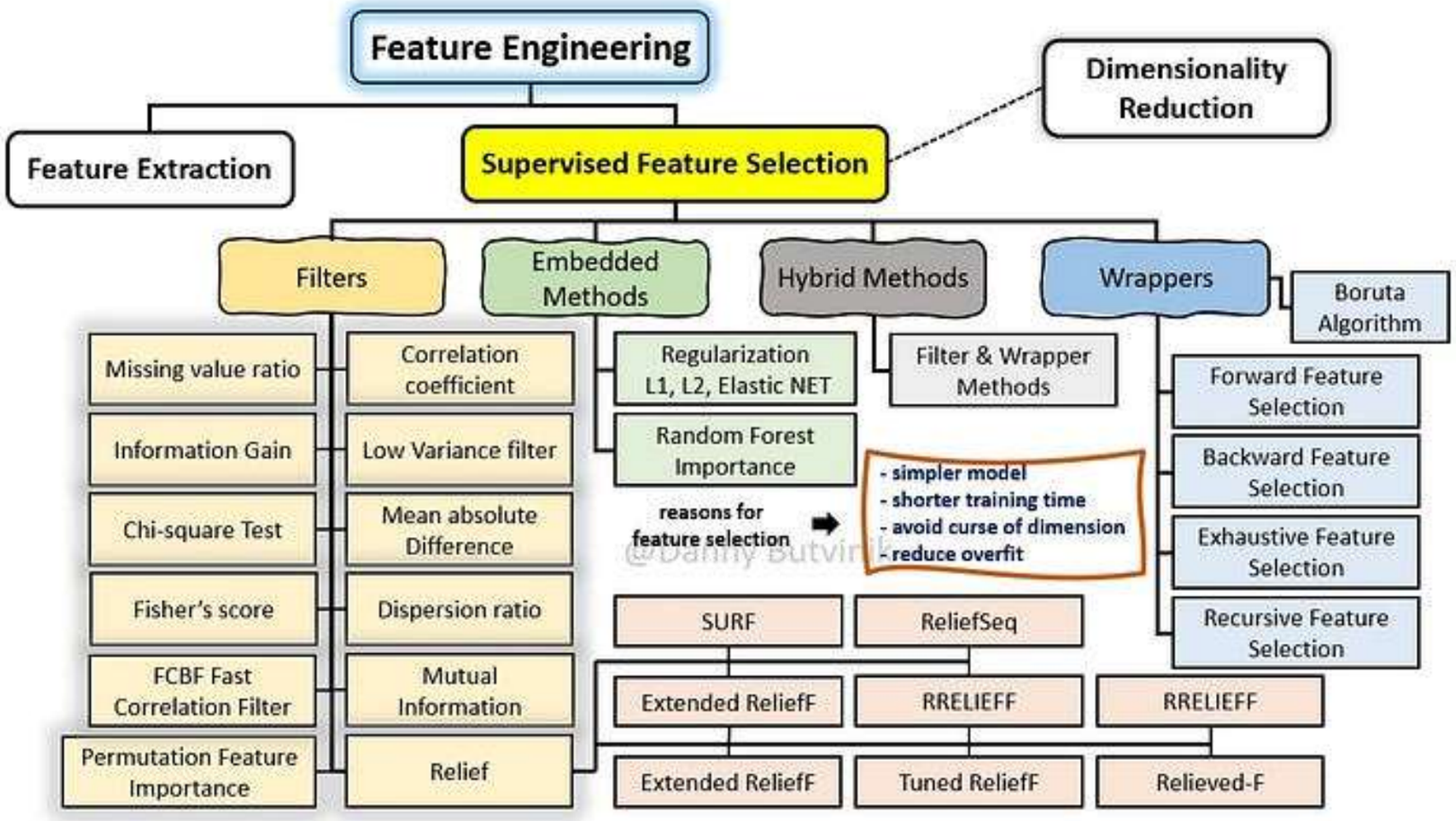
Latent feature representations used as predictors for the actual classification model

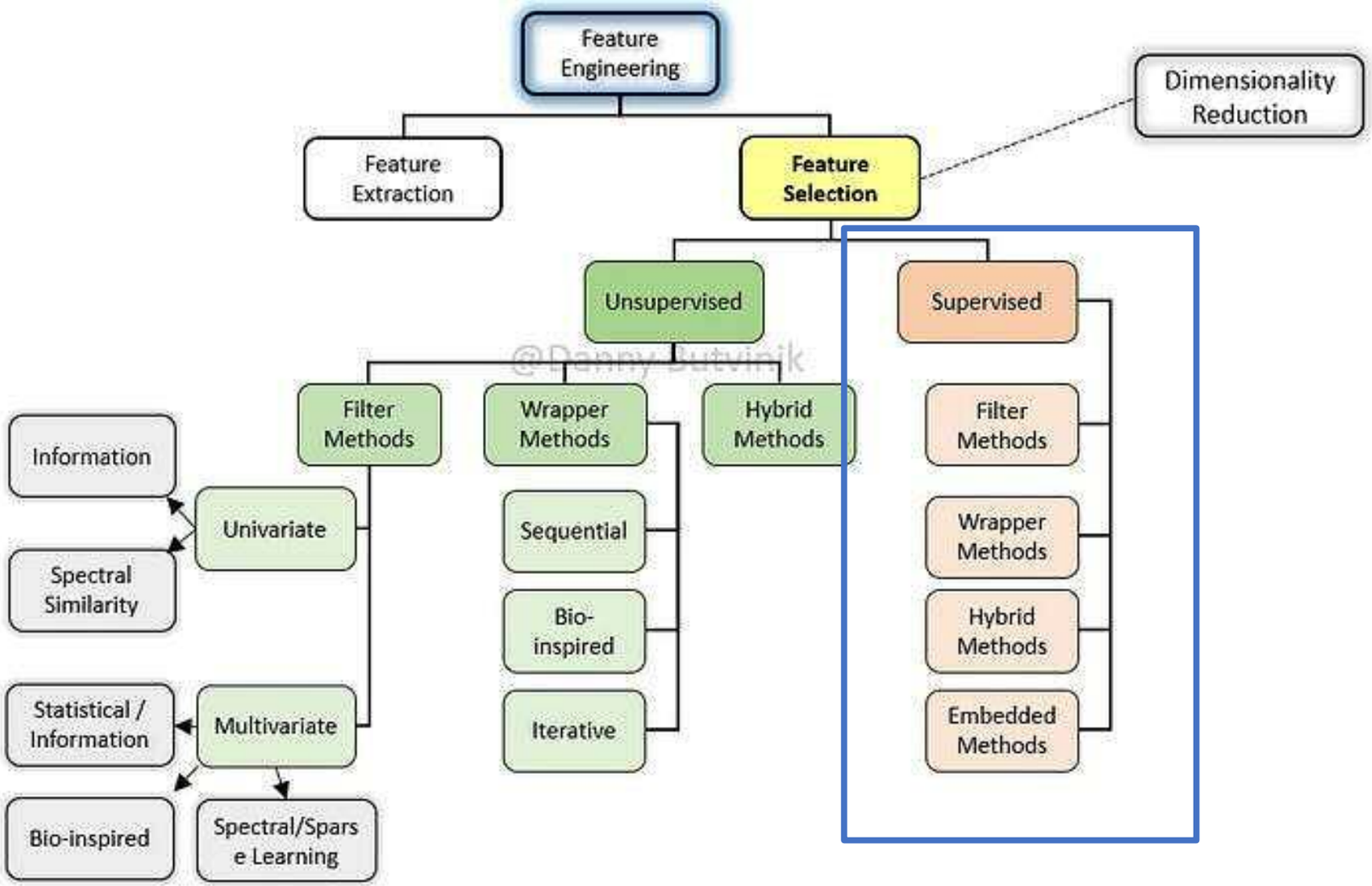
Training – Complex Model

Latent feature representation then used as predictors for the actual classification model



Feature Engineering




Feature Engineering
Feature Extraction
Feature Selection
Dimensionality Reduction
Unsupervised
Supervised
@Danny Butvinik
Filter Methods
Wrapper Methods
Hybrid Methods
Information
Univariate
Spectral Similarity
Statistical / Information
Multivariate
Bio-inspired
Spectral/Spars e Learning
Sequential
Bio-inspired
Iterative
Filter Methods
Wrapper Methods
Hybrid Methods
Embedded Methods