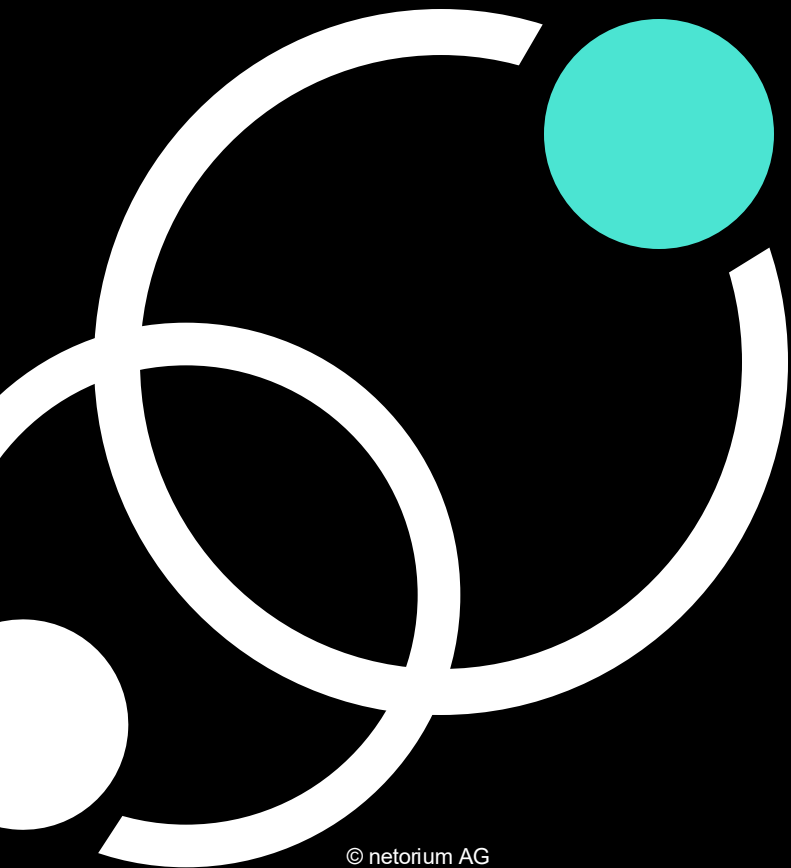




Wenn alles wie ein Nagel aussieht

*Warum nicht jedes Problem KI
braucht*



“ it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail.

Abraham Maslow (1908-1970)





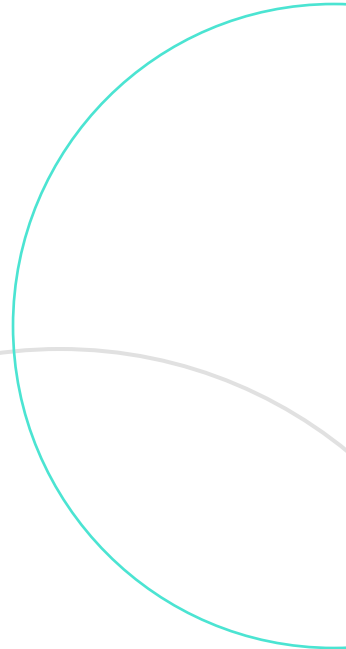
“wer nur einen Hammer besitzt, ist versucht, in allem einen Nagel zu sehen.



Abraham Maslow (1908-1970)



Nicht jedes Problem braucht Intelligenz



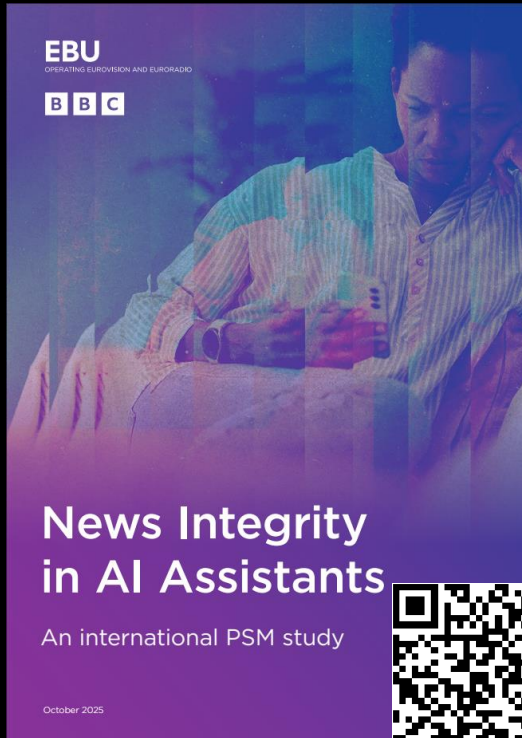
AGENDA

- Hype
- was KI kann
- wo ein Algorithmus reicht
- Entscheidungskriterium

DER HYPE IN ZAHLEN



DER HYPE IN ZAHLEN



GPT-4 or Reciting? Exploring the Capabilities and Limitations of Language Models Through Counterfactual Tasks

Wu¹, Linlu Qiu², Alexis Ross², Ekin Akyürek², Boyuan Chen²,
Jilin Wang², Najoung Kim², Jacob Andreas², Yoon Kim²
¹MIT ²Boston University
zfw@csail.mit.edu

arXiv:2307.02477v3 [cs.CL] 28 Mar 2024



Figure 1: GPT-4’s performance on the default version of various tasks (blue) and counterfactual counterparts (orange). The shown results use 0-shot chain-of-thought prompting [34; Kojima et al., 2023]. GPT-4 consistently and substantially underperforms on counterfactual variants compared to default task instantiations.

Abstract

The impressive performance of recent language models across a wide range of tasks suggests that they possess a degree of abstract reasoning skills. Are these skills general and transferable, or specialized to specific tasks seen during pretraining? To disentangle these effects, we propose an evaluation framework based on “counterfactual” task variants that deviate from the default assumptions underlying standard tasks. Across

a suite of 11 tasks, we observe nontrivial performance on the counterfactual variants, but nevertheless find that performance substantially and consistently degrades compared to the default conditions. This suggests that while current LMs may possess abstract task-solving skills to an extent, they often also rely on narrow, non-transferable procedures for task-solving. These results motivate a more careful interpretation of language model performance that teases apart these aspects of behavior.

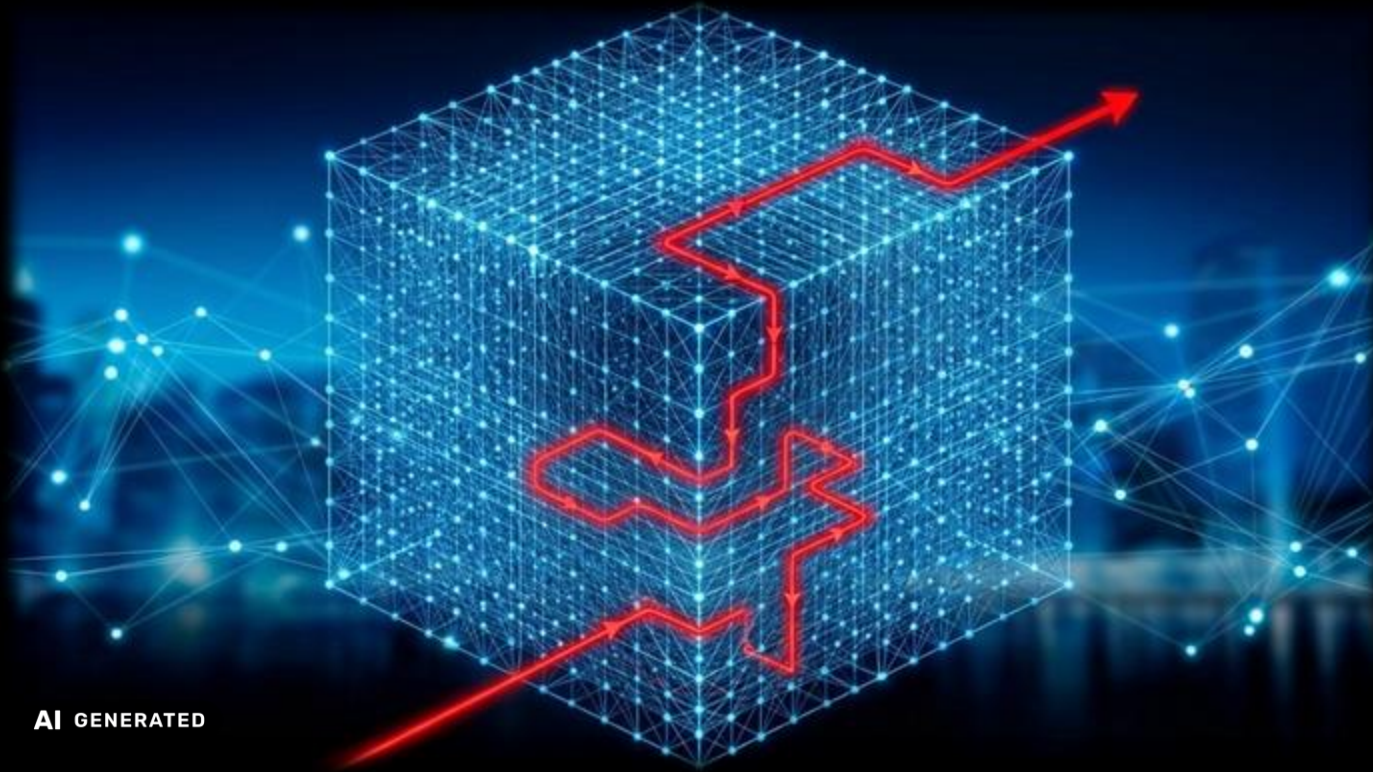
DER HYPE IN ZAHLEN



DER HYPE IN ZAHLEN

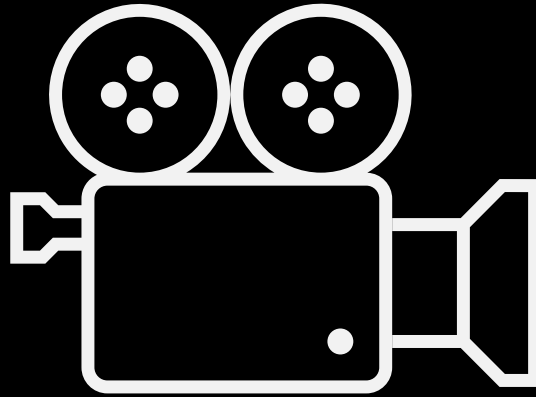


WAS KI WIRKLICH GUT KANN

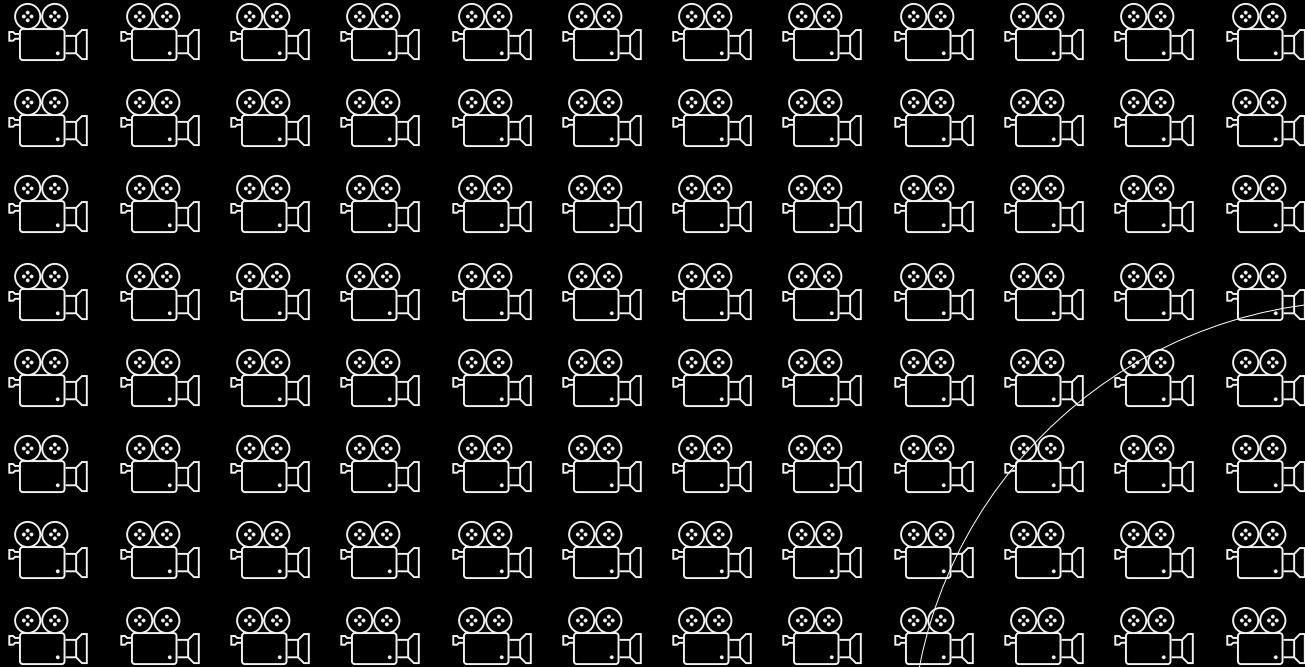


AI GENERATED

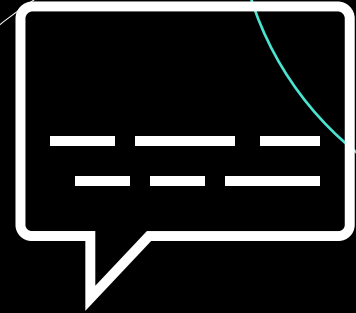
WAS KI WIRKLICH GUT KANN



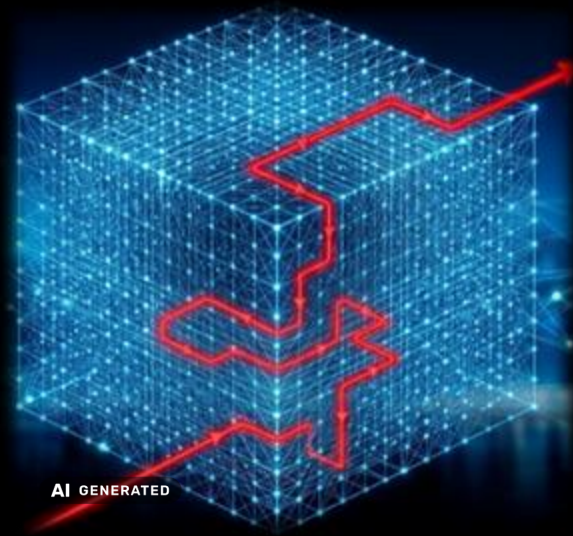
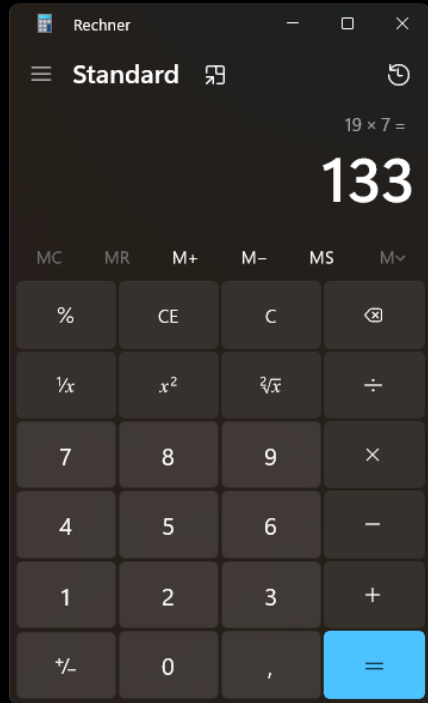
Was KI WIRKLICH GUT KANN



WAS KI WIRKLICH GUT KANN



ALGORITHMUS VS. KI



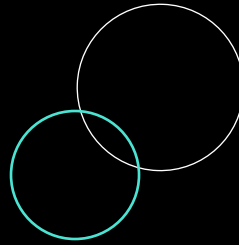
ALGORITHMUS VS. KI

- EBU R128 / ITU-R BS.1770

$$\Gamma_r = -0.691 + 10 \log_{10} \sum_i G_i \cdot \left(\frac{1}{|J_g|} \cdot \sum_{J_g} z_{ij} \right) - 10 \text{ LKFS}$$

- MD5/SHA

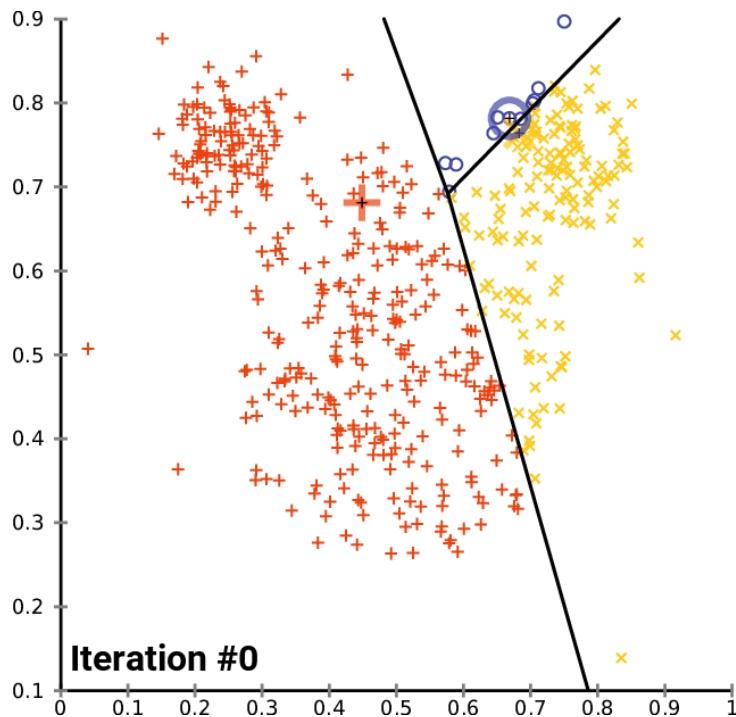
$$\begin{aligned} F(X, Y, Z) &= (X \wedge Y) \vee (\neg X \wedge Z) \\ G(X, Y, Z) &= (X \wedge Z) \vee (Y \wedge \neg Z) \\ H(X, Y, Z) &= X \oplus Y \oplus Z \\ I(X, Y, Z) &= Y \oplus (X \vee \neg Z) \end{aligned}$$



FAIR IS FOUL, AND FOUL IS FAIR



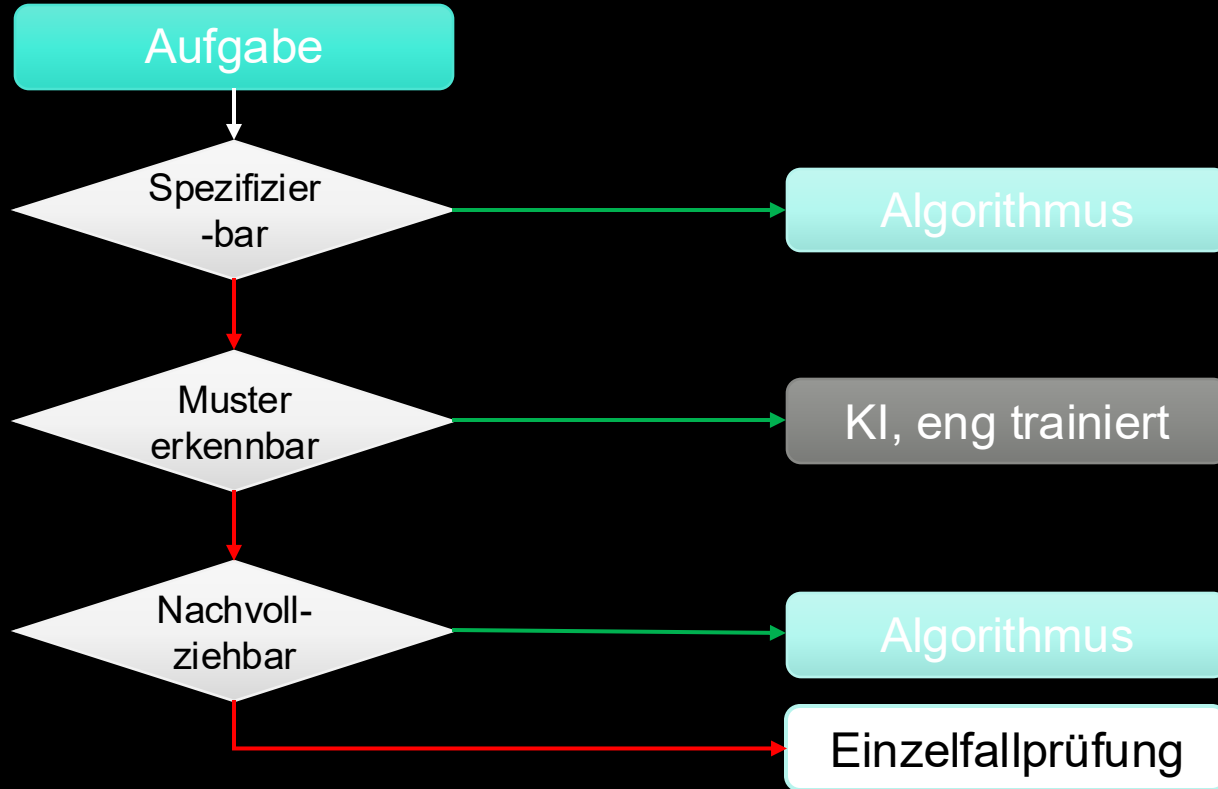
K-MEANS



REALITÄT

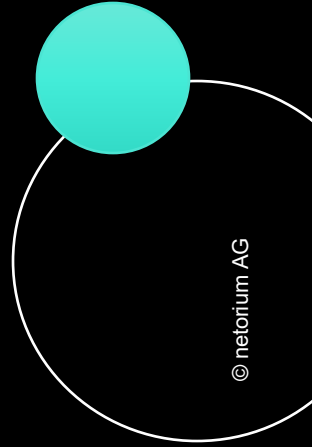
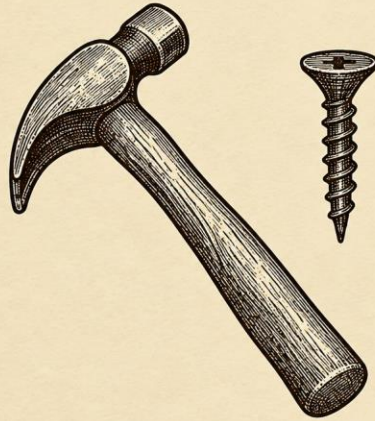


ENTSCHEIDUNGSBAUM



MANCHMAL REICHT LOGIK

The Law of the Instrument



© netorium AG



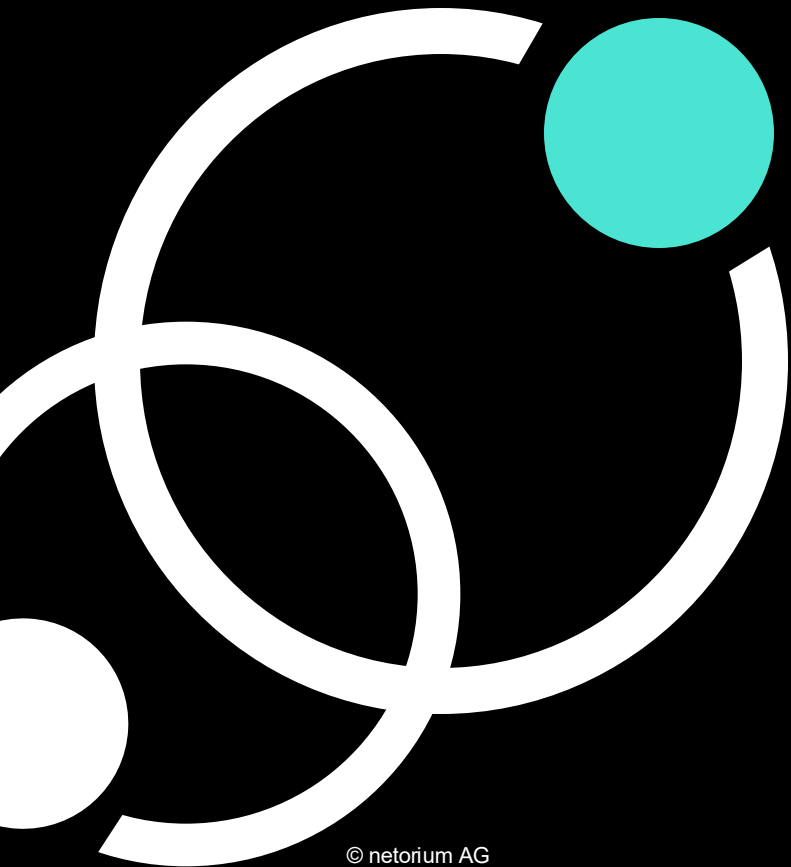
Computer Simulation of Personality: Frontier of Psychological Theory (1963)

“ the tendency of jobs to be adapted to tools,
rather than adapting tools to jobs

Silvan Tomkins

“ The computer program may be our current
hammer, but it must be tried. One cannot
decide from purely armchair considerations
whether or not it will be of any value.

Kenneth Mark Colby



Matthias Letsch

Technischer Leiter

matthias.letsch@netorium.de

+49 (6122) 170 96 45