

**Thomas Jäger, Wien: Faking the Middle Ages – Reflexionen über den Einsatz
synthetischer Daten in der Mediävistik**

Synthetische Daten sind keine bloßen Hilfsmittel, sondern kontrollierte Formen der Fälschung. In Informatik und KI-Engineering werden sie eingesetzt, um fehlende Daten zu simulieren, Verzerrungen auszugleichen und Modelle unter kontrollierten Bedingungen zu entwickeln. Am Beispiel mittelhochdeutscher Reimpaarverse zeigt sich, dass aktuelle Sprachmodelle zwar mittelhochdeutsche Texte verstehen, jedoch keine tragfähigen mittelhochdeutschen Texte erzeugen: Ihre Ergebnisse wirken nur oberflächlich plausibel und scheitern an den Anforderungen historischer Sprach- und Verssysteme. Dieses Scheitern verweist auf ein zentrales Problem mediävistischer Arbeit: die Notwendigkeit, lückenhafte und verzerrte Überlieferung modellbasiert zu ergänzen. In Editionsphilologie und Geschichtswissenschaft geschieht dies in Form von Konjekturen, Rekonstruktionen und Hypothesen. Der Vortrag argumentiert, dass synthetische Daten als explizite Fortsetzung dieser Praxis verstanden werden können: Sie machen implizite Modellierungen sichtbar und erlauben ihre systematische Variation und Überprüfung.

Anhand von Experimenten zur Erzeugung mittelhochdeutscher Reimpaarverse sowie von Normalisierungs- und Variantenpaaren wird gezeigt, dass synthetische Daten nicht primär der Materialvermehrung dienen, sondern der Modellierung von Überlieferungsbedingungen. Gerade das Scheitern aktueller KI-Modelle macht sichtbar, welches Strukturwissen zu Metrik, Syntax und Überlieferungsvarianz explizit gemacht werden muss. Erste Ergebnisse zeigen, dass sich einzelne Teilaspekte bereits modellieren lassen.

Synthetische Daten schaffen Testumgebungen für Verfahren der Normalisierung, Variantenanalyse und textnahen Recherche. Der Beitrag verortet diese Ansätze im Kontext mediävistischer Auseinandersetzungen mit Fälschung, Rekonstruktion und Normierung und schlägt vor, die gezielte Erzeugung von „Falschheit“ als methodisches Instrument zu begreifen: nicht als Ersatz der Überlieferung, sondern als Mittel, ihre Bedingungen explizit zu modellieren und kritisch zu prüfen.