# Hurra, hurra: Die Betaversion ist da

Andreas Gross am 13. August 2024 über das Projekt QualsKAT

### Was ist QualsKAT

QualsKAT ist die Abkürzung für das KAT von der Qualifikationsabteilung jeder Scientology Org oder Gruppe. KAT ist das Akronym für **K**SW 1, **A**rtificial Intelligence **T**ool.

Also ein Werkzeug auf der Basis künstlicher Intelligenz, damit Ron’s Aufforderung an jeden Scientologen umgesetzt werden kann, die Spreu vom Weizen bezüglich der technischen und administrativen Referenzen trennen zu können.

Denn jeder weiss: **Es ist nicht immer LRH drin, wo LRH drauf steht.**

Ich habe mir als Auditor und diplomierter Computer-Ingenieur die Aufgabe gestellt, mittels der modernsten KI-Technologie herauszufinden, welche Texte gesichert von Ron stammen und welche nicht. Es deutet viel darauf hin, dass wir das erreichen können, ohne einer Autorität glauben zu müssen.

### Aktuelle Neuigkeiten:

Nach wochenlanger Arbeit erlangten wir jetzt eine Beta-Version von QualsKAT deren Ergebnisse ich hier im Telegramforum <https://t.me/QualsKAT> veröffentlichen kann.

Wir haben die Sprachmodelle (näheres dazu weiter unten) mit tausenden Seiten von LRH trainiert, so dass sie jetzt seinen literarischen Fingerabdruck kennen und jeden weiteren Text beurteilen können.

Das QualsKAT-Skript scheint gut in der Lage zu sein, einzelne Dokumente sogar schon ab einer Seite Text daraufhin zu bewerten, ob es wahrscheinlich von LRH geschrieben ist oder nicht.

Natürlich können geringfügige Änderung/Revisionen an einer ursprünglich von LRH herausgegebenen Referenz nicht als solche erkannt werden, da bedarf es schon einer inhaltlichen Prüfung.

Aber es gibt eine ganze Reihe Staff Work, die herausgegeben wurden, die vielleicht durchaus Standardtech sind, auf LRH Vorträgen oder Instruktionen basieren oder auf C/S-Anweisung. Aber dann erkennt das Programm am Stil, dass es ist nicht von LRH selbst geschrieben wurde.

Ich habe in einem ersten Schritt die roten Volums aus den 1970ern und in der Ausgabe von 1991 und auch die Management Serie und OEC von 1991 in einzelne Seiten zerlegt und jede einzelne von dem künstlichen Intelligenz System untersuchen lassen.

Anbei eine Tabelle namens ergebnisse.csv mit den Ergebnisse   
von über 26.000 Seiten.

### Erläuterung der Tabelle ergebnisse.csv

Die erste Spalte enthält den Dateinamen der Seite.

z.B. 1965\_69\_Seite\_169.txt

Dieser setzt sich zusammen aus dem Namen der PDF aus dem sie kommt, wie z.B. hier die Datei 1965\_69.pdf, die zu den alten roten Volumes gehört und dann noch der Seitennummer, hier die Seite 169.

Wir haben fünf KI-Modelle eingesetzt und auf LRH-Texte fein abgestimmt, so dass sie quasi den literarischen Fingerabdruck von Ron zu kennen glauben. Diese fünf sind:

* roberta-base
* microsoft\_deberta-base
* distilbert-base-uncased
* roberta-large
* albert-base-v2

Hierbei handelt es sich um verschiedene vortrainierte Sprachmodelle, die auf der Transformer-Architektur basieren. Diese Modelle werden in der natürlichen Sprachverarbeitung (NLP) verwendet, um Aufgaben wie Textklassifizierung, Frage-Antwort-Systeme, Textgenerierung und mehr zu bewältigen.

Die nächsten fünf Spalten B bis F geben für jedes Modell die Prognose, ob der aktuelle Text von LRH stammt oder eher nicht. Die Angabe ist in Prozent, wobei unter 10% sehr unwahrscheinlich bedeutet, über 90% nahezu gesichert.

Es gibt jedoch Texte, die von einzelnen Modellen als 100% Ron gesehen werden und von anderen überhaupt nicht. Deswegen haben wir eine weitere Spalte G angefügt, die den Mittelwert über diese fünf Meinungen bildet. Dies als erste Annäherung zur Findung der Wahrheit.

Die Spalte H namens «Sorted» dient zum Sortieren der Tabelle nach den Bänden und Seitenzahlen, denn leider hatten wir zu Anfang die Seitenzahlen nicht 4-stellig mit führenden Nullen, sondern auf 1 folgte 10 gefolgt von 100, dann 101 bis 199 und dann 2, 20, 200 usw. Daher diese Spalte.

Da die KI einen Tag lang benötigt, um die 26000 Seiten zu bewerten, habe ich diesen Schönheitsfehler erst für künftige Versionen behoben.

Das Programm ist noch in der Entwicklungsphase, ich nenne es mal eine Beta-Version vom Entwicklungsstand 12. August 2024.

### Unerwartete Bewertungen

Die Ergebnisse sind von mir überhaupt noch nicht geprüft worden, was ja auch nicht einfach ist bei so vielen Seiten. Ich bitte die Ergebnisse daher mit Vorsicht zu genießen. Einzelne Stichproben zeigten, dass die Ergebnisse durchaus Sinn machen.

Es gab teilweise auch sehr unerwartete Ergebnisse, wie das «LRH-Vortragstranscript 501101 THE INTENSIVE PROCESSING PROCEDURE.pdf

Dieses wurde von allen 5 Modellen einhellig als Nicht-LRH beurteilt. Leider gab es keine Audiodatei zu diesem Text. Dieses unerwartete Ergebnis hat sich aber letzten Endes doch als richtig erwiesen. Denn es war eben kein 20-seitiges Transkript einen ganzen Vortrags, sondern eine Zusammenfassung auf dreieinhalb Seiten, offenbar von jemandem anderen als LRH.

### Euer Feedback ist essentiell

Ich freue mich auf eure kritischen Prüfungen meiner Arbeit, auf eure Erkenntnisse und euren Input und Feedback, auf Verbesserungsvorschläge oder auch auf vernichtende Beweise, dass das Skript nichts taugt, ich bin für jede Möglichkeit offen. Ich denke aber, dass das angestrebte Ziel im Sinne von KSW (keeping Scientology working) erreicht werden kann: die Spreu vom Weizen trennen zu können.

Ist es wert, sich mit diesem Projekt intensiv zu befassen und es zu einem vollen Erfolg zu bringen? Ich denke schon, gemäss dem HCOPL hängt die ganze Zukunft von Scientology davon ab, dass dieses Ziel erreicht wird.

### Unterstützng nötig

Das Projekt ist weit davon entfernt abgeschlossen zu sein, das ist jetzt gerade mal an der Oberfläche gekratzt. Wir werden das noch weiterhin verbessern. Wer mithelfen will, das zu machen ist herzlich willkommen.

#### Finanziell

Wir brauchen Unterstützung darin, dieses Projekt zu finanzieren. Denn wir mussten speziell für dieses Projekt der künstlichen Intelligenz einen Power-Computer anschaffen. Mit schneller GPU, 64 MB RAM, 2 TB SDD, eine Hardware, die ich normalerweise nicht einsetzen brauch, da ich sonst mit einem normalen Bürocomputer hinkomme. Aber für diese Entwicklung ist ein Power-PC unerlässlich, die schnellsten Systeme einzusetzen. Alleine das kostet schon ein paar tausend Dollar.

Darüber hinaus muss meine Arbeitszeit zumindest zum Teil finanziert werden. Denn andernfalls muss ich mich für die laufenden Kosten um andere Einkommensquellen kümmern und so den Fortschritt dieses Projekts vernachlässigen. Das wäre schade.

Eine gute Investition in die Zukunft von Scientology: <https://www.gofundme.com/f/scientologists-help-develop-tool-to-preserve-l-ron-hubbard>

#### Promotion

Weitere Unterstützungsmöglichkeiten sind in erster Linie das Projekt unter Scientologen breit bekannt zu machen. Auch bei Scientologen, die noch aktiv in der Scientology Kirche tätig sind. Aber auch bei Feld-Auditoren.

#### Testing

Weiter ist es hilfreich, das Skript auf die Probe zu stellen. Das beinhaltet zum Beispiel die Tabelle ergebnisse.csv anzuschauen und zu gucken, ob die Ergebnisse stimmig sind oder zu recherchieren, wo sie unstimmig erscheinen. Ob es Fehleinschätzungen der AI gibt:

* es gibt zwei Möglichkeiten von Fehlern, die das Programm macht nämlich False Positives, das heißt das Programm hat eine Textseite als gültigen LRH-Text bezeichnet, wo es offensichtlich nicht der Fall ist.
* Und dann gibt es False Negativ. Wo das Programm einen Text   
  als nicht-LRH bezeichnet hat, obwohl er definitiv von Ron zu sein scheint.

Letztendlich können wir nur sicher sein bei Texten, die Ron als Vortrag   
geliefert hat, also wo ein Audio Tape zur Verfügung steht, von dem es ein Transkript gibt bzw. wir ein Transkript erstellen können (dafür haben wir ebenfalls eine Software auf AI-basis erstellt).

Nur diese Texte können als gesichert gelten, da wir Ron hören können, da all diese Audios schon seit Mitte letzten Jahrhunderts auf dem Markt sind und damals konnte man noch nicht – wie in diesem Jahrhundert – Audios unbemerkt fälschen. Das wäre ein weiteres Projekt für uns, auch Audio-Dateien mittels AI und Sprachmodellen auf den Sprecher LRH hin zu bewerten.

Aber alle Bücher, Bulletins, Policies und Artikel: da muss mit gerechnet werden, das Staff-Work eingang gefunden haben. Texte die Ron vielleicht vorgelegt wurden und von ihm genehmigt wurden (oder auch nicht). Die aber nicht von ihm geschrieben sind. Was nicht unbedingt heißt, dass es Out-Tech ist. Aber es könnte - da es nicht von Ron ist - sich auch nach inhaltlicher Prüfung als Out-Tech erweisen. Auch andere Autoren haben das Recht zu schreiben. Aber es gab mal eine Zeit, als durch BTBs und HCOBs die Urheberschaft eindeutig dargestellt wurde. 1991 wurden viele BTBs aus der damaligen Zeit als HCOBs neu herausgegeben.

Soweit erstmal zum aktuellen Stand des Projektes und der vorgelegten Tabelle.   
Input für diese Tabelle waren ein paar Dutzend PDF-Dateien von den besagten roten und grünen Bänden, die seit Jahren digitalisiert vielen Scientologen vorliegen, auf ihren Computern gespeichert sind und genutzt werden.

Ich gebe hier noch mal die genauen Dateiennamen an:

Die Roten Volumes aus den 1970ern wurden 1995 erstellt und haben diese Dateinamen:

* 50\_53.PDF
* 54\_56.PDF
* 57\_59.PDF
* 60\_61.PDF
* 62\_64.PDF
* 65\_69.PDF
* 70\_71.PDF
* 72\_76.pdf
* 76\_78.PDF
* 78\_79.PDF

Die Roten Bände, die 1991 vom RTC publiziert wurden, heissen (am 29. 9. 2006 digitalisiert und gespeichert):

* Subject Volume 1.pdf
* Subject Volume 2.pdf
* Subject Volume 3.pdf
* Subject Volume 4.pdf
* V01 1950-1952.pdf
* V02 1953-1954.pdf
* V03 1955-1956.pdf
* V04 1957-1958.pdf
* V05 1959-1960.pdf
* V06 1961-1962.pdf
* V07 1963-1965.pdf
* V08 1966-1969.pdf
* V09 1970-1971.pdf
* V10 1972-1976.pdf
* V11 1977-1979.pdf
* V12 1980-1984.pdf
* V13 1985-1991.pdf

Seither gibt es neue HCOBs, sogar aus den 2000er Jahren. Ich freue mich über Kopien von Dir, damit ich auch diese einer Prüfung unterziehen kann. Sie müssen nicht von vornherein Nicht-LRH-Texte sein, sondern könnten Auszüge aus Vorträge oder Büchern sein, die heute zum HCOB oder HCOPL aufgewertet wurden.

Dann habe ich noch diese neuen Grünen Bände von 1991 mit den Richtlinien der AI zur Bewertung vorgelegt (sie tragen ein Scan-Datum vom 15. 1. 2007):

* Management Series V1.pdf
* Management Series V2.pdf
* Management Series V3.pdf
* V00 Basic Staff Hat.pdf
* V01 HCO Division.pdf
* V02 Dissemination Division.pdf
* V03 Treasury Division.pdf
* V04 Technical Division.pdf
* V05 Qualifications Division.pdf
* V06 Public Division.pdf
* V07 Executive Division.pdf

Alle diese Dateien habe ich mit dem anliegenden Python-Skript namens   
pdf2singlePages.py erstellt (anbei). Jeder mit Python vertraute Programmierer kann also dieses Programm selbst verwenden, um die zigtausend Einzel Dateien zu erstellen. Aus verständlichen Gründen, um rechtliche Angriffe wegen mutmaßlichen oder behaupteten Copyright-Verstößen zu entgehen, werde ich diese Einzeldateien oder auch die gesamten Dateien nicht zur Verfügung stellen können, bitte versteht und respektiert diese Entscheidung, nimmt das nicht persönlich wenn ich etwaige Bitten abschlagen muss.

Es gibt bestimmt viele Dutzend Leute, die in der Lage sind, diese einzelnen Dateien mit dem Script zu erstellen. Oder du kannst PDF-SAM verwenden, eine kostenlose Software.

Und dann in dem Verzeichnis mit den 26000 Textdateien das folgende Skript verwenden, das csv2odsConvert-Sorted.py heißt. Das wird aus der ergebnisse.csv eine excel-ähnliche Tabelle erschaffen: ergebnisse.ods

Ähnlich der Tabelle in der Anlage, mit der man die einzelnen Seiten, die einen interessieren, direkt über einen Mausklick öffnen kann.

Wenn es noch Fragen zu diesem Vorgehen gibt oder Input: bitte hier im Telegramm Forum <https://t.me/QualsKAT> da schaue ich regelmäßig herein, veröffentlichte den neuesten Stand, neue Erkenntnisse und versuche die aufkommenden Fragen zu beantworten. Das Forum ist öffentlich. Dieser Text darf auch verbreitet werden, um das Projekt bekannt zu machen.

**Andreas Gross**

Adresse: Althusweg 12, CH-6315 Morgarten, Schweiz.

Email: Andreas @ FreieScientologen.de

### Anhang

* ergebnisse.csv
* ergebnisse.ods
* pdf2singlePages.py
* pdf2singlePages.md
* csv2odsConvert-Sorted.py
* csv2odsConvert-Sorted.md