

# Hurray, hurray: The beta version is here

Andreas Gross on August 13, 2024 about the QualsKAT project

## What is QualsKAT

**QualsKAT** is the abbreviation for the **KAT** of the **Qual**ification Department of each Scientology Org or Group. **KAT** is the acronym for **KSW 1**, **Artificial Intelligence Tool**.

In other words, a tool based on artificial intelligence so that Ron's call to every Scientologist to be able to separate the wheat from the chaff in terms of technical and administrative references can be implemented.



As we all know: **Just because it says LRH on the tin, doesn't mean it's LRH inside.**

As an auditor and qualified computer engineer, I have set myself the task of using the latest AI technology to find out which texts are verified as coming from Ron and which are not. There are many indications that we can achieve this without having to believe an authority.

## Latest news:

After weeks of work, we have now obtained a beta version of QualsKAT, the results of which I can publish here in the Telegramforum <https://t.me/QualsKAT>.

We have trained five language models (more on this below) with thousands of pages of LRH, so that they now know his literary fingerprint and can assess any further text.

The QualsKAT script seems to be well capable of evaluating individual documents, even from as little as one page of text, as to whether or not they are probably written by LRH.

Of course, minor changes/revisions to a reference originally issued by LRH cannot be recognized as such; a substantive review is required.

But there is a whole series of staff work that has been published, which may well be standard tech, based on LRH lectures or instructions or on C/S instructions. But then the program recognizes by the style that it was not written by LRH himself.

As a first step, I broke down the red volumes from the 1970s and the 1991 edition and also the Management Series and OEC from 1991 into individual pages and had each one examined by the artificial intelligence system.

Enclosed is a table called `ergebnisse.csv` with the results from over 26,000 pages.

## **Explanation of the table `results.csv`**

The first column contains the file name of the page.

e.g. `1965_69_page_169.txt`

This is made up of the name of the PDF from which it comes, e.g. here the file `1965_69.pdf`, which belongs to the old red volumes, and then the page number, here page 169.

We used five AI models and fine-tuned them to LRH texts so that they virtually think they know Ron's literary fingerprint. These five are:

- `roberta-base`
- `microsoft_deberta-base`
- `distilbert-base-uncased`
- `roberta-large`
- `albert-base-v2`

These are various pre-trained language models based on the Transformer architecture. These models are used in natural language processing (NLP) to handle tasks such as text classification, question-answer systems, text generation and more.

The next five columns B to F give the prediction for each model as to whether the current text originates from LRH or not. The information is given as a percentage, with less than 10% meaning very unlikely and more than 90% almost certain.

However, there are texts that are seen as 100% Ron by some models and not at all by others. We have therefore added a further column G, which forms the average of these five opinions. This is a first approximation for finding the truth.

Column H called "Sorted" is used to sort the table according to the volumes and page numbers, because unfortunately at the beginning we did not have the page numbers in 4 digits with leading zeros, but on 1 followed 10 followed by 100, then 101 to 199 and then 2, 20, 200 and so on. Hence this column.

Since the AI needs a day to evaluate the 26000 pages, I have fixed this flaw for future versions.

The program is still in the development phase, I'll call it a beta version as of August 12, 2024.

## **Unexpected ratings**

The results have not been checked by me at all, which is not easy with so many pages. I would therefore ask you to treat the results with caution. Individual samples have shown that the results do make sense.

There were also some very unexpected results, such as the LRH lecture transcript «501101 THE INTENSIVE PROCESSING PROCEDURE.pdf»

This was unanimously assessed as non-LRH by all 5 models.

Unfortunately, there was no audio file for this text. However, this unexpected result turned out to be correct in the end. It was not a 20-page transcript of an entire lecture, but a three-and-a-half page summary, apparently by someone other than LRH.

## **Your feedback is essential**

I look forward to your critical reviews of my work, to your insights and your input and feedback, to suggestions for improvement or even to

devastating evidence that the script is no good, I am open to every possibility. However, I think that the desired goal can be achieved in the spirit of KSW (keeping Scientology working): to be able to separate the wheat from the chaff.

Is it worth working hard on this project and making it a complete success? I think so, according to the HCOPL the whole future of Scientology depends on this goal being achieved.

## **Support needed**

The project is far from complete, we have only just scratched the surface. We will continue to improve it. Anyone who wants to help with this is very welcome.

### ***Financial***

We need support to finance this project. Because we had to buy a power computer especially for this artificial intelligence project. With a fast GPU, 64 MB RAM, 2 TB SDD, hardware that I don't normally need to use, as I can usually get by with a normal office computer. But for this development, a power PC is essential to use the fastest systems. That alone costs a few thousand dollars.

In addition, my working hours must be at least partially financed. Otherwise, I will have to look for other sources of income to cover the running costs and neglect the progress of this project. That would be a shame.

A good investment into the future of Scientology:

<https://www.gofundme.com/f/scientologists-help-develop-tool-to-preserve-l-ron-hubbard>

### ***Promotion***

Other possibilities for support are primarily to make the project widely known among Scientologists. Also among Scientologists who are still active in the Church of Scientology. But also with field auditors.

### ***Testing***

It is also helpful to put the script to the test. This includes, for example, looking at the results.csv table and seeing whether the results are

consistent or researching where they appear inconsistent. Whether there are misjudgements by the AI:

- There are two possible errors that the program makes, namely false positives, i.e. the program has designated a page of text as valid LRH text where this is obviously not the case.
- And then there is False Negative. Where the program has labeled a text as non-LRH, even though it definitely appears to be from Ron.

Ultimately, we can only be sure with texts that Ron has delivered as a lecture, i.e. where an audio tape is available from which there is a transcript or we can create a transcript (we have also created AI-based software for this).

Only these texts can be considered reliable, as we can hear Ron, because all these lectures have been published since the middle of the last century and at that time it was not yet possible - as in this century - to falsify audios without being noticed. This could be another project for us to evaluate audio files for the speaker LRH using AI and language models.

But all books, bulletins, policies and articles: it must be assumed that staff work has found its way in. Texts that may have been submitted to Ron and approved by him (or not). But which are not written by him. Which does not necessarily mean that it is out-tech. But it could - since it's not from Ron - turn out to be out-tech even after checking the content. Other authors also have the right to write. But there was a time when BTBs and HCOBs clearly showed the authorship. In 1991, many BTBs from that time were republished as HCOBs.

So much for the current status of the project and the table presented. The input for this table was a few dozen PDF files from the aforementioned red and green volumes, which have been available in digital form to many Scientologists for years, are stored on their computers and are used.

I'll give you the exact file names again here:

The Red Volumes from the 1970s were created in 1995 and have these file names:

- 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

The Red Volumes, published by the RTC in 1991, are called (digitized and saved on 29. 9. 2006):

- 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

Since then there have been new HCOBs, even from the 2000s. I would be happy to receive copies from you so that I can check these too. They don't have to be non-LRH texts from the outset, but could be excerpts from lectures or books that have been upgraded to HCOB or HCOPL today.

Then I submitted these new Green Volumes from 1991 with the AI guidelines for evaluation (they have a scan date of January 15, 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

I created all files with the Python script called pdf2singlePages.py (attached). Any programmer familiar with Python can use this program to create the thousands of single files. For understandable reasons, in order to avoid legal attacks for alleged or claimed copyright violations, I will not be able to provide these single files or even the entire files, please

understand and respect this decision, do not take it personally if I have to refuse any requests.

There exist many dozens of people who are able to create these individual files with the script. Or you can use PDF-SAM, a free software.

And then in the directory with the 26000 text files use the following script called csv2odsConvert-Sorted.py. This will create an Excel-like table from the results.csv: results.ods

Similar to the table in the appendix, with which you can open the individual pages that interest you directly with a mouse click.

If there are any questions about this procedure or input: please visit the Telegram forum <https://t.me/QualsKAT> where I regularly check in, publish the latest status, new findings and try to answer any questions that arise. The forum is public. This text may also be distributed to publicize the project.

## **Andreas Gross**

Address: Althusweg 12, CH-6315 Morgarten, Switzerland

Email: Andreas @ FreieScientologen.de

## **Appendix**

- results.csv
- results.ods
- pdf2singlePages.py
- pdf2singlePages.md
- csv2odsConvert-Sorted.py
- csv2odsConvert-Sorted.md



