

Introducing eSearch Pro

Update: 2025-02-21 Issue 12

The first public beta
30-day trial was launched

30 September 2024



eSearch Pro 1.0 is a work in progress and still in the beta test phase.
Some functionality may be limited.

Introduction

eSearch Pro is a comprehensive information retrieval application featuring an AI Mode that can connect to multiple online AI services, including all Perplexity web search models. It has the unique ability to be translated by users. German and French sample language files are provided, along with a free Language File Translator, enabling quick edits to these or other partially translated language files.

eSearch Pro is being developed as a valuable alternative to dtSearch Desktop for our European customers, with the hope that it will also attract global interest.

We have been distributing dtSearch products in Europe for 30 years, but they are only available in English. Although recent versions of dtSearch utilize our multilingual stemming files, licensing restrictions prevent us from offering translated versions.

While eSearch Pro lacks some of the eForensic features of dtSearch, we aim to provide functionality and new features that make it suitable for applications where dtSearch or others may be less effective. Here we present a few ideas...

Contents

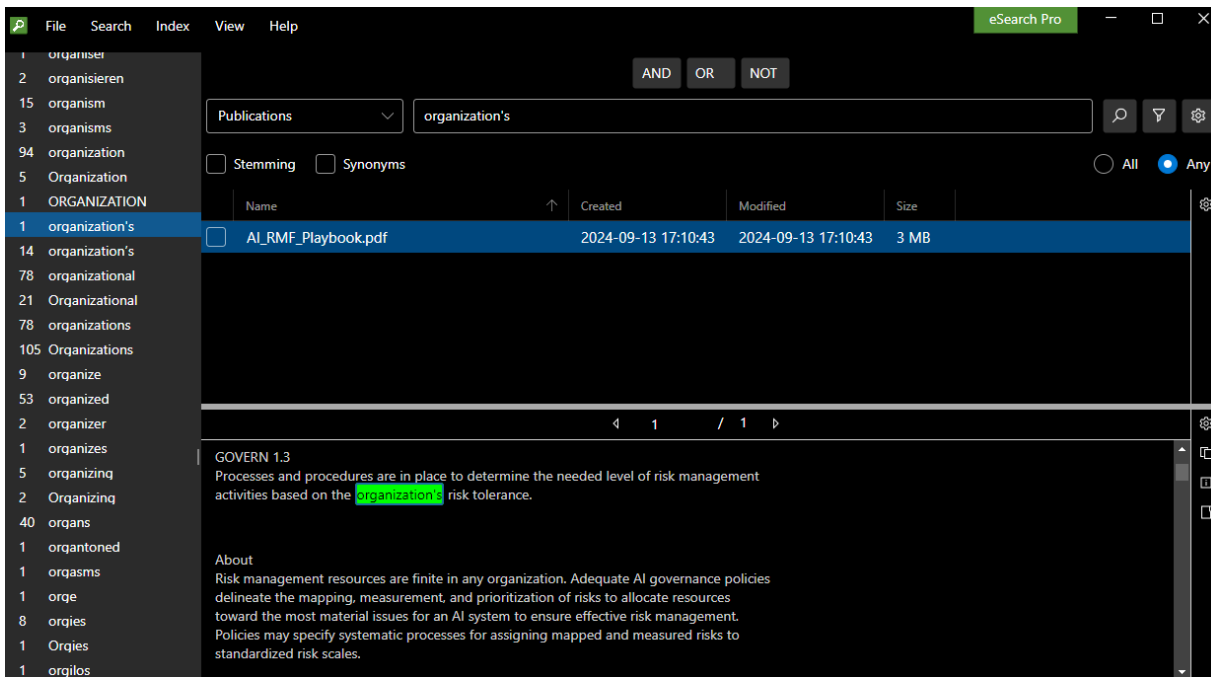
Introduction	2
Viva la difference!.....	3
Proof Reading.....	4
Geolocation in images	5
Locate Numeric Patterns.....	6
Search for Email addresses.....	6
Translation.....	7
Software Engineering	8
Large CSV Datasets	9
Example - ICIJ files.....	9
Example - TREC 2.....	10
Research or Publish Newsletters and Notebooks.	11
AI Online Search	13
Perplexity Sonar & Sonar Pro LLMs.....	13
Perplexity Sonar Reasoning	14
Multiple AI services.....	15
ChatGPT	15
Custom AI	15
30-day trial now available!.....	16

Viva la difference!

Here are a few notable differences between eSearch Pro and dtSearch:

- A valuable feature that dtSearch has, but others lack is a word list of all the words in the index. eSearch Pro takes the feature several steps further:
 - Full length fully scrollable word list on the main window.
 - Sorted alphabetically for accented characters & uppercase letters.
 - Retains punctuation within numbers, the possessive apostrophe and contractions.
- Image formats (such as .png .jpg .gif .bmp .svg .tif) are displayed in the viewer; the viewer can also display a marker at the geolocation coordinates of an image on a built-in map view.
- Options for ‘Search as you Type’ and ‘List Index Contents on Empty Query’ (You can add words to limit the results to just the ones you need).
- Viewer plays browser compatible audio and video file formats like .WebM, .Ogg, .Wav, MP3. Just click ‘open in folder’ to play other formats in their default application or the recommended app.
- Source code can be searched and displayed with syntax highlighting and line numbers, 40+ source code file types supported.
- A filetype tree-view selector makes it easy for you to limit your index contents to just the filetypes of interest without having to move the files.
- ePub, Jupyter Notebook, Markdown file support.
- Portable version is self-contained and will run on a USB flash drive. You can index & search files stored on the host machine’s hard drive or a portable library on the USB flash drive itself for your own use or to publish search-only copies to others.
- Search results include not only a file’s Modified date and time, but also the Created and last Accessed date and time, all in ISO 8601 format (YYYY-MM-DD).
- Hit navigation includes current hit number as well as total count.
- Did we mention AI functionality or dark mode as default?

Proof Reading

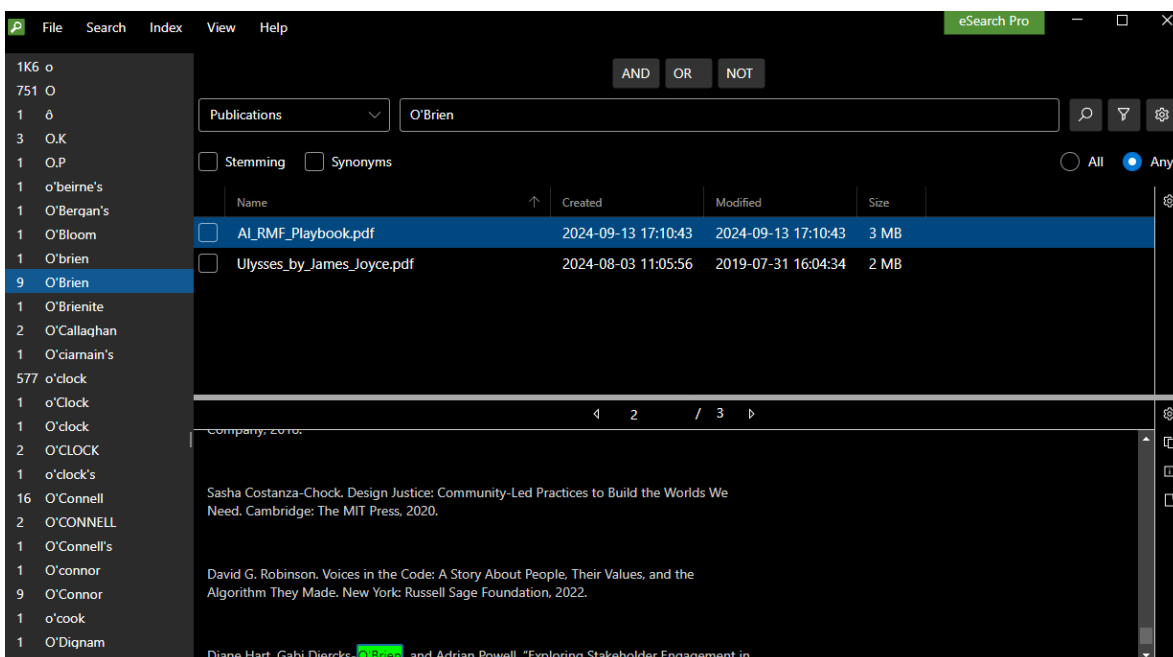


Whether you are a professional proofreader, a translator, or a just a nitpicker, you will love the scrollable word list that shows every word in the documents. Simply double-click on a word to highlight it.

The default indexing method is 'accent sensitive' so that cafe doesn't pass for café.

Case Sensitive indexing option ensures that references to a Polish delicacy doesn't get mistaken for a kind of wax polish.

Indexing preserves punctuation within numbers and words like dates, IP addresses, library formats, sums of money, or abbreviations. Possessive apostrophes with different symbols (U+0027 or U+2019 as used in the NIST AI RMF Playbook shown above) are easily spotted.



Geolocation in images

eSearch Pro displays images and automatically extracts metadata such as geolocation, camera settings and more. You can choose the specific types of metadata you want to see in your results columns. Additionally, you can click and drag header columns to reorder them and sort the results based on the selected metadata key.

The screenshot shows the eSearch Pro interface with a search for 'Germany*'. The results table is as follows:

Name	Exif/DateTimeOrig	Exif/Latitude	Exif/Longitude	Modified	Type
germany-allgaeu-fliegenpilz.jpg	30/09/2018 13:43	47.46690555555555	10.20375	2024-08-24 17:32:35	jpg
germany-english-garden.jpg	04/11/2018 12:09	48.14572777777778	11.58768888888889	2024-08-24 17:32:49	jpg
germany-garching-heide.jpg	29/08/2018 19:31	48.26827499999999	11.60336111111111	2024-08-24 17:32:11	jpg

A red box highlights the 'Image' icon in the top right corner of the results table. Below the table, a preview of a waterfall image is shown.


You can also choose to display the geolocation on a map!

The screenshot shows a list of search results with geolocation data. The results table is as follows:

Size	Name	Exif/Latitude	Exif/Longitude	Exif/Altitud	FNumber	Model
836 KB	japan-katsura-river.jpg	35.014377	135.669015	-	2	MI 5
506 KB	turkey-bodrum.jpg	37.02995277777778	27.41326388888889	79.19	1.8	Pixel 2
794 KB	italy-garda-lake-sailing-club.jpg	45.87763055555555	10.85716111111111	71.95	1.8	Pixel 2
1 MB	germany-allgaeu-fliegenpilz.jpg	47.46690555555556	10.20375	1601.74	1.8	Pixel 2
2 MB	germany-english-garden.jpg	48.14572777777778	11.58768888888889	558.85	1.8	Pixel 2
627 KB	germany-garching-heide.jpg	48.26827499999996	11.60336111111111	540.05	1.8	Pixel 2
582 KB	england-london-bridge.jpg	51.50410555555554	-0.074575	77.88	1.8	Pixel 2
678 KB	ireland-dingle.jpg	52.139276657230475	-10.274594797178132	-	4	DMC-FX60

Below the table, an OpenStreetMap preview is shown, displaying the geolocation of the selected image (turkey-bodrum.jpg) on a map of Bodrum, Turkey.

Displays images and lists metadata in these formats: .bmp .gif .jpg .pgm .png .svg .tiff.

 TIP: Keep images in separate indexes and press the Search Metadata button to display the image metadata in the wordlist.

Locate Numeric Patterns

eSearch indexes numbers by default and preserves punctuation within numbers. This makes it possible to search for patterns within them using Regular Expressions.

eSearch includes an example regular expression in the Sample Numeric Patterns synonyms files to identify IPv4 internet addresses. You can also add your own regular expressions, it uses the same Lucene syntax as Microsoft Azure AI Search.

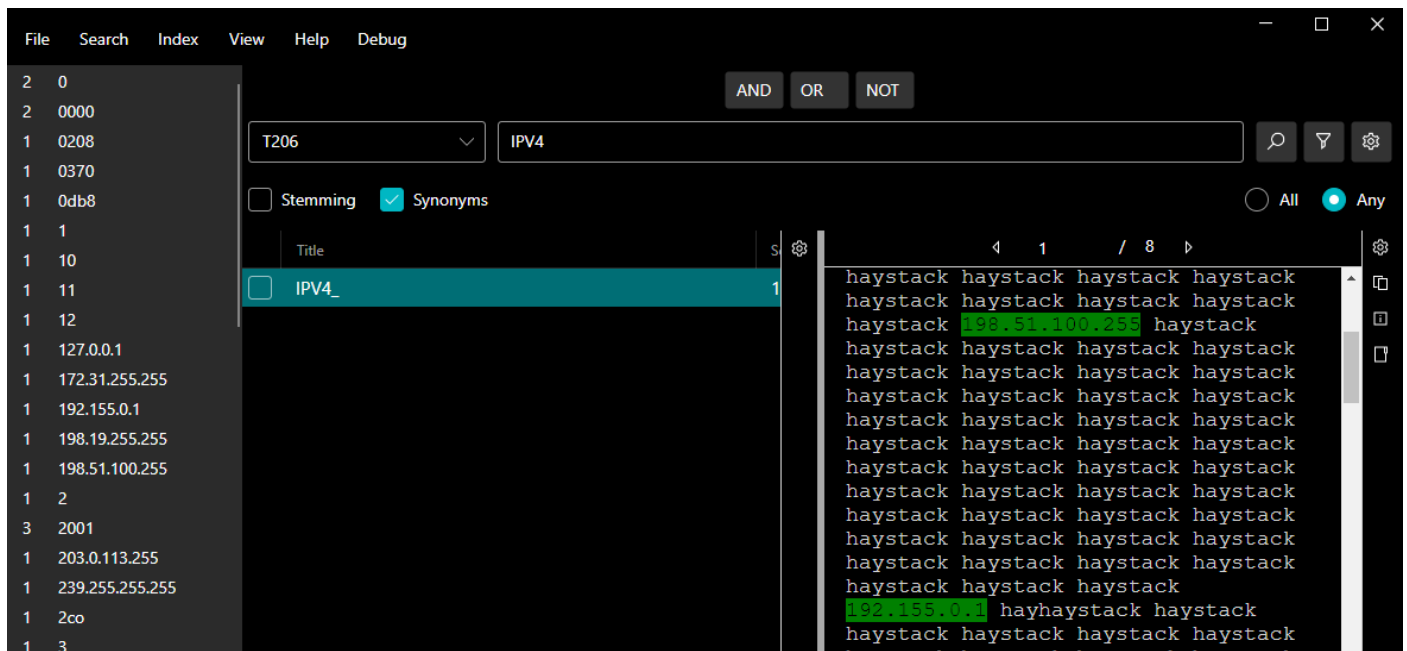


Image above shows IPV4 search query used to find and highlight all IPv4 addresses in a file.

Search for Email addresses

Email addresses are split at the @ symbol, so domains and email aliases appear separately in the word list.

You can search for all users at contoso.com by entering that domain as the search query (without using a wildcard), or you can enter the alias jon.doe to find all such emails at any domain, e.g., jon.doe@gmail.com or jon.doe@contoso.com

To search for the exact email use both parts in quotes including the @ sign like "jon.doe@contoso.com".

Unlike dtSearch, there is no need to specify an option in the index settings to index emails, it is the default behaviour.

Translation

eSearch Pro supports .xliff (1.2 & 2.0), .tbx and .tbm file formats used by translation software. The index is accent sensitive by default and Index Settings offers a Case Sensitive option, so that for example Language, language and LANGUAGE get indexed and appear in the word list as three separate words.

Multilingual stemming means that plurals and other word variations are found in many European languages.

Pre-defined synonym files are included that cover irregular verbs and nouns in Danish, Dutch, English, French, German, Italian, Norwegian, Spanish, and Swedish. You can choose to use several synonym files during a search. Synonym files also include multilingual days and months.

The word list displays contractions like d'un and possessive apostrophes.

The screenshot shows the eSearch Pro interface. On the left is a word list with 'd'un' selected. The main area shows search results for 'd'un' in a file named 'sample-1-xliff2.xliff'. The results are displayed in a monospaced font, showing the original text and the translated text.

Name	Created	Modified	Size
sample-1-xliff2.xliff	2024-09-04 16:43:12	2024-09-04 16:43:12	656 B

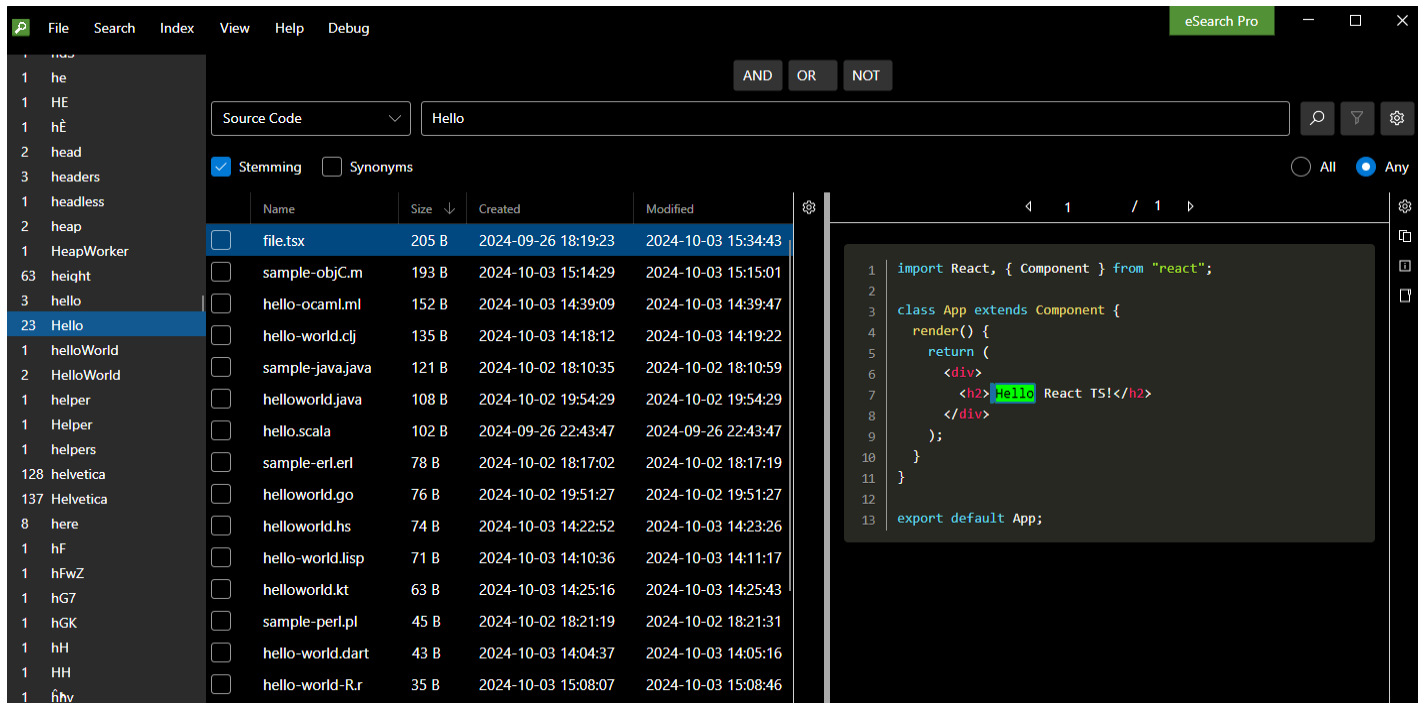
```
Hello world
Bonjour le monde
This is an XLIFF document.
Il s'agit d'un document XLIFF.
```

*As of 1.0.9176 (beta) release it is only suitable for small files, eSearch Pro currently does not have a way to segment large files.

The free GSsplit 3.0 by G.D.G. Software SARL can be used for this purpose. <https://www.gdgsoft.com/gsplit> The UI is available in many languages.

Software Engineering

Over 40 source code file formats are supported, the viewer displays source code with hit highlighting, syntax highlighting & line numbers.



Example shows Created and Modified date & time all in ISO 8601 format (YYYY-MM-DD), file size in Bytes and filename.

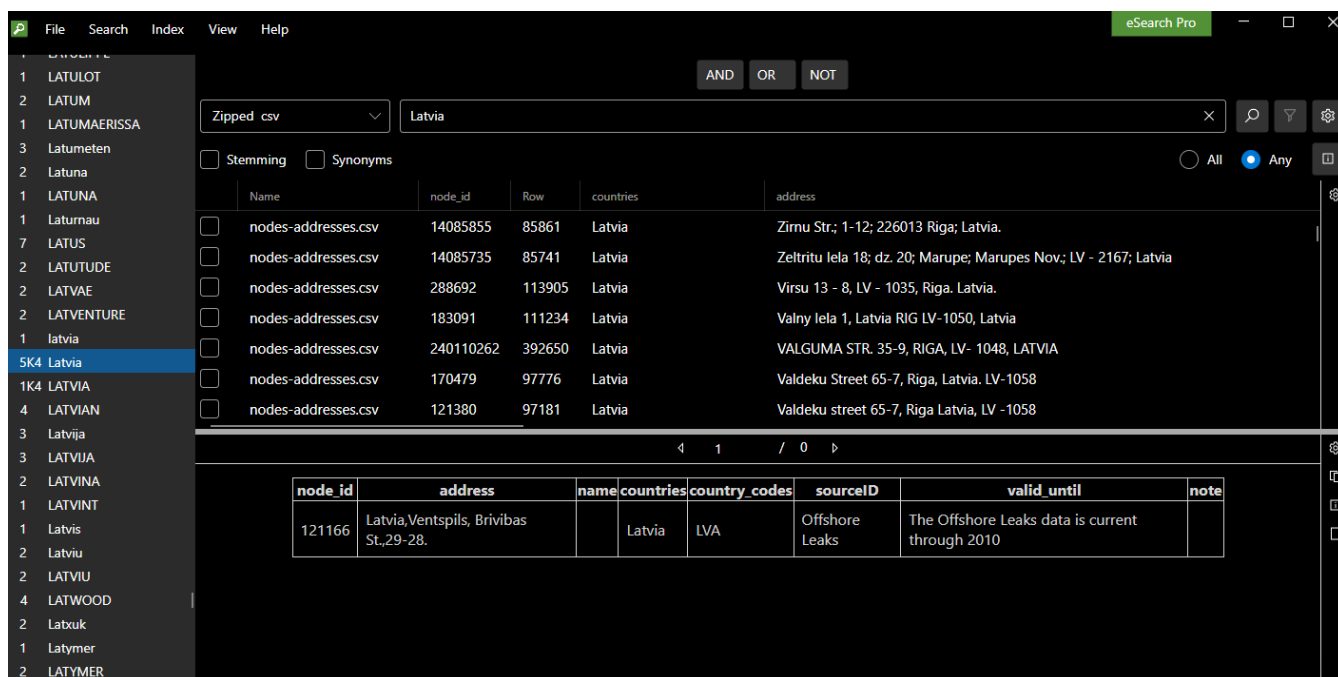
eSearch Pro 1.0.9120 (beta) supports .asm .bat .c .clj .coffee .cpp .cs .dart .elm .erl .fs .go .groovy .hs .java .jl .js .jsx .kt .lisp .lua .m .ml .php .pl .pro .ps1 .py .r .rb .rkt .rs .sass .scala .scss .sh .sql .swift .ts .tsx .v .vb .vhd

Large CSV and JSONL Datasets

eSearch Pro can index .csv files inside an archive such as Zip. It will automatically index all metadata and display each record separately. You can select which columns to display in the search results.

Update: JSONL files are now supported with each row indexed as a separate document, eSearch can handle large datasets such as <https://nexla.com/cord-19-research-data-different-formats-avro-parquet-jsonl/>

Example - ICIJ files.



The screenshot shows the eSearch Pro interface with a search for 'Latvia' in a 'Zipped csv' dataset. The search results are displayed in a table with columns: Name, node_id, Row, countries, and address. The results list several records from 'nodes-addresses.csv' with various addresses in Latvia. A detailed view of a record is shown at the bottom, with columns: node_id, address, name, countries, country_codes, sourceID, valid_until, and note.

node_id	address	name	countries	country_codes	sourceID	valid_until	note
121166	Latvia,Ventspils, Brivibas St.,29-28.		Latvia	LVA	Offshore Leaks	The Offshore Leaks data is current through 2010	

Example above from International Consortium of Investigative Journalists Offshore Leaks zip file. <https://offshoreleaks.icij.org/pages/database>

The 70MB 'full-olddbLATEST.zip' file includes Offshore Leaks (2013) Panama Papers (2016) Bahamas Leaks (2016) Paradise Papers (2017) Pandora Papers (2021), a total of 5,356,936 records. On a Windows 11 Pro Intel i7 32Gb RAM machine a case sensitive index of that zip file with no stop words removed takes approximately 4 minutes to index.

Example - TREC 2.

The screenshot shows the eSearch Pro interface with the search query "Musk AND battery" and the dataset "TREC 2". The search results table is as follows:

Name	Row	msg_dt	message
trc2.csv	1588416	2009-01-13 17:55:50	AUTOSHOW-UPDATE 1-Tesla to supply battery for electric Smart
trc2.csv	1297941	2008-11-03 14:45:21	Tesla secures \$40 million in convertible financing
trc2.csv	1588233	2009-01-13 17:04:58	BRIEF-Tesla partnering with Daimler on electric Smart-Musk
trc2.csv	1588776	2009-01-13 19:50:41	AUTOSHOW-UPDATE 2-Tesla to supply battery for electric Smart
trc2.csv	1207751	2008-10-15 20:02:44	Tesla to delay new model, cut jobs to save cash
trc2.csv	1286239	2008-10-31 05:53:20	Tesla sees new financing, founder offers support
trc2.csv	1208014	2008-10-15 20:52:24	UPDATE 1-Tesla to delay new car, cut jobs to save cash

The detailed view of the selected record (row 1297941) shows the following story text:

DETROIT, Nov 3 (Reuters) - Tesla Motors Inc said on Monday that its board had approved \$40 million in convertible debt financing to push ahead with stepped-up manufacturing of its battery-powered Roadster sports car. Last week, Tesla founder Elon Musk had said the electric car start-up was moving to close a new financing round from existing investors to bolster a cash balance that had dropped to \$9 million. San Carlos, California-based Tesla said it had commitments for the new convertible debt financing from most of its current, major investors. 'Forty million is significantly more than we need,' Musk said in a statement. 'However, the board, investors and I felt it was important to have significant cash reserves.' Tesla markets the \$109,000 Roadster, the first battery-powered performance car and a vehicle that has become a symbol of both the promise of electric cars and the financial risks of rushing the new technology to the market. Tesla has delivered

TREC 2. News content copyright Thompson Reuters.

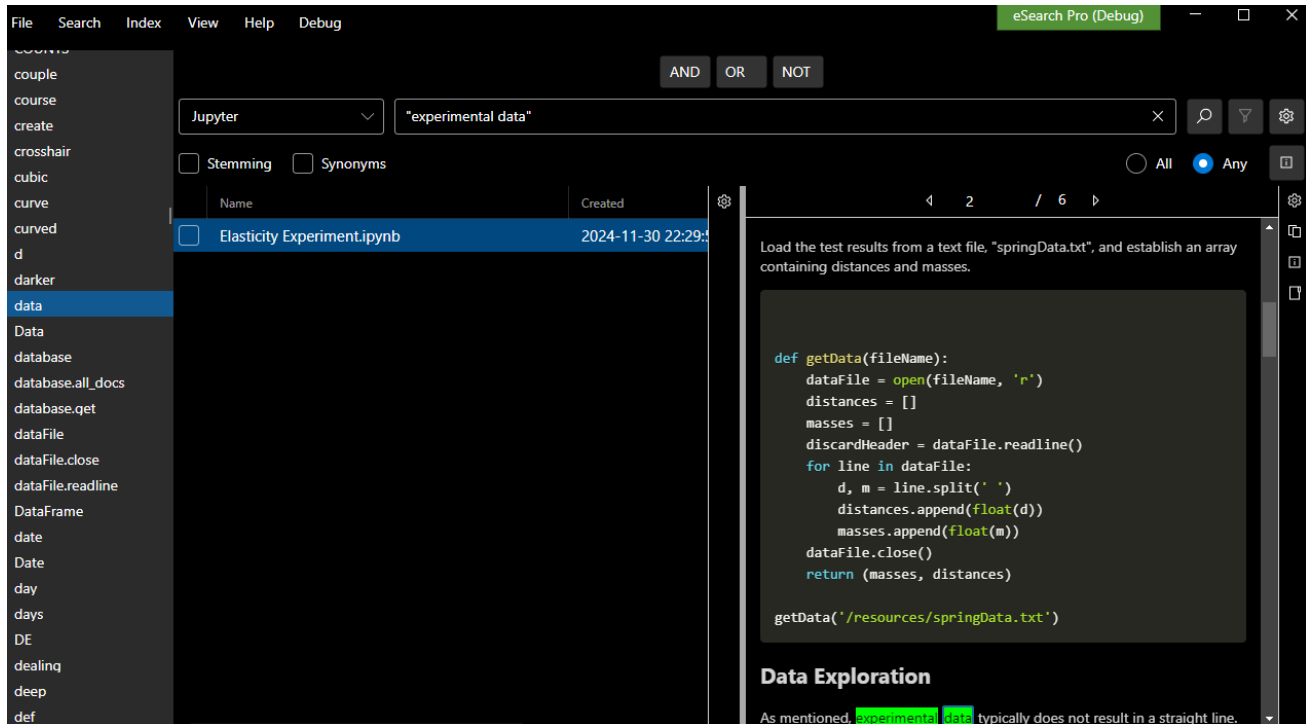
Test results:

eSearch Pro 1.0.9077. 8 minutes to index 1,800,370 records (2.67GB .csv file). Stop words = None. Case Sensitive = yes. No limit on word length or file size. Lenovo Think Station. Processor: 12th Gen. Intel Core i9.12900 2.4GHz. Windows 11 Pro 23H2 Storage 951GB, 680GB free. RAM 32GB (31.7 usable)

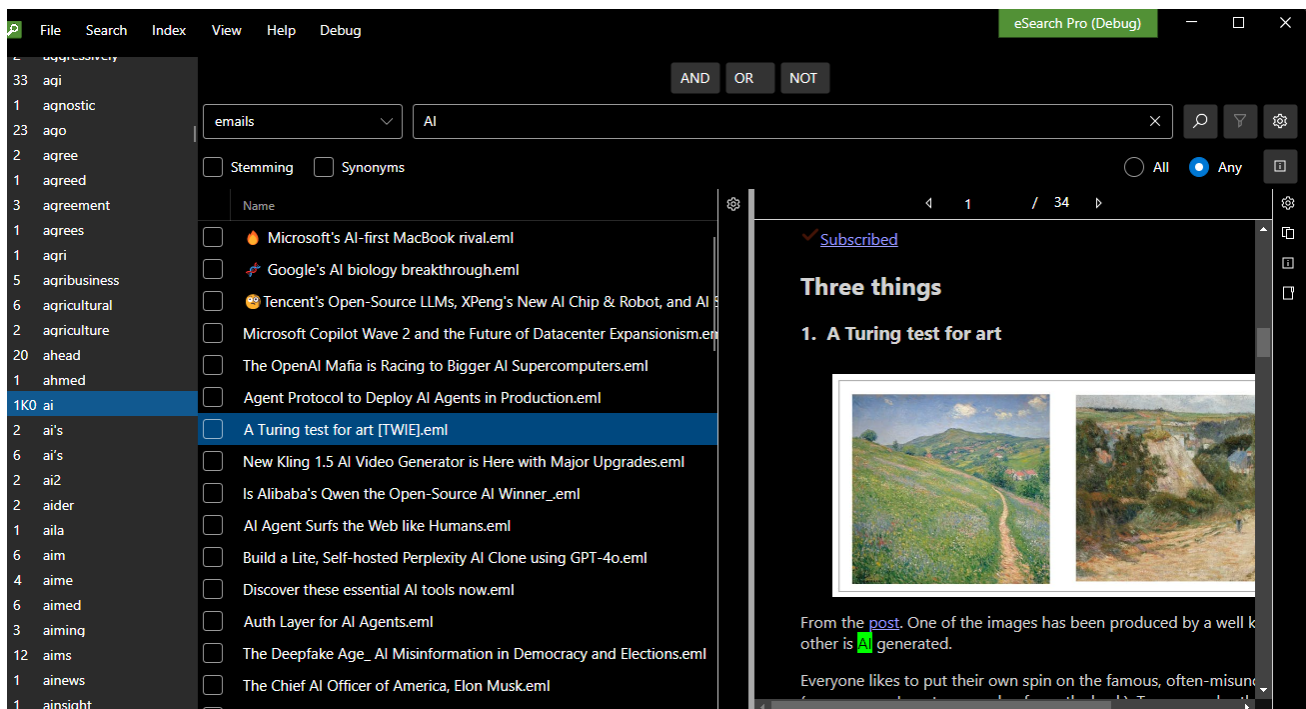
Research or Publish Newsletters and Notebooks.

There is an explosion of online material available through newsletters and Notebooks created by content creators on platforms like Substack, Bluesky, BeeHiiv, and others. Additionally, over 20,000 free papers are published each month on arXiv.org and other pre-print services.

eSearch Pro includes support for Jupyter Notebooks and Markdown files, as well as .eml files generated by Outlook's Save-As feature.



Example Jupyter Notebook (.ipynb) file display showing hit highlighting.



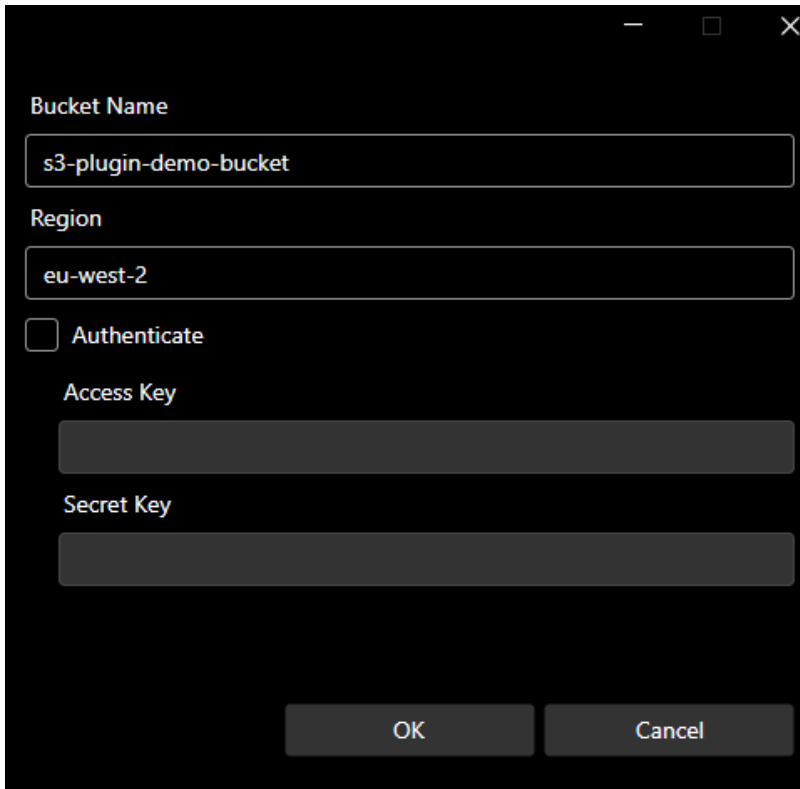
Example email newsletter research. eSearch Portable – Search-only license lets you publish your own searchable Notebook collection.

Plugins

eSearch Pro supports plugins to extend its functionality.

For example, it's possible to index files outside of the local file system using a plugin.

A free plugin that can index files on an Amazon S3 store is available from here: <https://searchcloudone.com/plugins-for-esearch-pro/>



Bucket Name
s3-plugin-demo-bucket

Region
eu-west-2

Authenticate

Access Key
[Empty text box]

Secret Key
[Empty text box]

OK Cancel

Acknowledgement: The files in the S3 demonstration bucket are a collection of files from the open-source and free course by Decoding ML here: <https://github.com/decodingml/second-brain-ai-assistant-course>

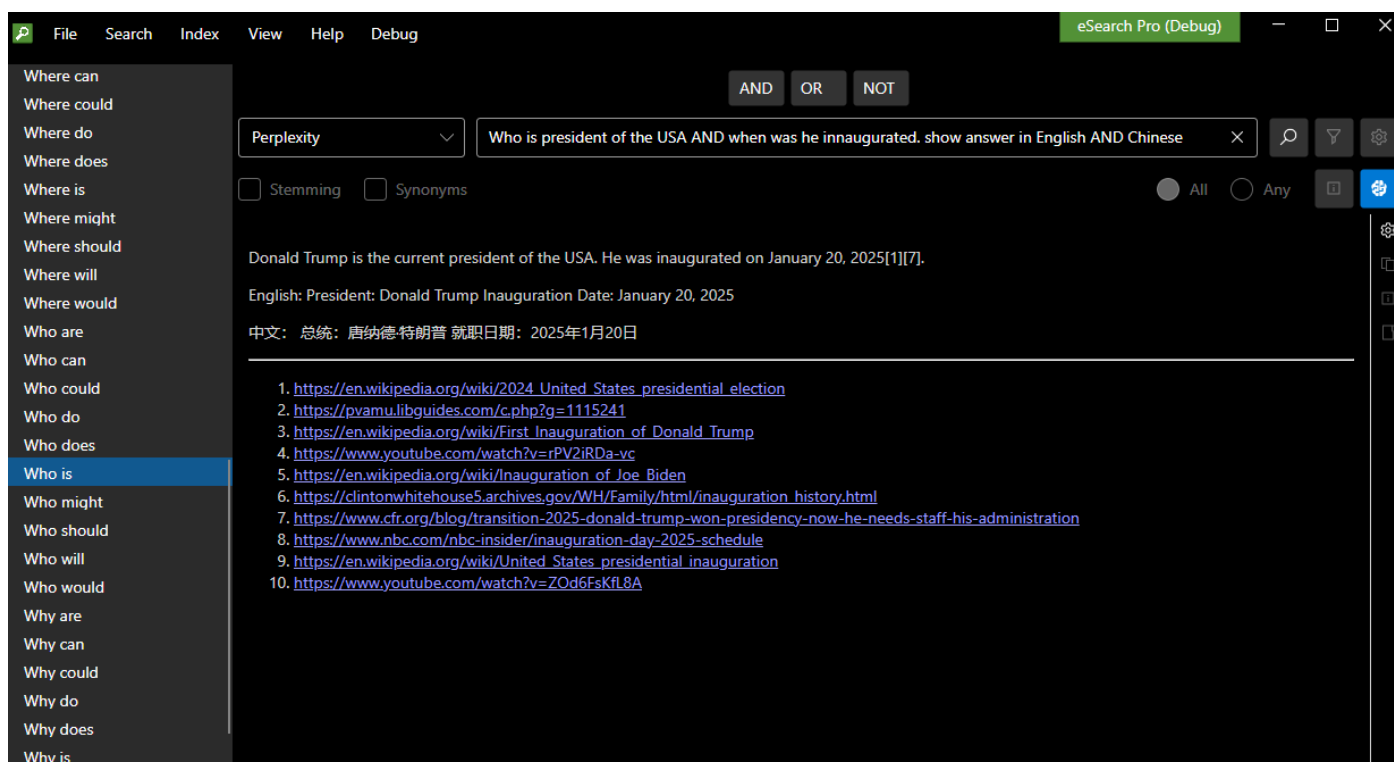
The API will be available so that developers can create new plugins.

AI Online Search

Get up-to-date information from the web without leaving eSearch Pro by using natural language queries instead of keyword searches.

The word list with suggested question stems enhances your productivity. You can add instructions to provide answers in any language, all without having to open a search in your browser or wasting time copying and pasting text into Google Translate or other online translation services.

Perplexity Sonar & Sonar Pro LLMs



Above image (1.0.9162 beta) shows AI mode using Perplexity Sonar Pro model which has a 200K context window. Currently we also include the Sonar model which has a 127K context menu and gives 5 citations and a less detailed response.

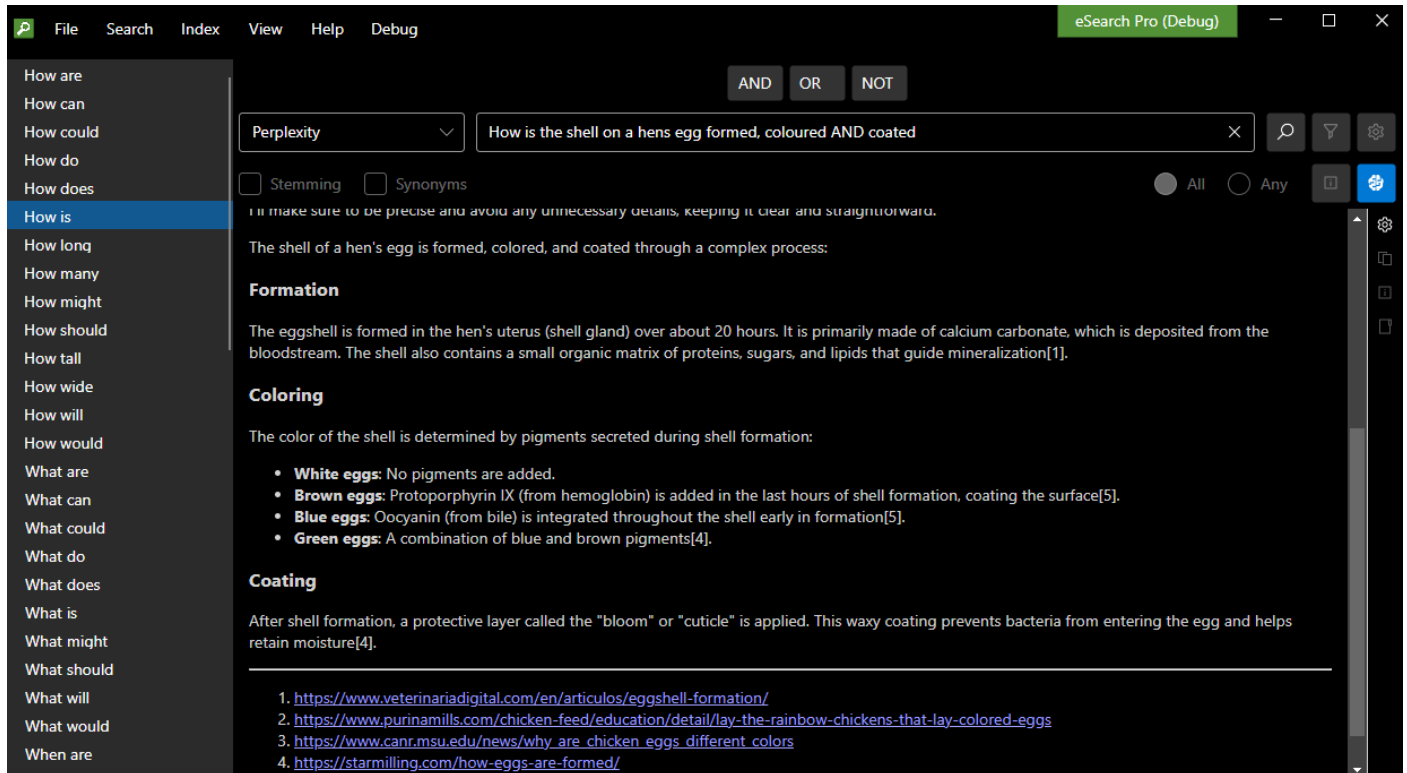
SimpleQA is the prevailing benchmark for factuality, designed to evaluate whether large language models can answer short, fact-seeking questions. Sonar Pro is the best-performing model on factuality because it combines the summarization power of LLMs with access to real-time information rather than relying on stored training data to answer questions. Sonar Pro leads this benchmark with an F-score of 0.858, while Sonar received an F-score of 0.773. <https://sonar.perplexity.ai/>

*Requires a Perplexity.ai API key. This is a pay-as-you-go service, a Perplexity Pro subscription is NOT required.

Perplexity Sonar Reasoning

eSearch Pro AI Mode now includes the latest Sonar Reasoning models, these use the DeepSeek R1 open-source AI model, hosted on servers in the USA.

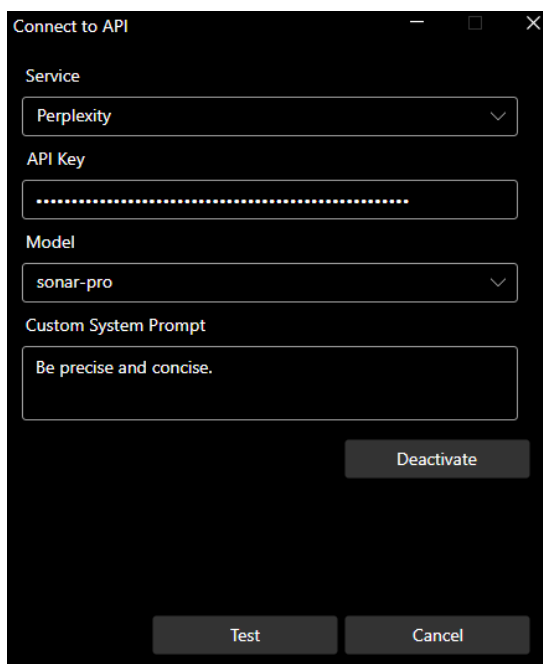
They have a 127k context length and the output includes CoT (chain of thought) steps, as well as links to the web sources used.



Above shows Perplexity Sonar Reasoning Model. (From 1.0.9162 beta.)

Like the Sonar and Sonar Pro models, reasoning models combines a powerful LLM with near real-time web search to give you answers in a few seconds, including an explanation of the steps it took (CoT) to reach its conclusion.

You can carry out research that would normally involve you having to visit each of the websites to extract the information manually. Click on any of the links to the web sources to verify or get further information.

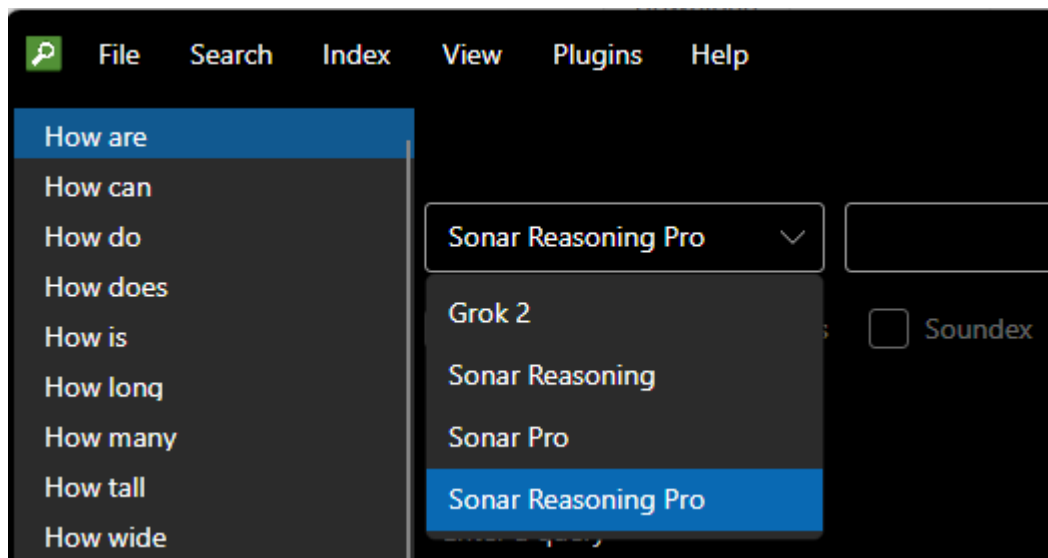


This requires a Perplexity.ai API key. This is a pay-as-you-go service, no 'Pro' subscription is required.

Easy setup. Select the Perplexity service and enter the API-key for each model.

Multiple AI services

Apart from the Perplexity AI web search services, you can connect to multiple AI services of your choice, including OpenAI ChatGPT models and Custom setups.

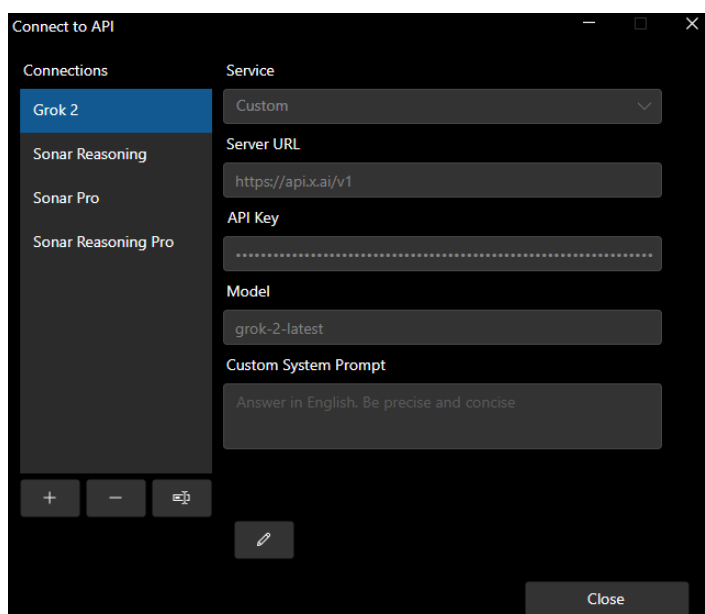


ChatGPT

The OpenAI ChatGPT service has many models available offering a variety of capabilities at different price points. The knowledge cutoff date varies depending on the specific ChatGPT model: GPT-3.5: June 2022 GPT-4: December 2023 GPT-4 Turbo: April 2023. These dates represent the latest information each model was trained on. OpenAI have indicated that these non-reasoning models will be phased out in 2025.

Custom AI

This allows you to connect to another OpenAI Compatible service such as a local LLM or another provider to get the latest models. You can also enter a system prompt to meet your organisations specific search needs.



Setup allows you to rename connections to identify services and models or associate them with projects or cases to control costs.

A security best practice is to rotate API-keys, setup allows you to edit the API-key.

More AI is in the pipeline... We aim to continue to offer the latest online and offline AI capabilities to improve your desktop information retrieval experience and productivity!

30-day trial now available!

- You can download a fully functional trial version. After 30-days it will revert to Lite mode. eSearch Lite has some limitations on functionality, but you can continue to use it for free, for both private or commercial use, for as long as you wish.
- eSearch Pro is a work in progress, check back often for [new releases](#).
- eSearch Pro and Lite versions do not track usage or pass any data to us or any third party. There is error logging, which you can send to us if you need support.
- While in beta testing, both eSearch Pro and Lite are free of charge. You can download new trial versions of eSearch Pro at any time.
- Contact us with your feedback and requirements if you have specific Retrieval-Augmented Generation (RAG) or Agentic AI needs or would like early access to private betas for testing.
- If you have feedback, suggestions, questions, or need a different 'search' problem solved, please contact support@electronart.co.uk

Go to SearchCloudONE.com for free downloads & latest release notes on eSearch Pro!

ElectronArt Design Ltd.
124 City Road
London
EC1V 2NX