



MediCrawl: an ML based system to automate the search of illegal products in ecommerce

11. September 2024

Dr. Nicolás Löffler-Pérez

Schweizerisches Heilmittelinstitut
Institut suisse des produits thérapeutiques
Istituto svizzero per gli agenti terapeutici
Swiss Agency for Therapeutic Products

Hallerstrasse 7, 3012 Bern, Schweiz
www.swissmedic.ch

Swissmedic is the National
authorisation and supervisory
authority for drugs and medical
products

It ensures that medical products are
of **high quality, safe, and effective**



Among our core competencies

- the authorisation of medicinal products
- market monitoring of medicinal products and medical devices
- clinical trials and laboratory testing regarding the quality of medicines



SWISSmedic 4.0



7

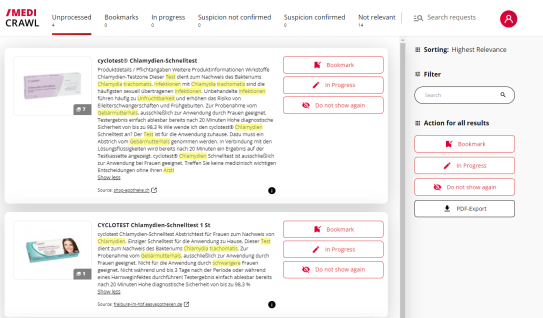


2020 - 2025



Swissmedic 4.0 is an innovation lab, an experimental field for change and innovation. The aim of the initiative is to promote interdisciplinary work and to design new digital business models.

What we did we do?



We are developing a platform to improve market monitoring

Why did we do it?



The market is growing rapidly, and e-commerce is difficult to regulate effectively.

How do we evaluate impact?



User centered approach. It is inspectors who define KPIs for success

Images 2+3: generated through gpt-4o

How does the application look?

The screenshot displays the /MEDI CRAWL application interface. At the top, there is a navigation bar with the following status counts: Unprocessed (4), Bookmarks (0), In progress (0), Suspicion not confirmed (0), Suspicion confirmed (0), and Not relevant (14). A search bar and a user profile icon are also present.

The main content area shows two search results for Chlamydia tests:

- cyclotest® Chlamydien-Schnelltest**: Includes a product image, a description of the test, and three action buttons: Bookmark, In Progress, and Do not show again.
- CYCLOTTEST Chlamydien-Schnelltest 1 St**: Includes a product image, a description of the test, and three action buttons: Bookmark, In Progress, and Do not show again.

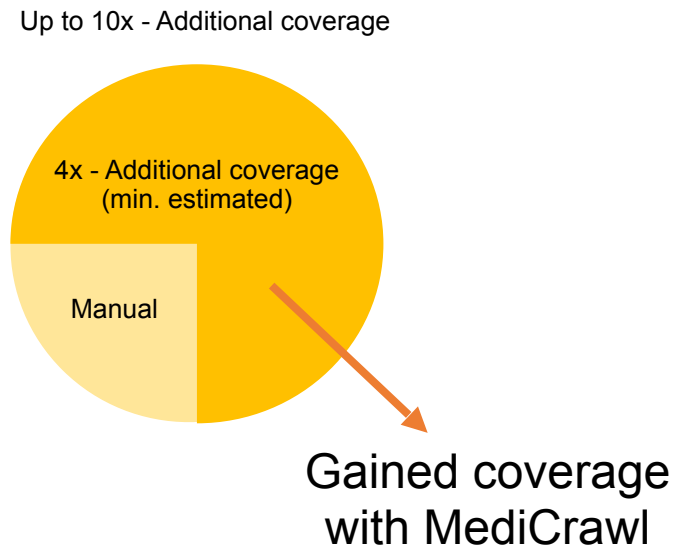
On the right side, there is a sidebar with the following options:

- Sorting: Highest Relevance
- Filter: Search bar
- Action for all results: Bookmark, In Progress, Do not show again, and PDF-Export.

Medicrawl has a clear goal: **to look for illegal products in online markets.**

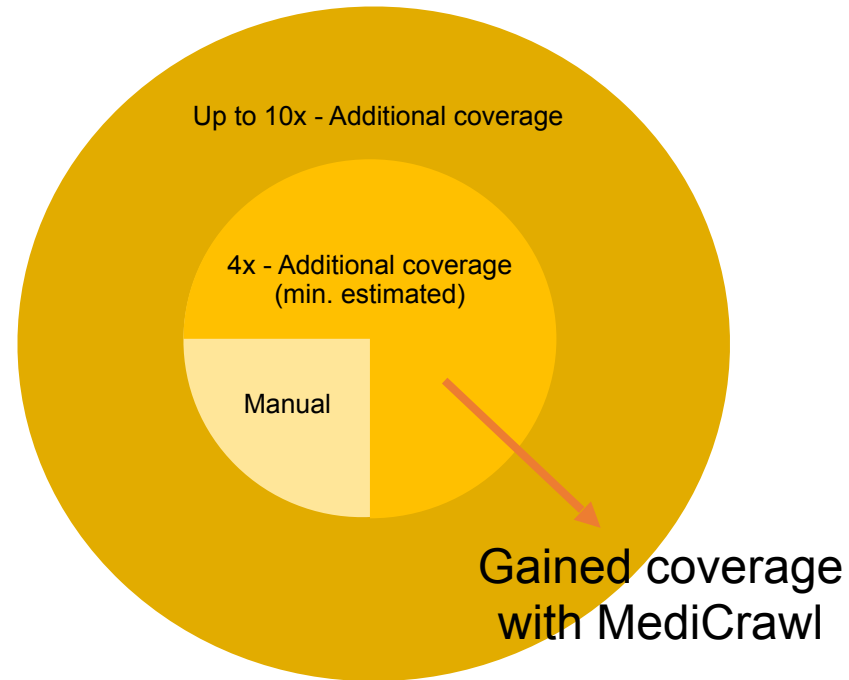
Medicrawl **evaluates whether a product is suspicious,** enabling experts to determine if it violates the law.

Our goal: to expand coverage on online markets



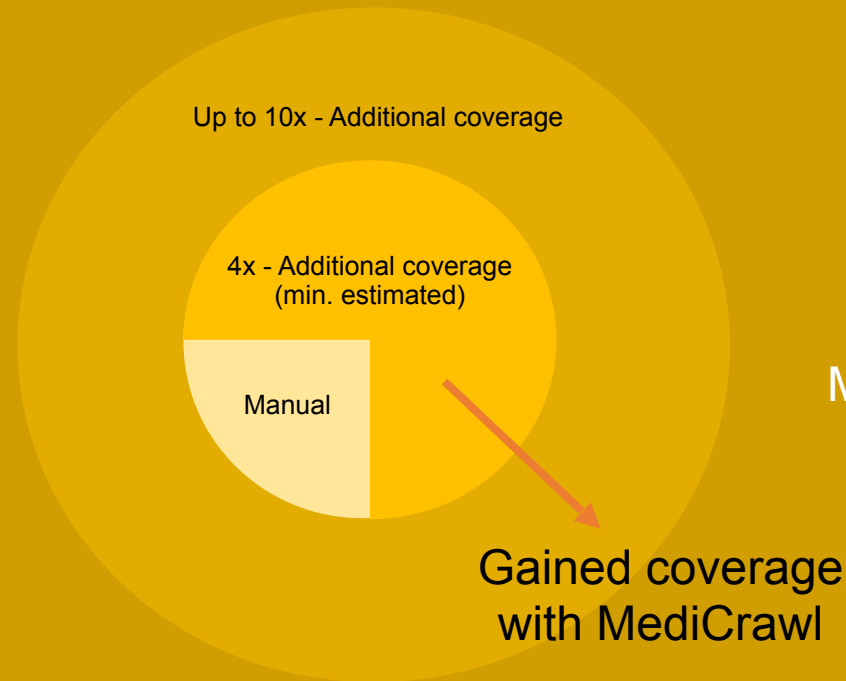
Status 2024: Regular use within Swissmedic to support and complement the work of our colleagues

Our goal: to expand coverage on online markets



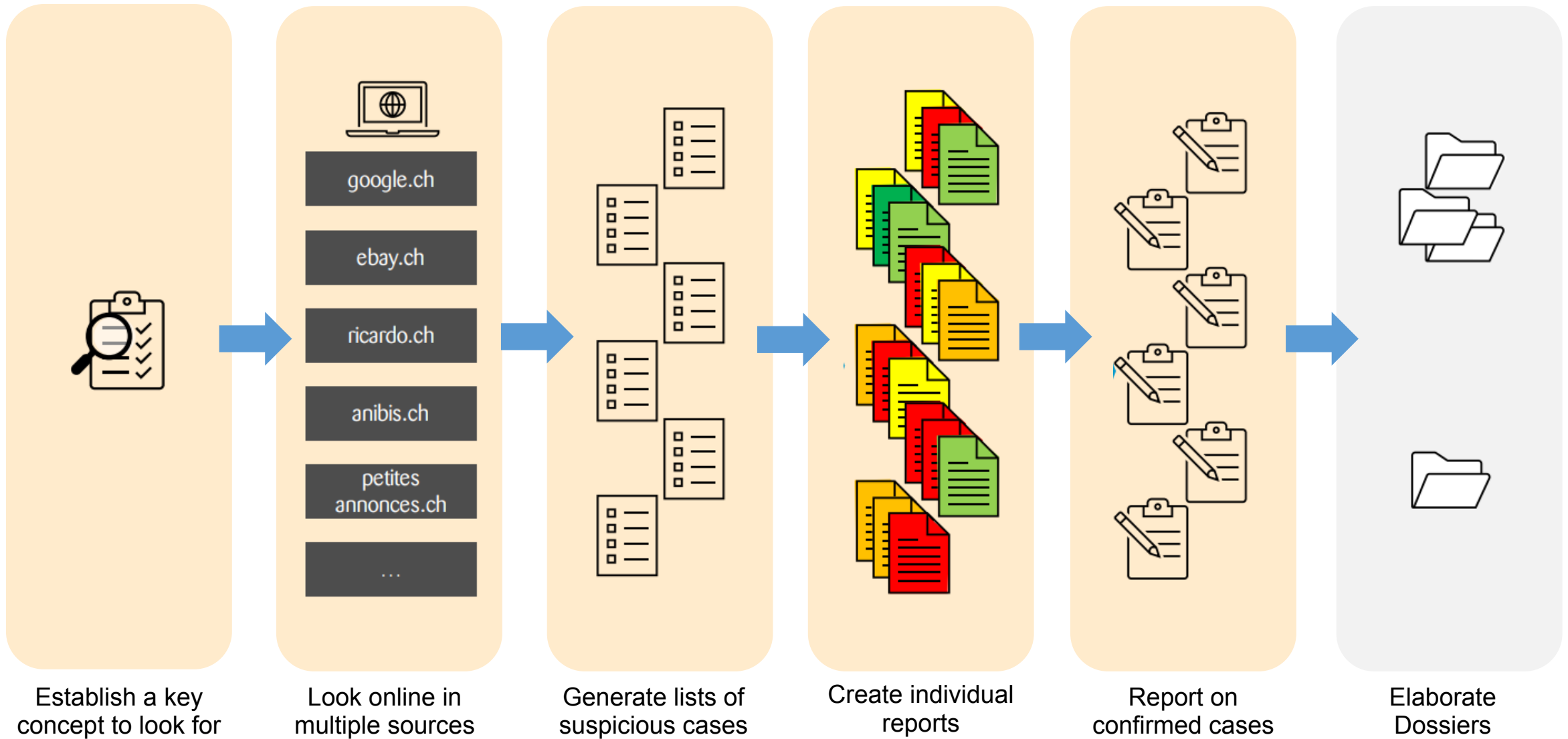
Status 2024: Regular use within Swissmedic to support and complement the work of our colleagues

Our goal: to expand coverage on online markets



The question remains...

How much can we push
MediCrawl to oversee the market?



In daily business, this is the part covered by MediCrawl



google.ch

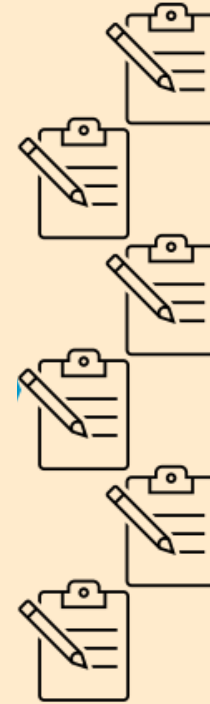
ebay.ch

ricardo.ch

anibis.ch

petites
annonces.ch

...



Establish a key
concept to look for

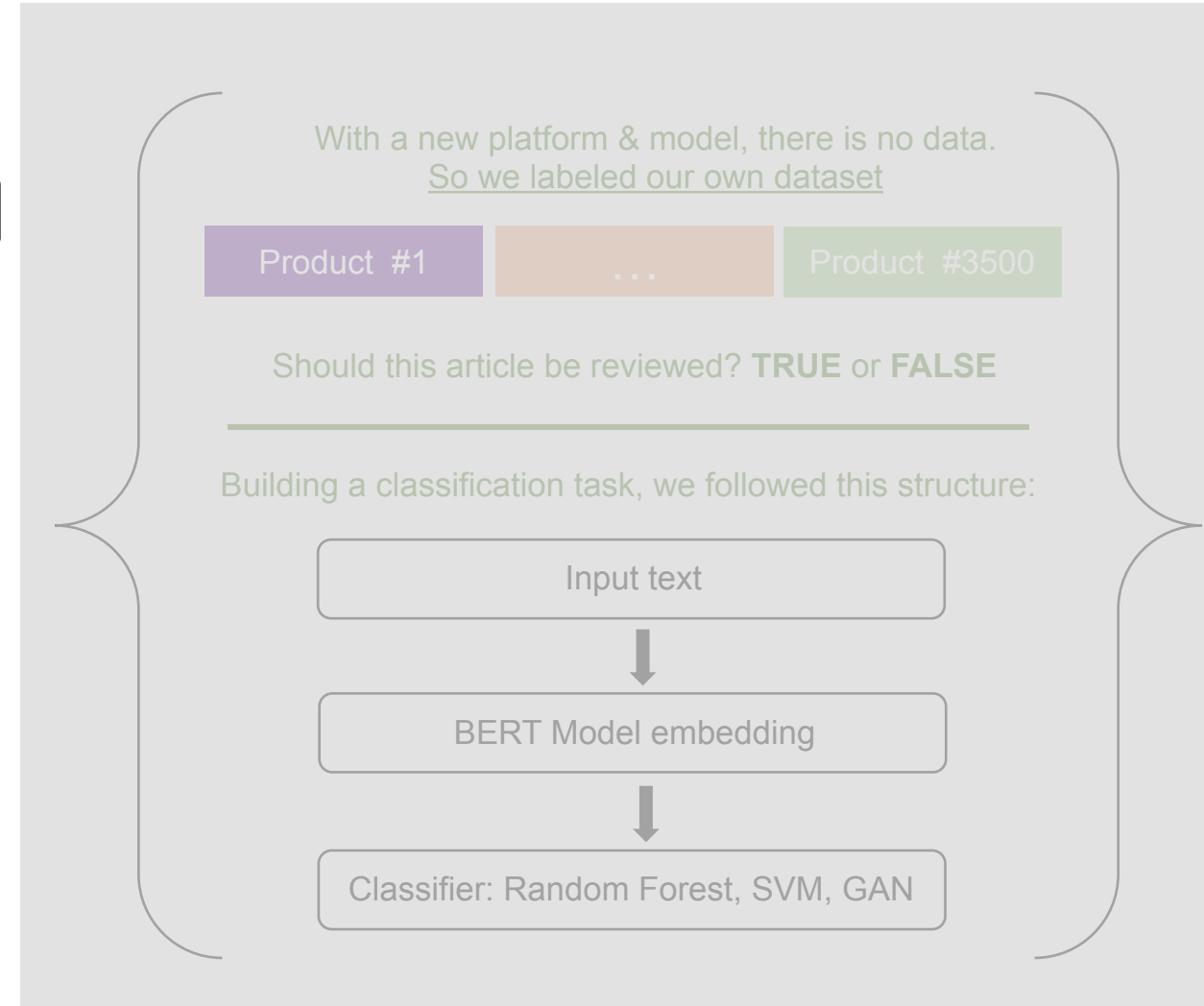
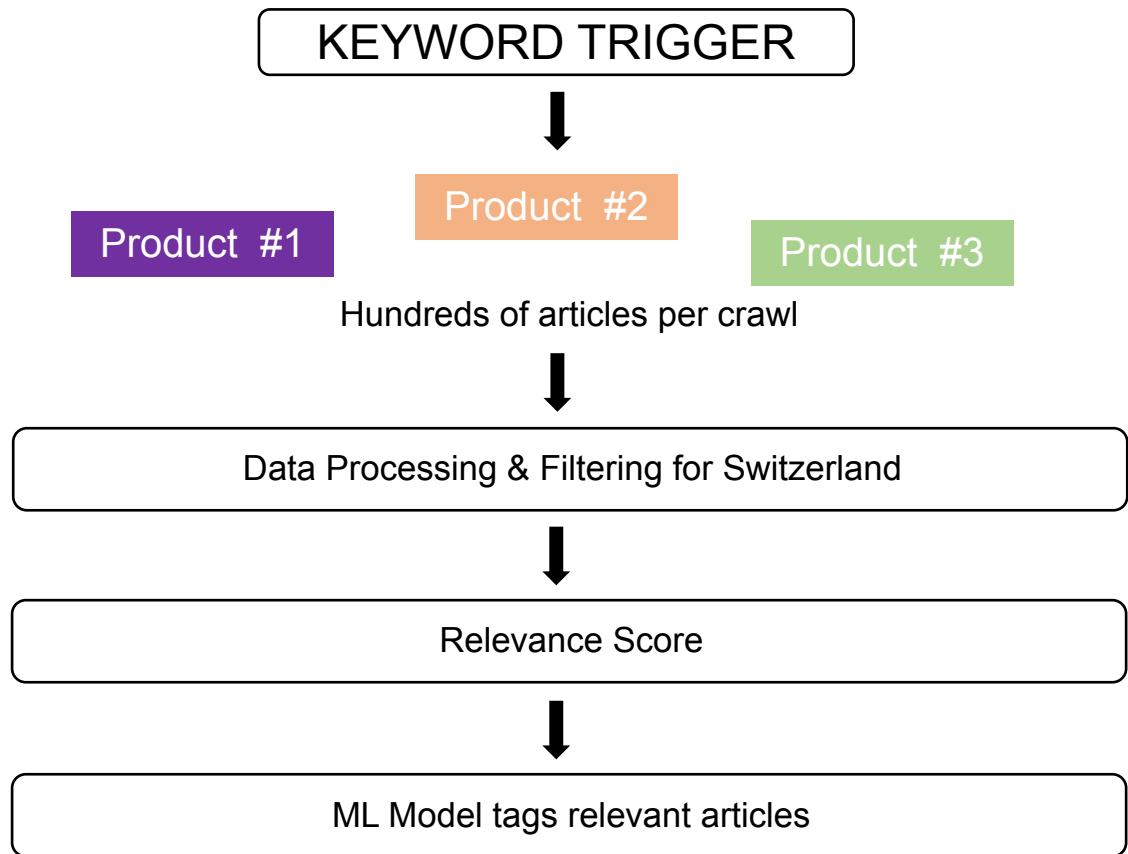
Look online in
multiple sources

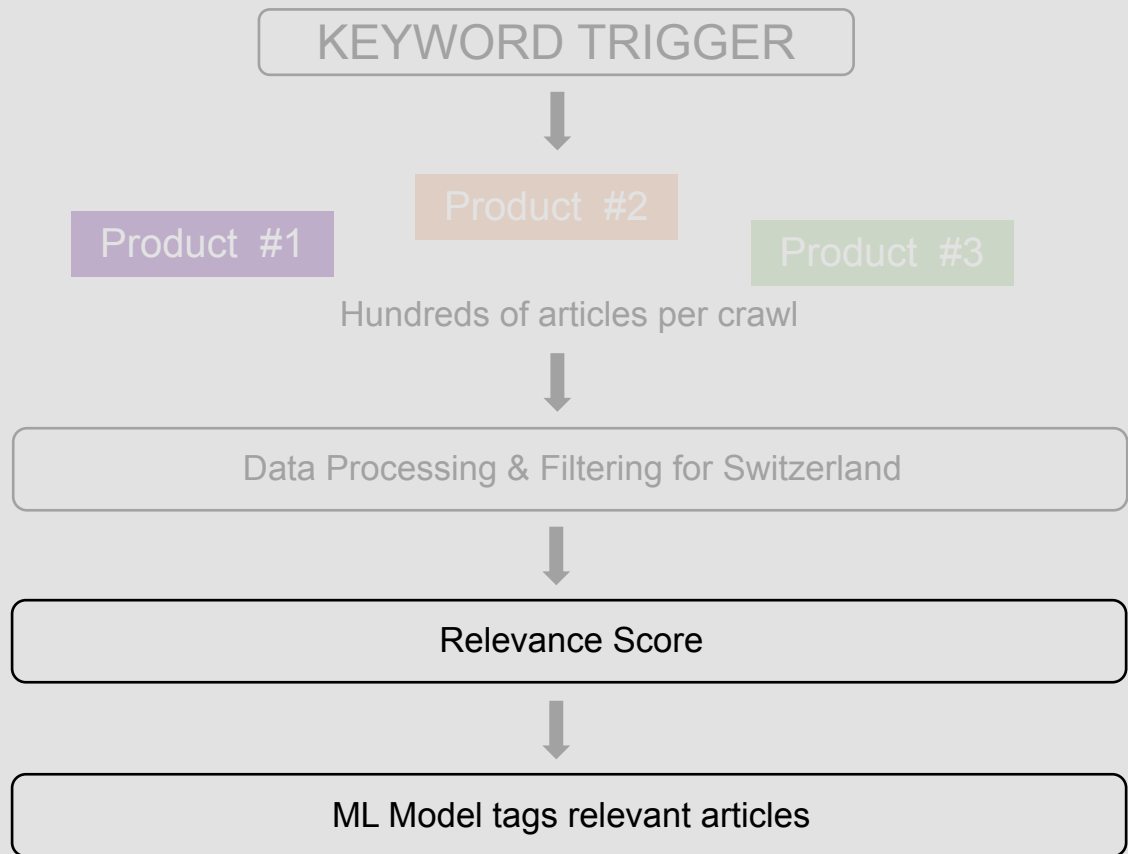
Generate lists of
suspicious cases

Create individual
reports

Report on
confirmed cases

Elaborate
Dossiers





With a new platform & model, there is no data.
So we labeled our own dataset

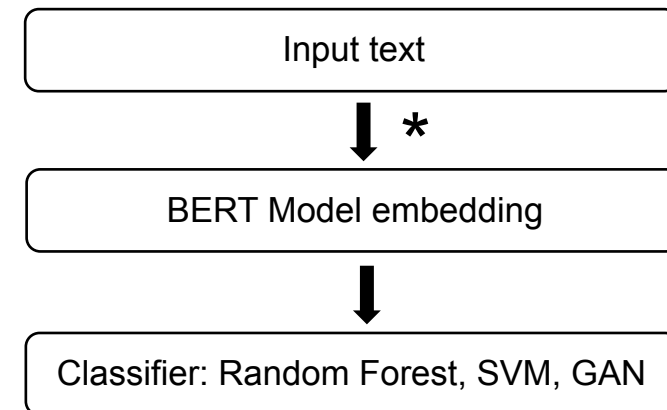
Product #1

...

Product #3500

Should this article be reviewed? **TRUE** or **FALSE**

Building a classification task, we followed this structure:



*lots and lots of processing

How big should the training dataset be?

		Precision	Recall	Total labels used for training
Development phase	2021 Q4	~0.8	~0.95	~1'000
First production deployment	2022 Q1	~0.6	~0.95	~2'000
Separation Models AM/MEP	2023 Q2	AM 0.74 MEP 0.80	AM 0.65 MEP 0.82	~50'000
Retraining of independent models AM/MEP	2023 Q4	No improvements	No improvements	~60'000

Important for experts:
RECALL

How many TRULY relevant articles do we catch?

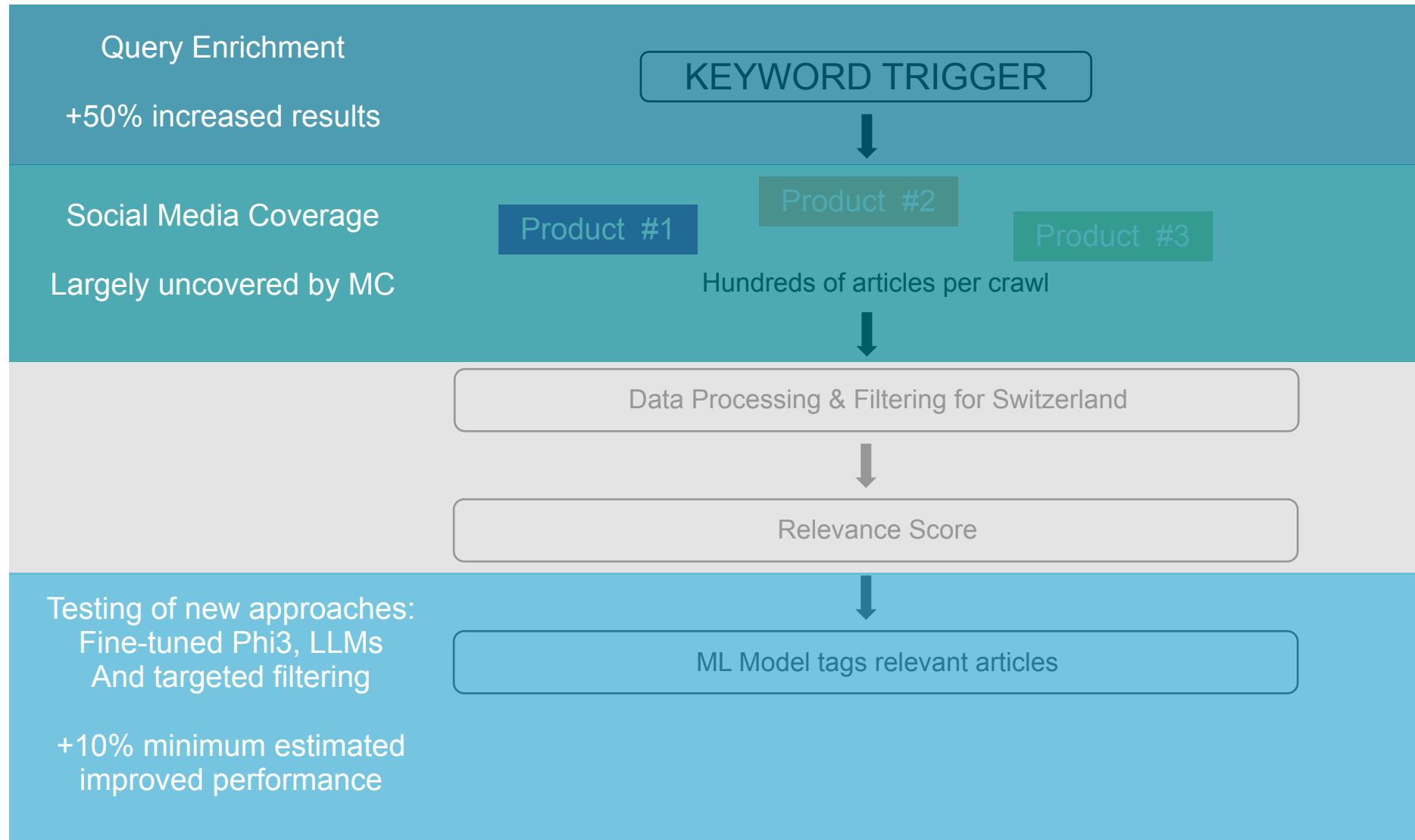
It did not improve, so how can we enhance it?

A valid question is:

BERT has been in the market for years,

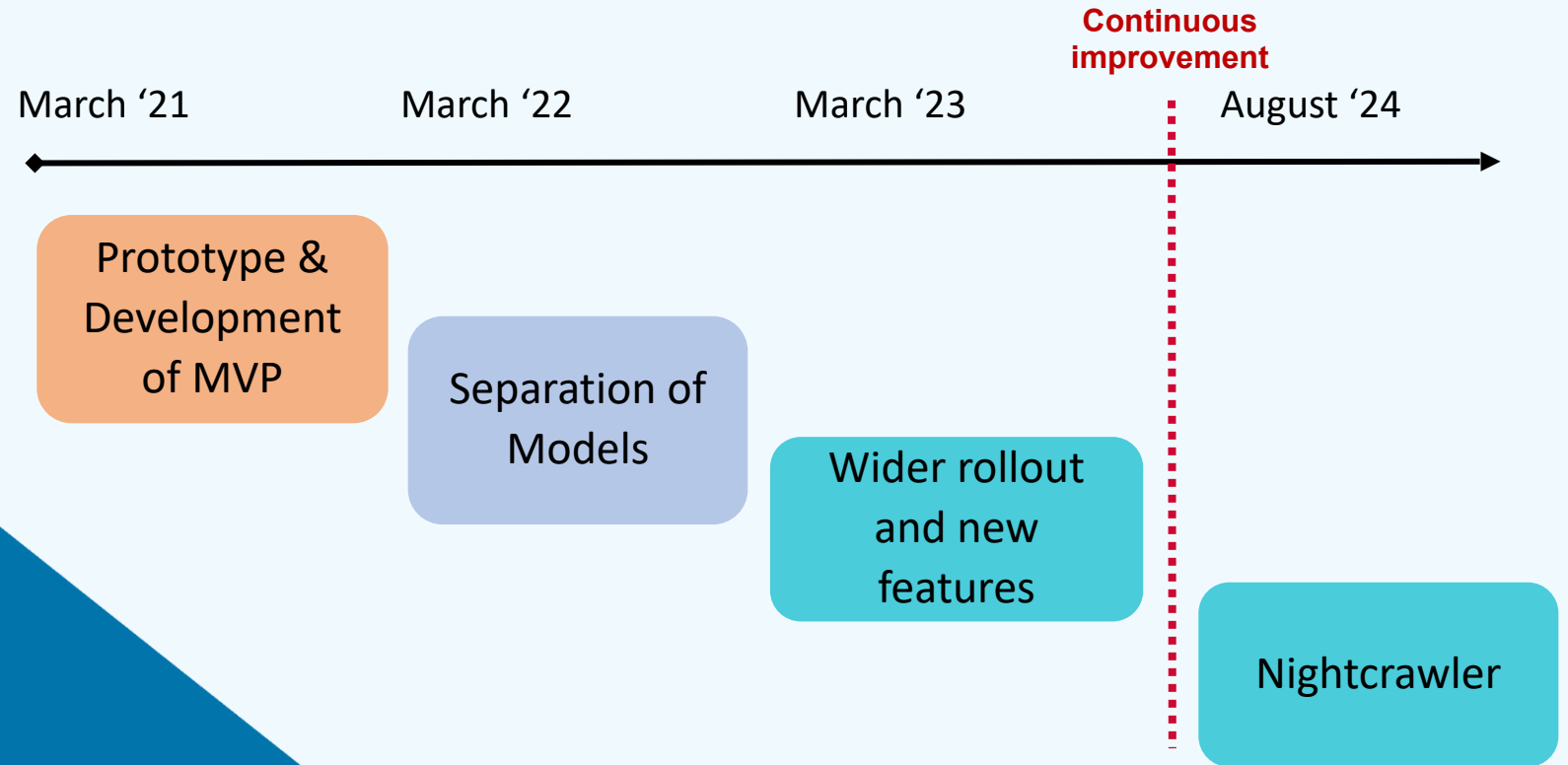
what about LLMs? Or VLMs?

Back to where we were & where we are going



What's the future of Medicrawl?

**/MEDI
CRAWL**



Takeaways

- User centered work beats everything. The more we know the user, the better tools we can develop.
- The product is never finished.
- Trust your data scientists with their choice of models. They choose the best for the given task.

Questions? Cool ideas to share?



Contact me at nicolas.perezgonzalez@swissmedic.ch or find me on 

Recall over precision matters to us

