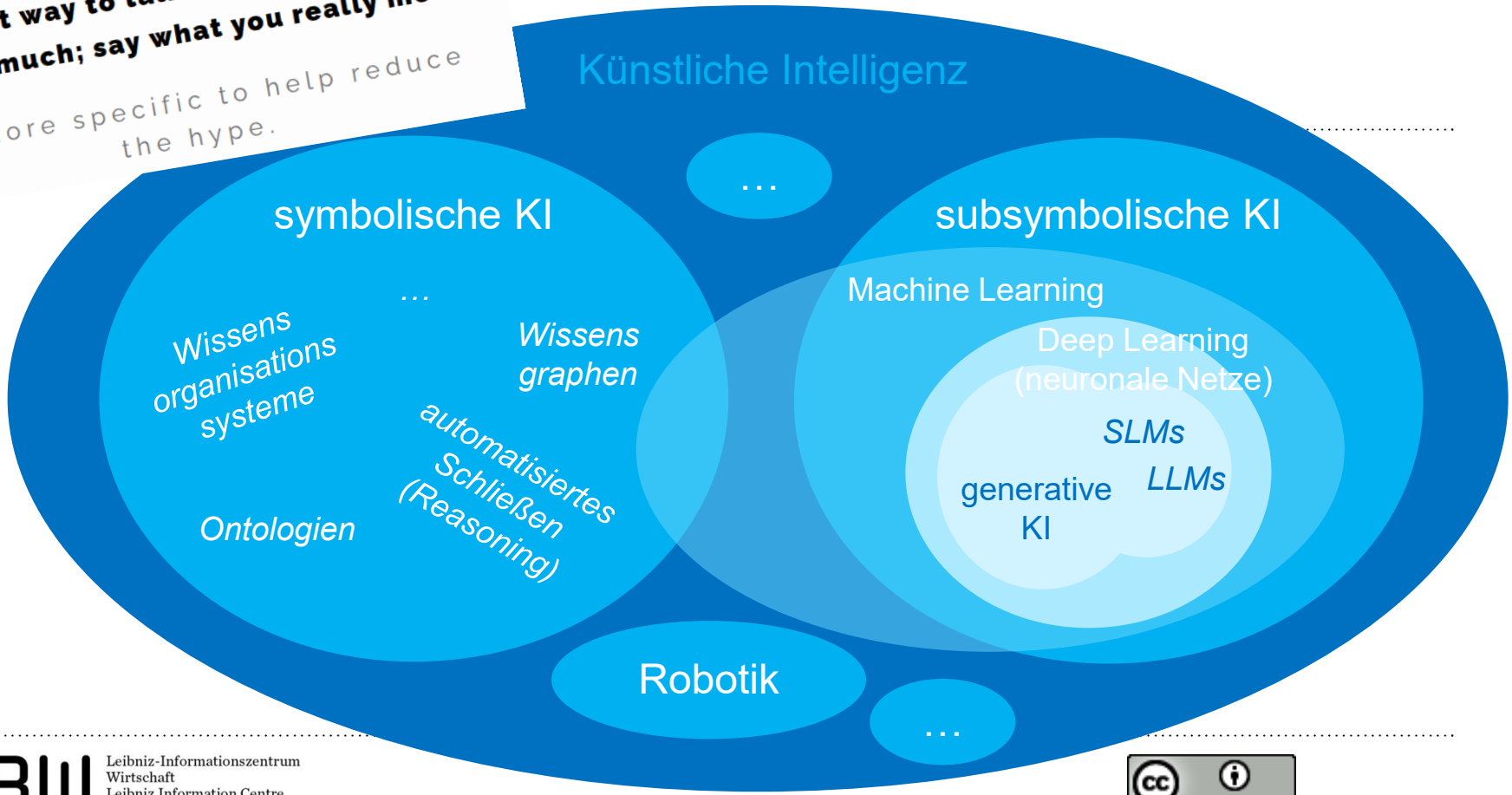
AutoSE: Automating subject indexing at ZBW

Dr. Argie Kasprzik
ZBW – Leibniz Information Centre for Economics
Latvian-German exchange "Digital transformation & AI", online, 15.06.2026

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

The best way to talk about AI: don't say 'AI' so much; say what you really mean

Be more specific to help reduce the hype.

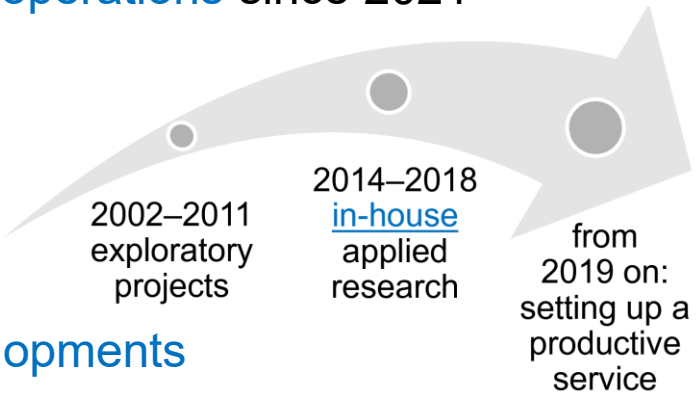


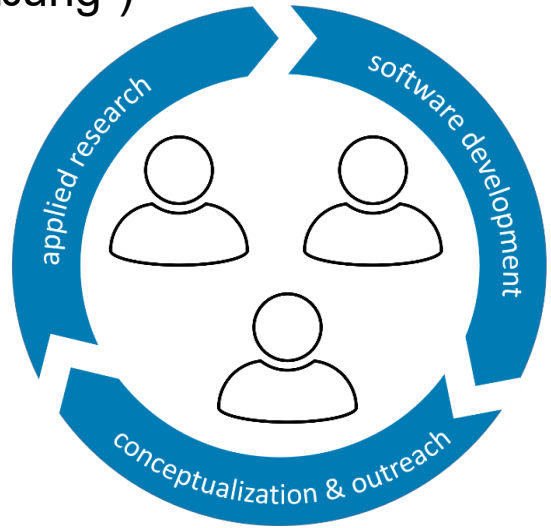
screenshot top left: https://www.bobdc.com/blog/stopsayingai/ ; graphics by A. Kasprzik

AutoSE: automated subject indexing as a productive service

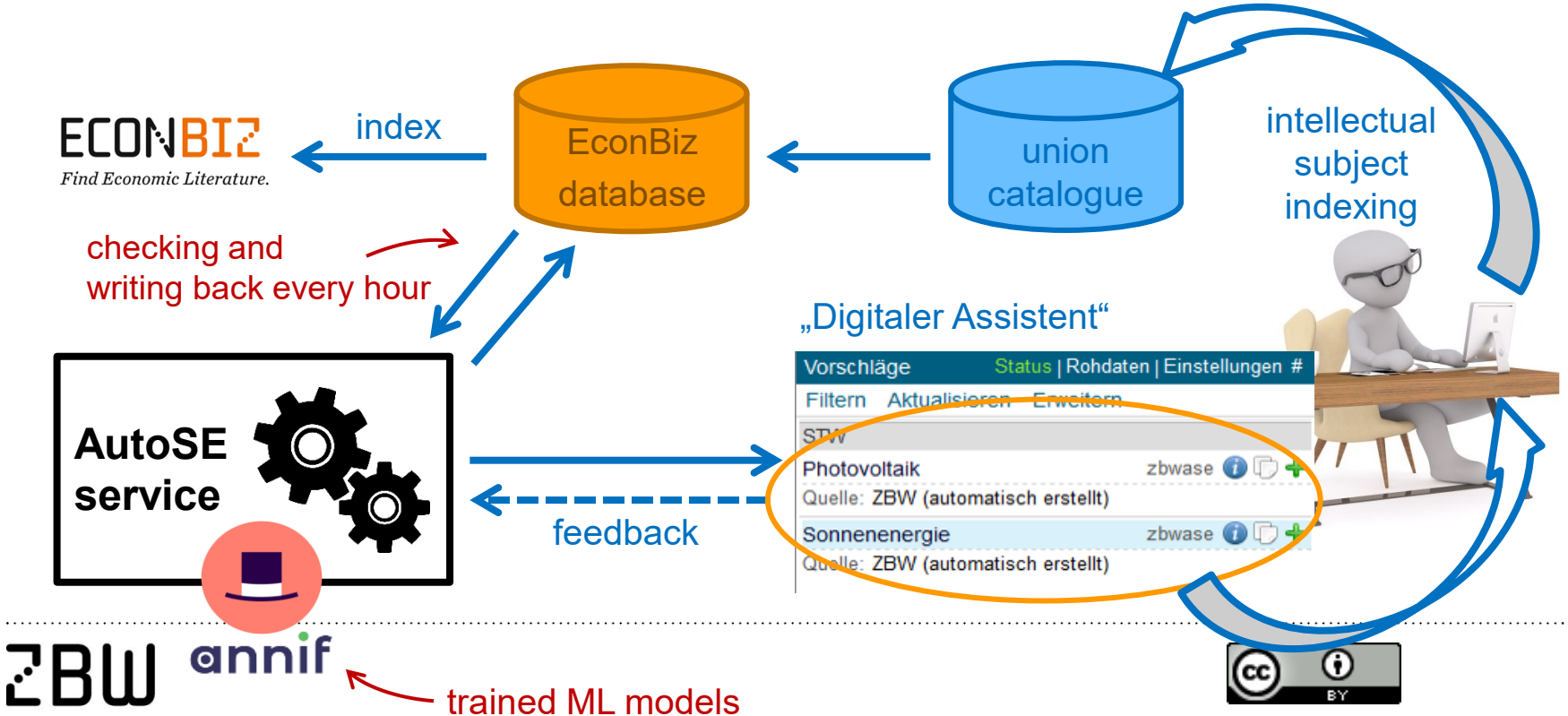
- previously: project AutoIndex (until 2018) – applied research & prototypes
- from 2019: AutoSE („Automatisierung der Sacherschließung“)
- service in productive operations since 2021
- Open Source!

- continuously: evaluating and integrating new developments from our applied in-house AI research



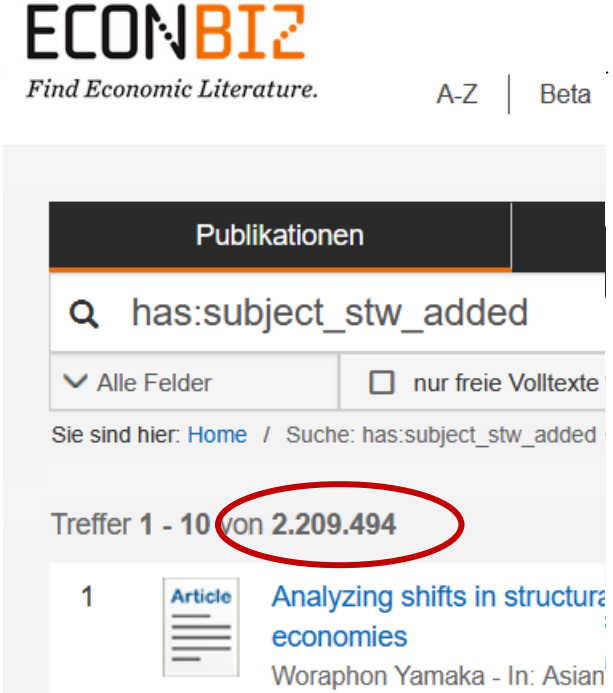


Fully automated and machine-assisted subject indexing

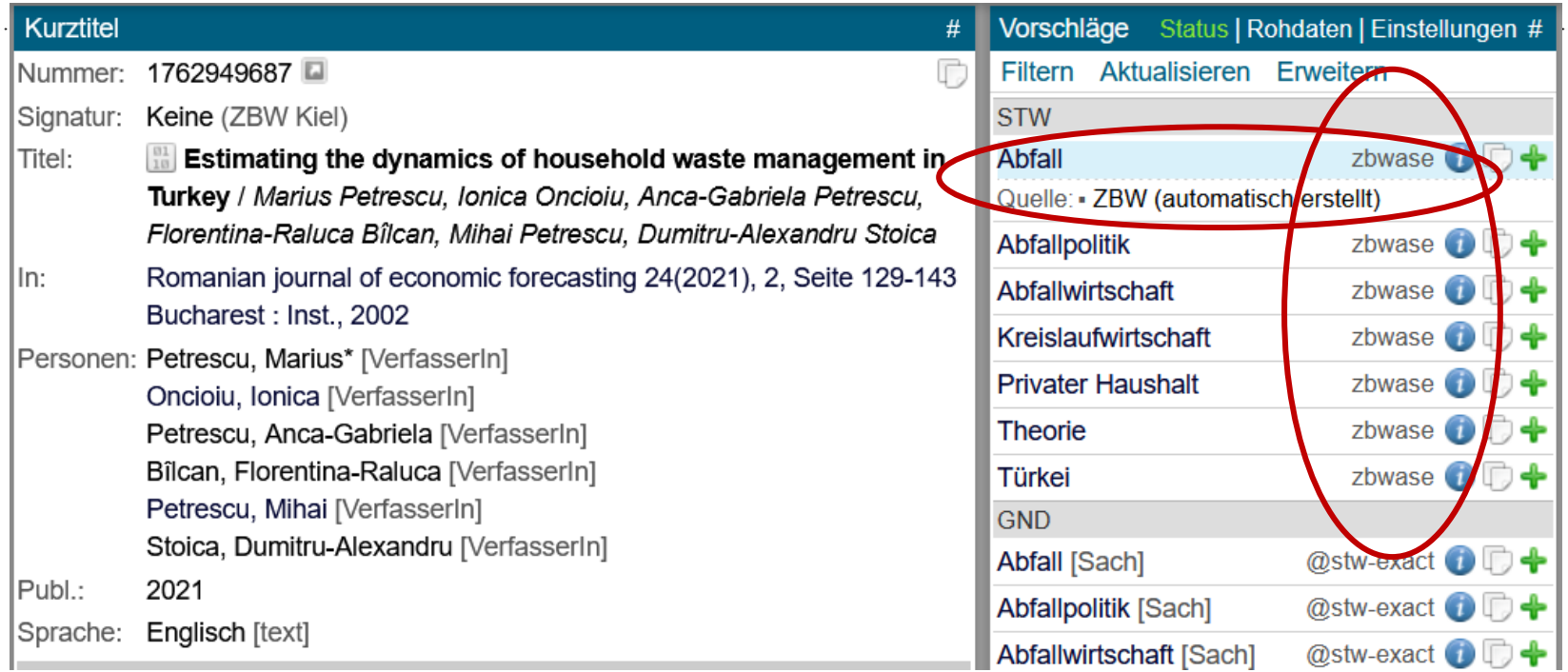


Facts & Figures

- currently for English publications, that is around two thirds of holdings (more languages planned)
- currently titles and author keywords (abstracts: experimental phase)
- May 2026: over 2,2 Mio. subject-indexed metadata records, i.e., a third of holdings and ca. 55% of English-language holdings

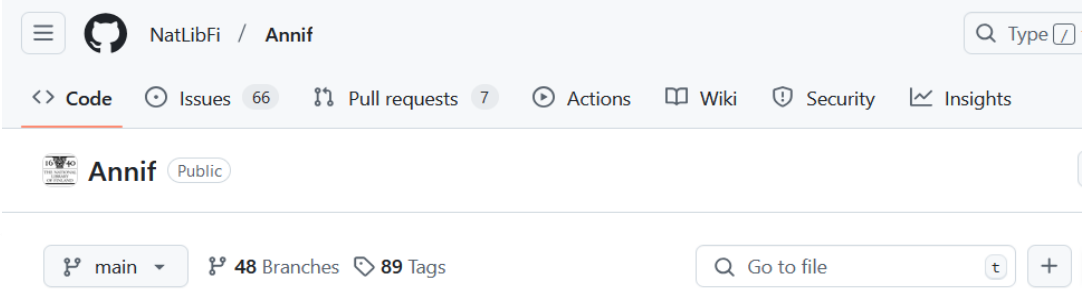


Displaying AutoSE suggestions in the DA-3 platform for machine-assisted intellectual subject indexing



… and collecting valuable feedback from human subject indexers

Open Source community cooperation: Annif



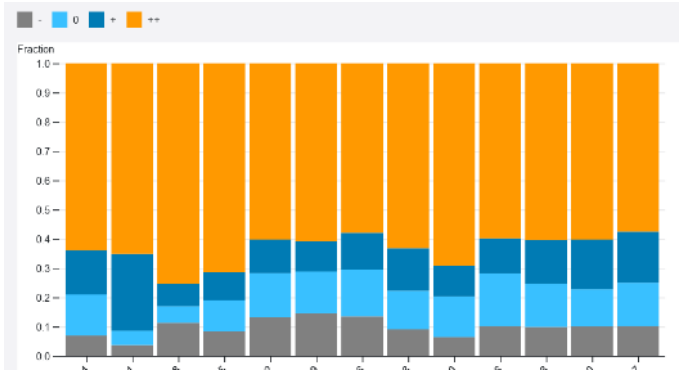
http://annif.org/

- Annif – developed by the National Library of Finland (NLF): a (comparatively) low-threshold toolkit for automated subject indexing, offers various models (backends)
- exchange of ideas between ZBW and NLF, ZBW contributes to open source development of Annif via GitHub and has provided own backend (*stwfsa*)
- AutoSE uses Annif as a core component
- challenge: to check periodically if requirements of productive systems such as AutoSE are still compatible with the "easy to use" ambitions
- challenge: some open source projects are not maintained anymore – requires community effort!

Methods currently used in production

- we combine established machine learning (ML) algorithms incl. a custom model developed at ZBW (stwfsa *) in a so-called *ensemble*
- complemented by a subsequent application of filters and rules
- separate search for optimal parameters
- inhouse development of an ML-based automated quality control (“*qualle*” **)
- overall performance: F1 score ~0.6 (given our training data: decent!)

omikuji
parabel *bonsai*
stwfsa *fastText*

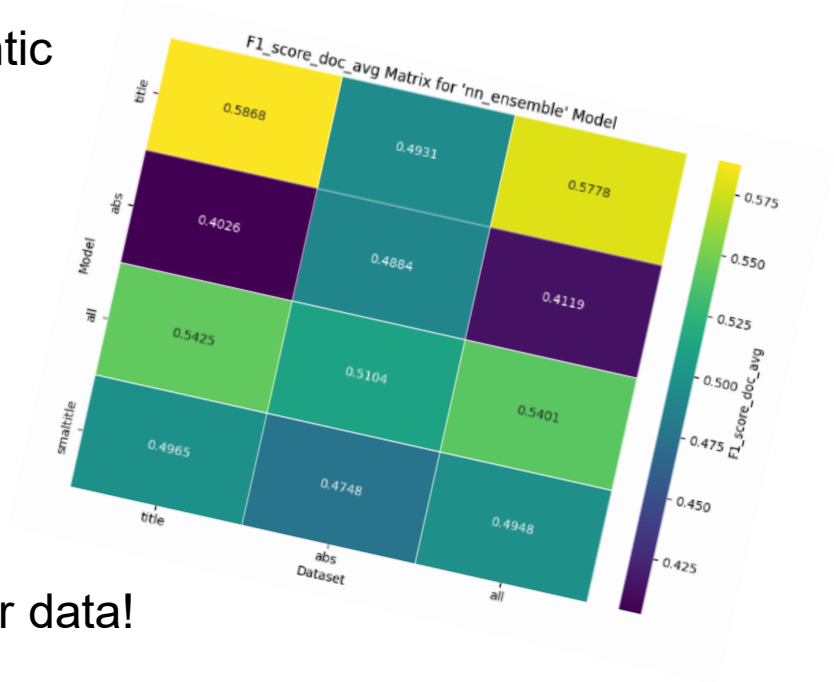


* https://github.com/zbw/stwfsapy

** https://github.com/zbw/qualle

Analyzing interaction of model choice with our data

- assumption: abstracts add valuable semantic information and therefore help in training
- however, less than 10% of our available metadata records have abstracts, and some are in the wrong language
- experimental result: in our case, abstracts in the training data actually hurt model performance! lesson: know your data!



Analyzing interaction with our controlled vocabulary (STW)

Tail Gap

~40%

labels with <100 training examples account for a disproportionate share of all misses

Facet Omission

93.4%

model catches one facet (e.g. "Economics") but entirely misses another (e.g. "Geography")

Specific → General

4.1%

model predicts a broader parent category, failing to pinpoint the precise sub-topic

Silent Failures

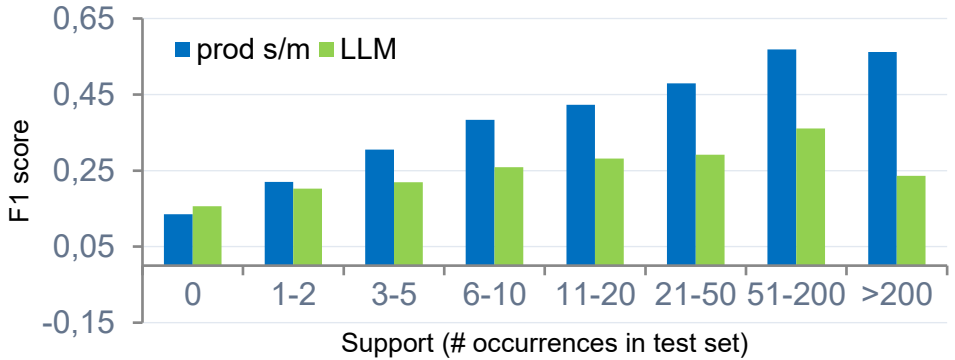
3.1%

model predicts nothing at all for a valid, labeled document

slide courtesy of Lakshmi Rajendram Bashyam (ZBW)

Comparing our productive models with LLMs for subject indexing

question: for the task of generating descriptors from a controlled vocabulary, which works better, our productive machine learning models or a (non-finetuned) open weight LLM?



preliminary results:

- overall our productive model wins
- LLM trades precision for recall (i.e., it takes more "wild guesses"), which sometimes pays off – but only on tail labels

in cooperation with Arben Hajra, ZBW

slide courtesy of Lakshmi Rajendram Bashyam (ZBW)

Current research roadmap for AutoSE

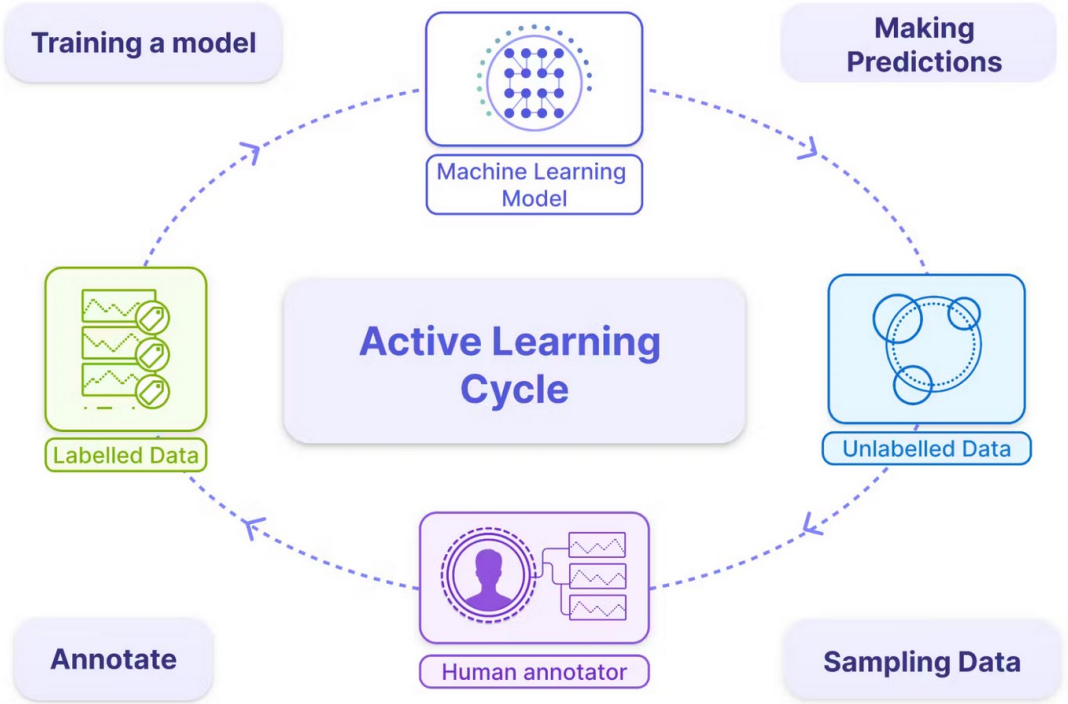
- experimenting with new Annif feature: *subject inclusion/exclusion*, train separate models and combine them (e.g., for geonames)
- particularly for tail labels:
  - experimenting with LLMs
  - evaluating various more recent models for AutoSE, such as X-Transformer and the *embedding*-based EBM (DNB)
  - data augmentation
- active learning

picture from wannapik.com

EBM: https://github.com/NatLibFi/Annif/pull/914
X-Transformer: https://github.com/NatLibFi/Annif/pull/798
https://github.com/NatLibFi/Annif/wiki/Subject-exclusion-and-inclusion

Active Learning

- an additional algorithm decides independently the annotation of which data records the model would benefit most from
- efficient use of intellectual subject indexers as a resource
- in production the system would push these records into a queue or pool and wait for humans to annotate them



https://encord.com/blog/active-learning-machine-learning-guide/

Some lessons learned

- productive operations need long-term commitment by the institution – not a series of tentative projects – **especially** if it involves AI
- for special use cases and data sets such as those in libraries there are no shelf-ready solutions yet (no matter what providers tell you) – personnel with varied expertise needed in-house
- even and especially generative AI needs to be combined with curated, high-quality (meta)data and document bases
- not everything needs an LLM, often (a combination with a) smaller specialized automation solution works just as well or better, and consumes less resources

Thank you!

AutoSE (incl. some more slide decks and publications)**:**
https://www.zbw.eu/en/about-us/knowledge-organisation/automation-of-subject-indexing-using-methods-from-artificial-intelligence

contact: a.kasprzik@zbw.eu