Managing the information explosion: the usefulness of machine learning in SIA – the NIPH example

Tiril Borge

INSIA Meeting May 2023,
Stockholm

Illustrations: Colourbox, DALL-E, CC

NIPH

- Cluster for Reviews and Health Technology Assessments (HTV)
- Our approach: pre-COVID vs during/post COVID
- What is machine learning (ML)
- The ML team at the Norwegian Institute og Public Health (NIPH)
  - Who we are
  - Our mandate
  - What we have done

Cluster for Reviews and Health Technology Assessments (HTV)

- ≈ 50 employees
- Reviews types
  - Health technology assessments (HTAs), single technology assessments
  - Full systematic reviews, rapid reviews, scoping reviews, and evidence and gap maps
- Topics
  - Health and welfare

Our approach pre-COVID

- 35-40 reviews/year
  - Search: 1500-15000
  - Included studies: 5-50
  - Time required for full systematic review: 800-1600 hours

➢ **No machine learning**

During COVID-19

- Number of reviews more than doubled (80 in 2020)
- The ever-increasing number of published studies made current practices unsustainable

➢ **Machine learning (ML) became a necessity**

What is machine learning?

Machine learning is a way for software to learn and make decisions without being explicitly programmed for every situation.

It's like teaching a software to think and solve problems on its own.

What is machine learning?

Instead of giving the software specific instructions, we provide it with data to learn from.

The software analyzes the information and looks for patterns and relationships in the data. Then, it uses these patterns to make predictions or decisions about new data it hasn't seen before.

What is machine learning?

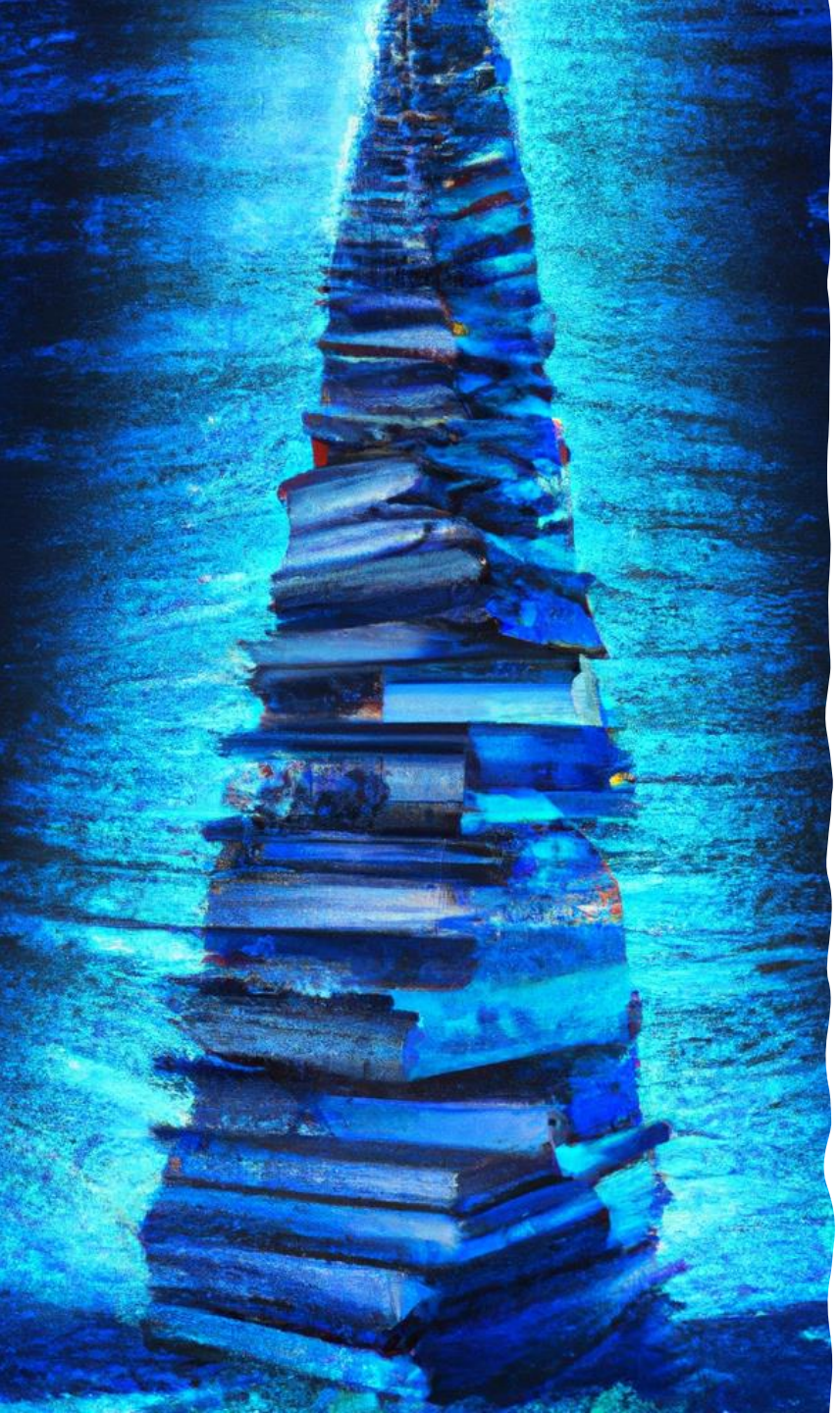
Imagine you want to teach a software to recognize cats. You would show it many pictures of cats and tell it, "These are cats." The software would look at the pictures and learn the features that make them cats, like pointy ears and whiskers.

Once it has learned these features, you can give it a new picture, and it will try to determine if it's a cat or not, based on what it has learned.

What is machine learning?

Machine learning is used in various areas, like recognizing faces in photos, predicting sales trends, or even helping self-driving cars navigate safely.

It's like having a smart assistant that can learn from examples and make decisions based on what it has learned.



As a result of COVID-19

- Experimenting with different ML functions
- Randomized experiment, summer 2020
  - No difference in precision between classifiers and humans
  - ≈ 40% less time used

➢ **Result: the ML team**

➢ Anchored in the NIPH strategy for the 2019- 2024 period concerning automation, workflow processes and methods innovation.

The ML team

- Researchers with varied backgrounds and expertise
- All working on systematic review products/HTAs
- Currently seven members
- **Non-programmers and ML naïve when we started**

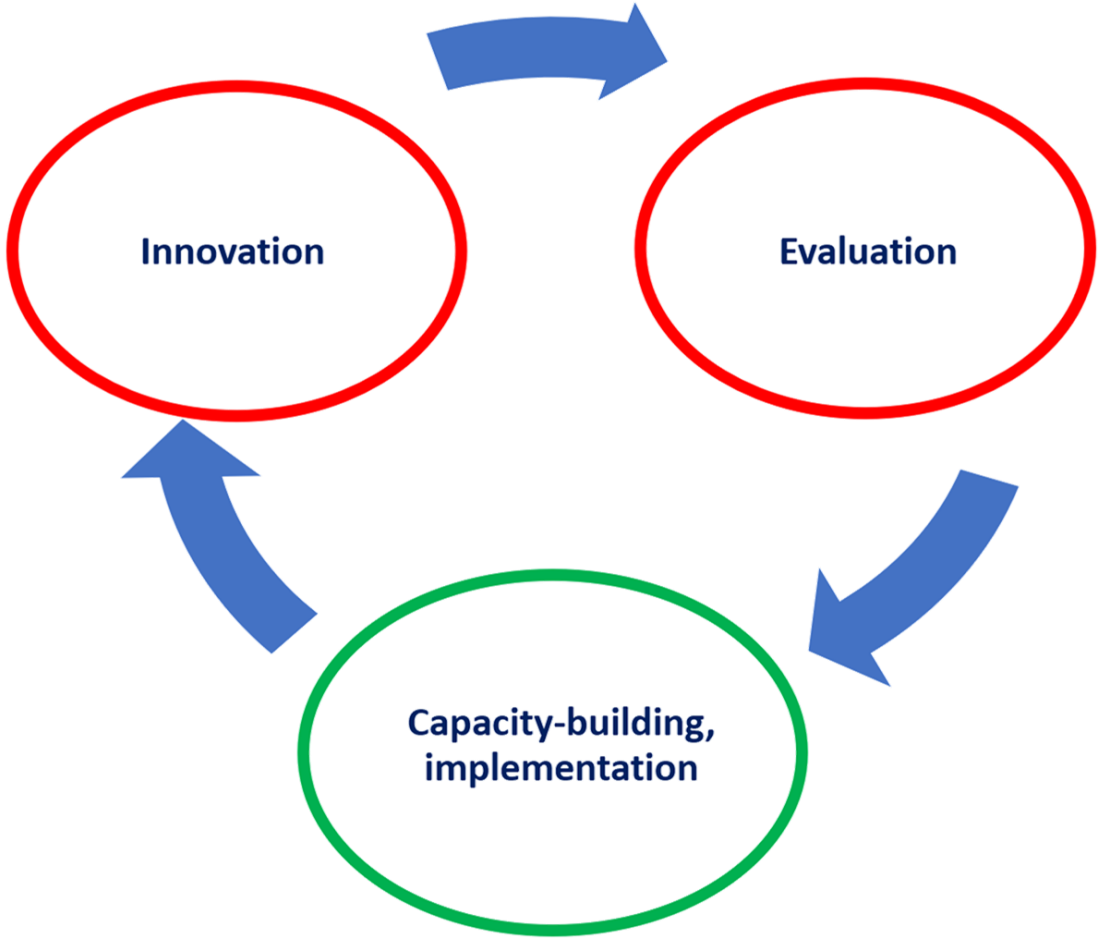
Team 1.0

- **Mandate:** to explore potential benefits of ML in evidence syntheses through piloting specific functions in a small number of projects
- 5 weeks' intensive peer-teaching
- Building our own «evidence base»

NIPH

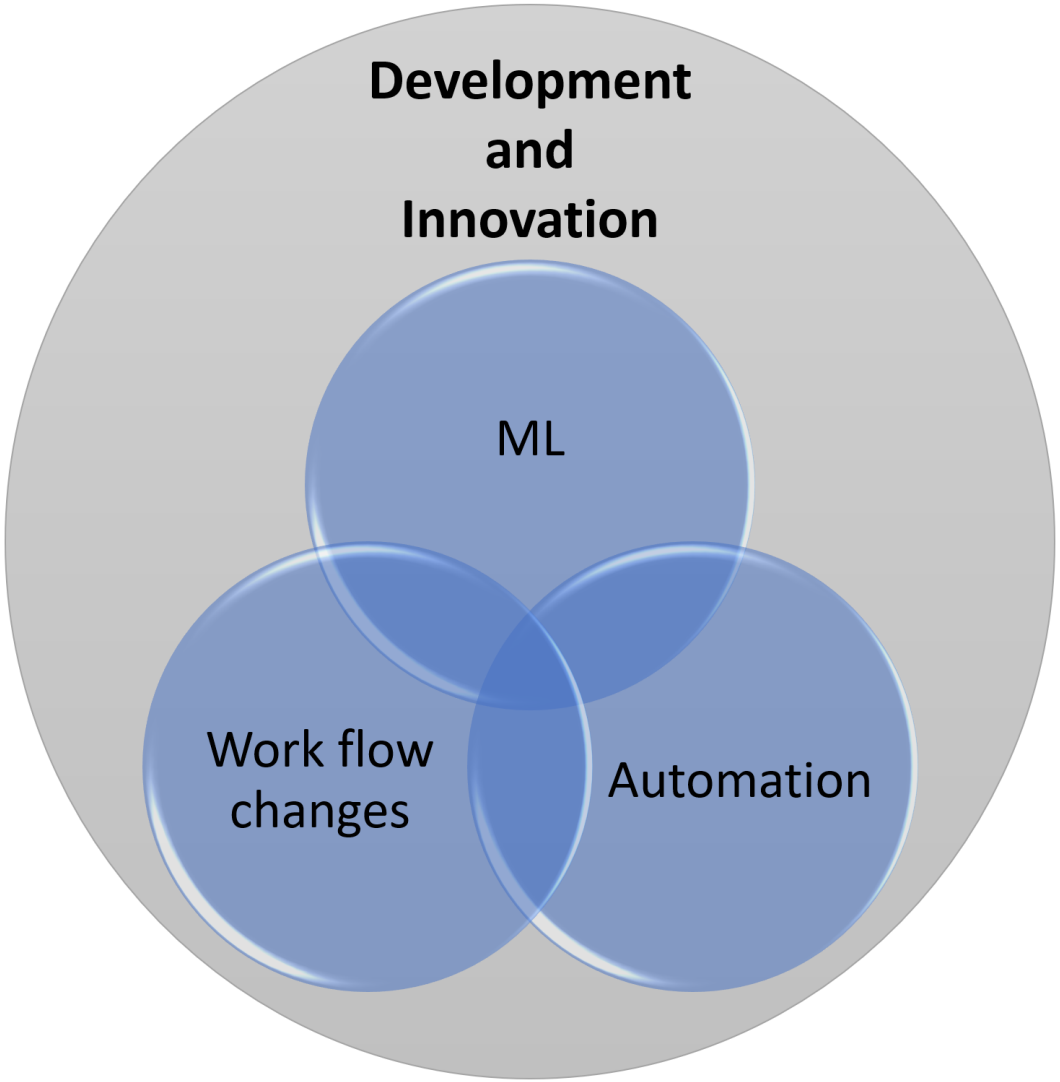
Team 2.0

- Based on the needs identified by Team 1.0
- **Mandate:** to scale up ML through identifying innovative and promising functions, evaluating them, and implementing those that were effective
- Focus on capacity building so that researchers could implement functions independently



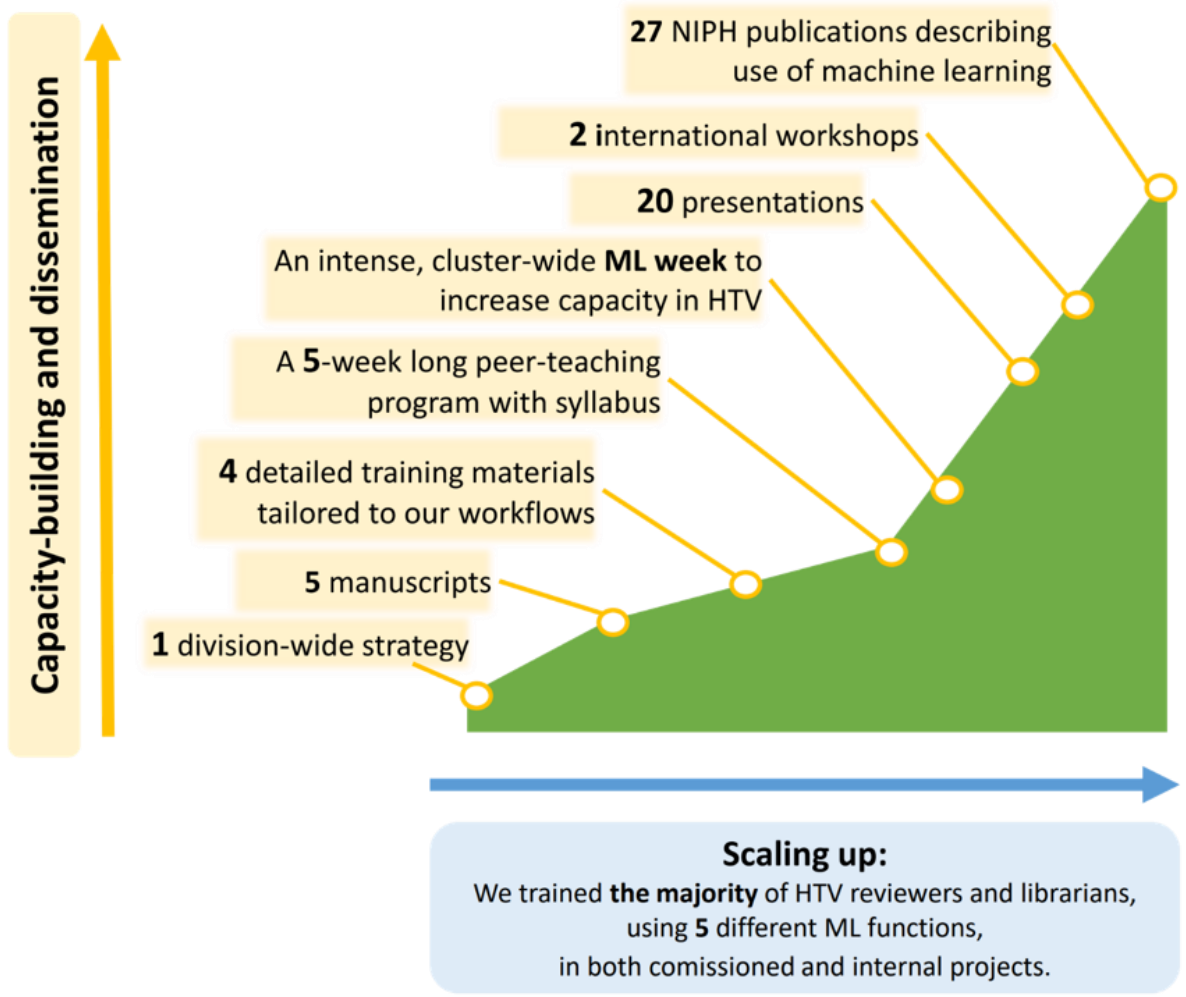
Current team

- Continued focus on implementation and capacity building
- ML is one part of a larger portfolio of activities related to innovation and development in our cluster



What have we found?

- 60-90% time saved in study selection, when using one or multiple ML functions
- 95% of all included studies found after screening 7.5-35% of retrieved studies
- Time savings were greatest when switching to a single screener and/or machine-screening studies after an indication that the majority of relevant studies had been identified



Key team characteristics

- A small, high-performing team that allowed risk-taking
- Curious and independent
- Open to change, new ideas and processes
- Interested in innovation and working more efficiently
- Having protected time in order to have room for thinking, reading, understanding, and being creative

Other facilitators for successful implementation

- Active support from leadership
- Getting early adopters on board
- Providing plain language teaching and guidance of how the ML functions work and how to implement them
- Providing a low threshold for support and questions
- In-house evaluations to «test» if ML functions are acceptable, feasible, and time-saving for our specific workflows

Some challenges

- Training other reviewers to be independent
- Finding out why researchers were hesitant to adopt ML and how to address this hesitancy
- Completely reliant on “off-the-shelf” products, as we are non-programmers
- Having protected time
- Reflexivity

«It is possible to start from nothing and have an impact in a relatively short period of time!»

Current projects – a selection

- **Planning:**
  - Evaluation of Chat GPT to perform RoB assessments
  - Attitudes, barriers and motivations for the use of ML
- **Ongoing:**
  - Development of a scalable training material for the use of machine learning functions in evidence synthesis
  - Evaluation of the effect of machine learning tools for evidence synthesis on resource use and time-to-completion
  - Retrospective evaluation of the use of an RCT classifier for study selection in QES's

Vision: Informal global network

- Institutions or groups that use or want to use ML and automation in evidence synthesis
- Fora where we can keep in contact and more easily:
  - Share ideas
  - Inform about exciting and new developments
  - Avoid duplicate work
  - Ask questions
  - Foster discussions
  - Bring about new collaborations

Results reports for Team 1.0 and 2.0

Please contact us with questions, comments or other enquiries: tibo@fhi.no or machinelearning@fhi.no

Some publications and preprints

- Jardim PSJ, Rose CJ, Ames HM, Meneses-Echavez JF, Van de Velde S, Muller AE. Automating risk of bias assessment in systematic reviews: a real-time mixed methods comparison of human researchers to a machine learning system. *BMC Med Res Methodol*. 2022 Jun 8;22(1):167. https://doi.org/10.1186/s12874-022-01649-y
- Muller AE, Ames HM, Jardim PSJ, Rose CJ. Machine learning in systematic reviews: Comparing automated text clustering with Lingo3G and human researcher categorization in a rapid review. Research Synthesis Methods 2022;13(2):229-41.
- Muller AE, Berg RC., Meneses-Echavez, JF. et al. The effect of machine learning tools for evidence synthesis on resource use and time-to-completion: protocol for a retrospective pilot study. Syst Rev 12, 7 (2023). https://doi.org/10.1186/s13643-023-02171-y
- Muller AE., Ames HM, Borge TC, Hestevik C, Meneses-Echavez JF, and Rose CJ. "A protocol to evaluate unsupervised text clustering to screen and categorize studies in systematic reviews." https://www.researchsquare.com/article/rs-1644531/v1
- Borge TC and Muller AE (in press). " Overdosevarslingssystemer – en kartlegging-oversikt med maskinlæring." Nordic Studies on Alcohol and Drugs