REPORT

2024

Implementing machine learning in an evidence synthesis group:

Recommendations based on a three-year implementation process

NIPH
Norwegian Institute of Public Health

Published by	Norwegian Institute of Public Health Division for Health Services
Norwegian title	Implementering av maskinlæring i en kunnskapsoppsummeringsgruppe: Anbefalinger basert på en treårig implementeringsprosess
English title	Implementing machine learning in an evidence synthesis group: Recommendations based on a three-year implementation process
Responsible	Guri Rørtveit, director
Authors	Heather Ames, Project Leader and Senior Researcher Tiril Cecilie Borge, Researcher Hans Bugge Bergsund, Senior Advisor Patricia Sofia Jacobsen Jardim, Senior Advisor
ISBN	978-82-8406-450-5
Publication type	Report
Number of pages	[42] (71 including appendices)
Comissioner	Cluster for Reviews and Health Technology Assessments
MeSH terms	Machine learning, implementation research, capacity building
Citation	Ames HMR, Borge TC, Bergsund HB, Jardim PSJ. Implementing machine learning in an evidence synthesis group: Recommendations based on a three-year implementation process [Implementering av maskinlæring i en kunnskapsoppsummeringsgruppe: Anbefalinger basert på en treårig implementeringsprosess.] –2024. Oslo: Norwegian Institute of Public Health, 2024.

Content

Hovedbudskap

Å utarbeide en kunnskapsoppsummering er en arbeids- og tidkrevende prosess, men bruk av maskinlæring (ML) kan bidra til å effektivisere denne prosessen uten å gå på bekostning av kvaliteten. Derfor etablerte Klynge for vurdering av tiltak (HTV) ved Folkehelseinstituttet (FHI) i 2020 et dedikert ML-team for å implementere ML i utarbeidelsen av kunnskapsoppsummeringer i HTV. Hovedmålet var å forbedre praksisen for utarbeidelse av kunnskapsoppsummeringer ved å kombinere menneskelig intelligens med ML for å optimalisere arbeidsflyten gjennom hele kunnskapsoppsummeringsprosessen.

Denne rapporten gir anbefalinger om hvordan man gjennomfører implementeringen av ML funksjoner i kunnskapsoppsummeringsprosesser i et ML-naivt arbeidsmiljø, basert på erfaringene med å implementere ML ved HTV. Den tilbyr forslag til beste praksis, forankret i våre refleksjoner rundt ML-implementeringen, med mål om å hjelpe andre ML-naive grupper eller institusjoner med å implementere ML-funksjoner i kunnskapsoppsummeringsprosesser. Veiledningen kan tilpasses ulike organisatoriske mål og målsettinger, samt er anvendelig for implementering av ulike ML-verktøy og -funksjoner.

Rapporten er strukturert i tre hoveddeler som samsvarer med ulike faser av implementeringen: preimplementering, implementering og opprettholdelse/evaluering. Vi bruker EPIS-rammeverket gjennom hele dokumentet som et verktøy for å forklare de ulike implementeringsfasene og viktige aspekter å vurdere i hver fase. Hver seksjon avsluttes med noen hovedpoeng basert på våre implementeringserfaringer, oppsummert som praktiske tips om viktige aspekter som vi anser som viktige å vurdere i implementeringen.

Tittel:
Implementering av maskinlæring i en kunnskapsoppsummeringsgruppe: Anbefalinger basert på en treårig implementeringsprosess
----
Hvem står bak denne publikasjonen?
Klynge for vurdering av tiltak, Område for helsetjenester, Folkehelseinstituttet
----
Godkjent av:
Rigmor Berg, avdelingsdirektør HTV1

Key messages

The evidence synthesis process is a labour- and resource intensive process but using machine learning (ML) is one way to expedite the evidence synthesis process without compromising quality. Therefore, in 2020, the Cluster for Reviews and Health Technology Assessments (HTV) at the Norwegian Institute of Public Health (NIPH) established a dedicated ML team to implement ML in evidence synthesis processes in HTV. The main aim was to enhance evidence synthesis practices by combining human intelligence with ML to optimize workflow changes throughout the evidence synthesis process.

This report provides recommendations on how to carry out implementation of ML functions in ML-naïve evidence synthesis groups, based on the experiences implementing ML at HTV. It offers "best practice" suggestions, rooted in our reflections on implementation, aiming to assist other ML naïve groups or institutions in implementing ML functions in the evidence synthesis process. The guide is adaptable to different organizational goals and objectives, while providing insights applicable to implementation of various ML tools and functions.

The report is structured into three main sections corresponding to different phases of implementation: pre-implementation, implementation, and sustainment/evaluation. We use the EPIS framework throughout the document as a tool to explain the different implementation phases and important aspects to consider in each phase. Each section concludes with a "Take home message" based on our implementation experiences, summarized as practical tips on important aspects that we believe are important to consider in the implementation process.

Title:
Implementing machine learning in an evidence synthesis group: Recommendations based on a three-year implementation process.

Publisher:
Cluster for Reviews and Health Technology Assessments, Division for Health Services, The Norwegian Institute of Public Health

----

Approved by:
Rigmor Berg, department director HTV1

Preface

In 2020, the Cluster for Reviews and Health Technology Assessments (HTV) at the Norwegian Institute of Public Health (NIPH) established a dedicated machine learning (ML) team to implement ML in the evidence synthesis products produced in HTV. Since its inception in late 2020, the ML team has positioned NIPH as a leader in implementing ML into evidence synthesis. Our experiences, best practices and lessons learnt have culminated in this implementation guidance document, which is aimed at being a resource for other institutes or groups that want to implement change involving implementation of ML in the evidence synthesis process.

Financing
The ML work from initiation in 2020 up until the end of 2022 was self-initiated and financed by HTV, Division for Health Services at NIPH. Much of the work during 2023, particularly relating to implementation activities, was externally funded, with the remaining work being financed by HTV.

Conflicts of interest
All authors declare they have no conflicts of interest.

With appreciation
The ML team's implementation efforts are due not only to the dedication of its members, past and present, but also to HTV leadership's investment and vocal support. There have also been numerous colleagues who have provided support, feedback, ideas, and opportunities including the HTV team of librarians. Outside of NIPH, James Thomas' mentoring and his team at EPPI Centre have continued to be instrumental to our understanding of ML and its potential to provide the most valuable evidence synthesis products to our commissioners.

We hope that institutions or groups embarking on the implementation of ML in the evidence synthesis process will find this guidance document a valuable resource to their implementation efforts.

Kåre Birger Hagen	Rigmor C Berg	Tiril C Borge	Heather Ames
Research director	*Department director*	*Team leader*	*Implementation leader*

Background for implementation guidance development

Since early 2020, the Cluster for Reviews and Health Technology Assessments (HTV) at the Norwegian Institute of Public Health (NIPH) recognized the potential benefits of employing machine learning (ML) in evidence syntheses. Consequently, a dedicated ML team was funded in late 2020, aligning with NIPH strategies for 2019-2024 focused on automation and workflow innovation. The ML team's work has been anchored in the NIPH strategy for the 2019-2024 period focusing on automation, increasing the speed of evidence syntheses, and implementing workflow and methods innovation (1). The ML team can be seen as a strategic innovation, an attempt to change the "business model" of evidence synthesis practices at NIPH to ensure a sustainable advantage over other evidence synthesis providers or groups. ML allows for the most effective use of scarce, valuable human resources while still maintaining high quality of the evidence synthesis products produced.

The ML team has since initiation become an international leader in implementing ML into evidence synthesis processes. Our work has so far culminated in three results reports (2-4), three strategy reports (5-7) three published papers (8-10), three protocols (11-13) (two as preprint), contributed to one book chapter (14), about 40 presentations both nationally and internationally as well as several international workshops.

Our setting for implementation

The overall goal of the ML implementation team is to use ML in a way that best combines human intelligence and ML, to enhance human activities, by figuring out how best to integrate ML and workflow changes, throughout the evidence synthesis process.

When ML implementation was initiated in 2020, HTV consisted of approximately sixty employees, including five leaders, eight librarians, one statistician. The remaining employees were researchers working with different evidence synthesis products of varied topics. The researchers had diverse backgrounds, featuring individuals with clinical backgrounds such as medical doctors, physiotherapists, and nurses, as well as others from non-clinical backgrounds like psychology, sociology, and anthropology. Years of experience with evidence synthesis work varied greatly, where some had worked with evidence syntheses for decades while others were evidence synthesis novices. All employees held a minimum of a master's degree, with the majority of the researchers having a PhD. The amount of technological experience and trust in new technology varied also.

Aim of implementation document

Throughout this report, we will offer overarching guidance drawn from our experiences during the implementation process within the HTV evidence synthesis group. Our suggestions for "take home messages" are rooted in reflections on what we executed well, and the lessons learned, providing insights into what we might approach differently if faced with implementing a new technology again. For more detailed examples and concrete descriptions of our implementation approach, please refer to the appendices, which are cited at relevant points in this report.

This implementation guide is tailored for evidence synthesis groups or institutions aiming to introduce ML into the evidence synthesis process within a ML-naïve work environment. Even though most of our implementation work has been centred around ML functions related to literature searching and screening, the guide aims to be sufficiently general for implementation of various ML functions and tools beyond the presented examples.

The document offers a practical framework for ML implementation into evidence synthesis processes, serving as a roadmap adaptable to each institution's organizational goals and objectives. Where we do not specify specifically that recommendations are made based on an implementation framework, model, or theory, it is based on our own experiences on what have worked/not worked.

Introduction

The evidence synthesis process is labour-intensive, aiming to gather and summarize all available information on a specific topic. While evidence synthesis is crucial for informed policymaking, the traditional evidence synthesis process cannot keep up with the speed policymakers demand, hence there is a growing need to expedite the production of evidence synthesis products without compromising on quality. Using machine learning (ML) functions during the evidence synthesis process is one way to do this. ML is not about replacing human effort but rather streamlining repetitive tasks while at the same time allowing more researcher time for 'thought-intensive' tasks. Recent estimates suggest significant resource savings with increased ML adoption (15-17).

The successful implementation of ML is not merely a technological endeavour; it is a strategic imperative. Beyond enhancing the efficiency of the evidence synthesis process, a well-executed integration of ML can catalyse transformative advancements in research capabilities, decision-making and work processes, and have an overall institutional impact. The adoption of innovative technologies, such as ML, is pivotal for research institutions aspiring to remain at the forefront of their fields. It goes beyond embracing a new function or tool; it entails cultivating a culture of innovation, continuous learning, and adaptability.

Theoretical background: The EPIS framework

Throughout this guidance document we will discuss how implementation research theories and frameworks can be used to structure and guide the implementation process.

The intricate process of implementing and sustaining innovations, such as ML functions or tools, requires a systematic approach, and various frameworks exist that aim to facilitate this by focusing on essential components for successful implementation and evaluation (18). Implementation frameworks are often dinstinguished into three types: *process*, *evaluation*, and *determinant* frameworks (19). One of the most widely used implementation process frameworks is the Exploration, Preparation, Implementation, Sustainment (EPIS) framework (20). EPIS offers a comprehensive view of the implementation process and aids in pinpointing areas requiring special attention (see Figure 1).

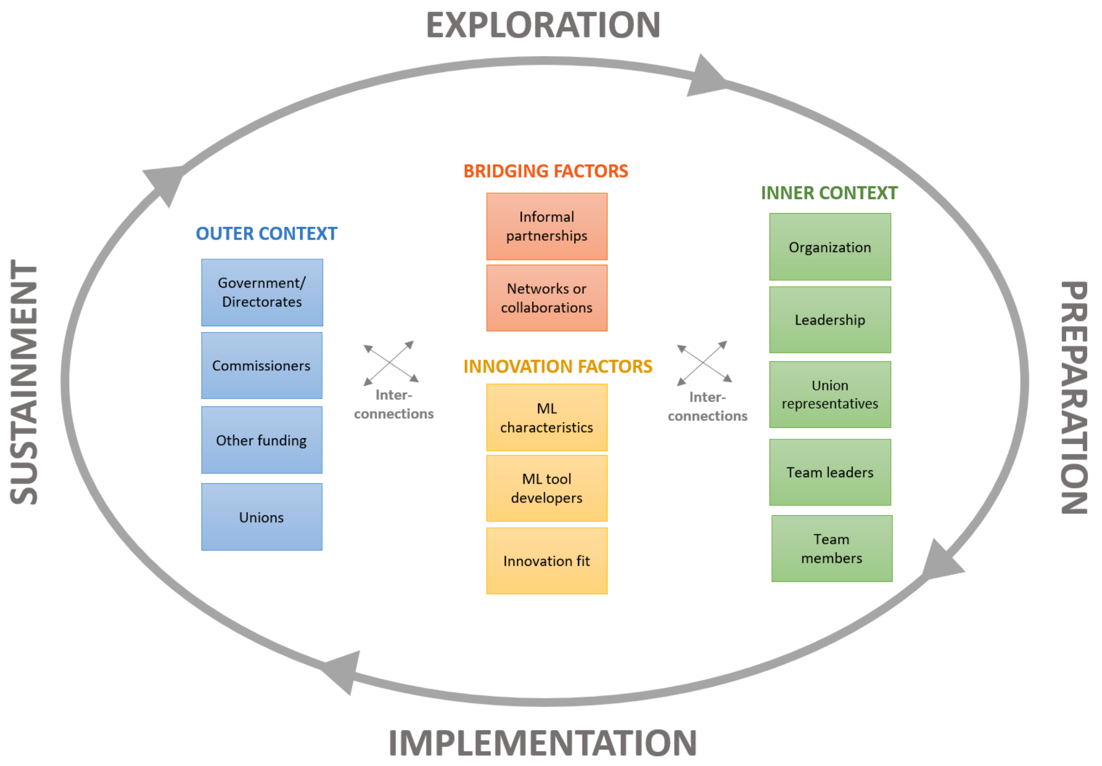

Figure 1. The EPIS framework. Figure adapted from original Figure in (21)

EPIS divides the implementation process into four distinct stages: Exploration, Preparation, Implementation, and Sustainment. The Exploration phase begins with awareness of a need, leading to considerations of addressing it. The Preparation phase involves identifying barriers and facilitators, deciding on implementation supports, and fostering a positive implementation climate. Implementation sees the initiation of the intervention, requiring continuous monitoring for adjustments, while the Sustainment phase involves maintaining the intervention's impact over time.

Recently, a few studies have highlighted the importance of the early stages of the implementation process like the Exploration and Preparation phase (22;23). The so-called *Pre-Implementation* stage may be decisive for successful implementation, with interventions showing rigorous pre-implementation efforts having higher levels of program start-up and competency (24). Due to the importance of this early stage of implementation, we have chosen to structure our report into three sections: Pre-implementation, Implementation and Sustainment.

Throughout this guidance we also build on concepts from the diffusion of innovation theory (25). This theory, proposed by Everett Rogers, explains how new ideas, products, and technologies spread through a population over time. The theory has four main elements: innovation, communication channels, time, and social systems. We refer to these when discussing the implementation process. We also lean on his conceptualisation of a five-step innovation decision process (knowledge, persuasion, decision, implementation and confirmation) as well as the various roles involved (Innovator, early adopter, early majority, late majority, laggards) (25).

Structure of main report contents

This report is divided into three main sections reflecting the phases of implementation: pre-implementation, implementation, and sustainment/ evaluation. In each section we will present key tasks that need to be carried out, discuss the roles different actors will play, including leadership, and suggest ways in which support can be given and feedback gathered. We will finish each section with a "Take home message" summarizing our reflections around our implementation experience on what worked well, and lessons learned for future implementation, which includes practical tips on important aspects to consider for that implementation phase.

Pre-implementation phase

The pre-implementation phase is crucial for ensuring a well thought out and grounded implementation of machine learning in a research organization. During this phase the implementation objectives will be decided, a team will be put together, a communication and implementation plan will be developed, and a timeline set. See Figure 2 for an overview of the activities to be carried out during the pre-implementation phase.

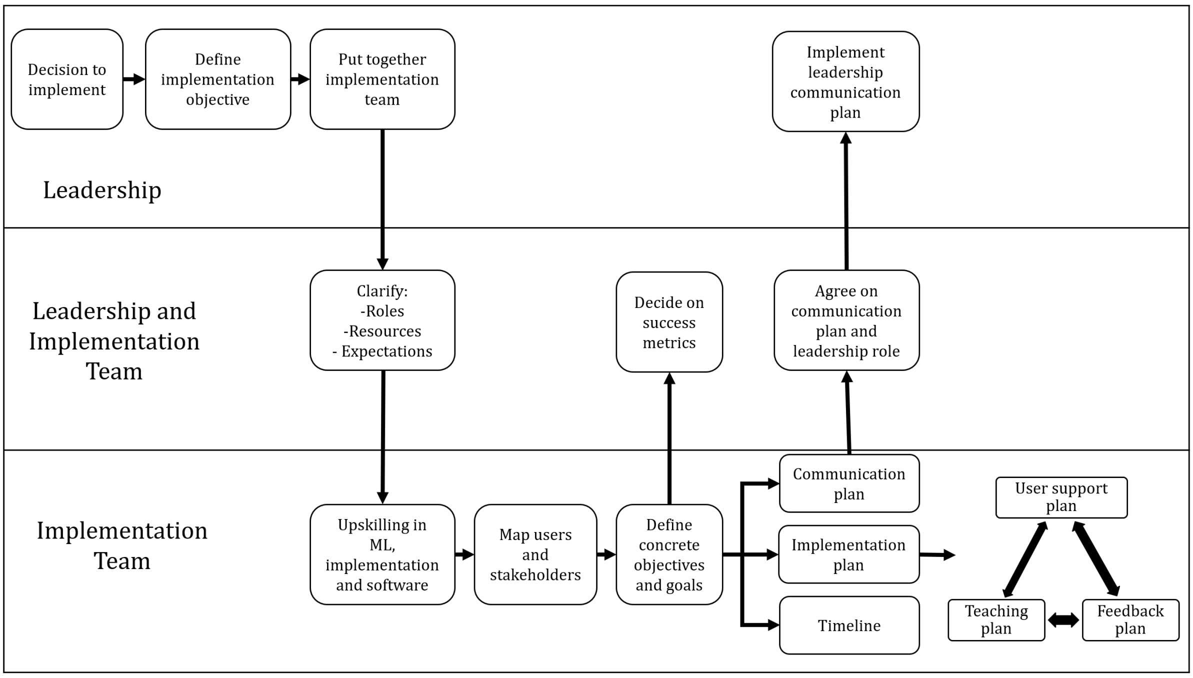

Figure 2: An overview of the pre-implementation phase and activities, based on our own experiences.

The following section will emphasize the Exploration and Preparation stages of the EPIS framework (see Figure 3). Most sections will focus on the inner context, with sections on learning objectives, implementation team, communication plan and implementation plan largely focusing on team members and team leaders. Leadership, such as top leaders, middle managers, and team leaders, will be covered in the section on change management, whereas commissioners in the outer context will be covered in the section on planning for workflow changes.

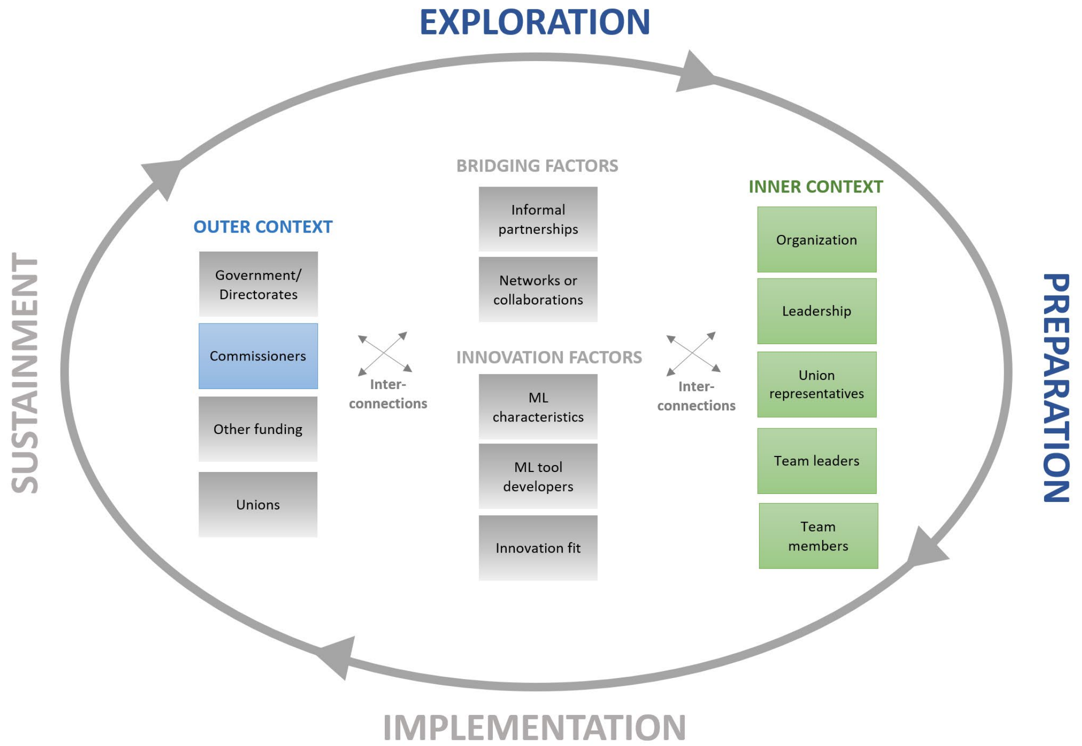

Figure 3: The EPIS framework with emphasis on the Pre-implementation stage

Application of the EPIS framework

Implementation frameworks can help inform and structure the implementation process by highlighting areas of the implementation process that would otherwise have not received sufficient attention. If an organization introduces an innovation without proper attention to factors influencing intial implementation (e.g. failing to identify barriers and facilitators), it may enhance the risk of detrimental consequences to the subsequent implementation stages (e.g. reduced acceptance, higher turnover etc.) (20). As outlined in Figure 3, Exploration and Preparation, are both part of the Pre-implementation phase and are important to address during intitial implementation efforts.

During Exploration, relevant stakeholders in the organization consider the needs of clients or employees and try to find the best innovation to address those needs. In particular, three areas are viewed as important when exploring different innovations: 1) readiness for change, 2) receptive context and 3) absorptive capacity (21).

First, an organization's readiness for change refers to employees' shared resolve to implement the change and their belief in capacity to do so (26). Second, receptive context may entail both a positive implementation climate (e.g. employees view the organization as welcoming of the innovation) and culture (e.g. employees experience that the organization promotes the use of innovations). Third, absorptive capacity concerns the organization's existing knowledge and skills, its ability to use new knowledge and mechanisms to to support sharing of knowledge.

The Preparation stage of the EPIS model is also referred to as the "Adoption Decision" stage and there are various factors that are likely to impact this decision, including organizational structure and leadership. An innovation with a high degree of fit with organizational structure, roles and responsibilities is assumed to more likely be adopted than an innovation with a poor fit. Moreover, a supportive leadership can play a decisive role in whether the innovation is implemented or not. Moreover, the presence of an innovation champion in the organization's leadership, is expected to enhance the likelihood that the innovation will carry through the Exploration and Preparation stage (27).

Other theoretical model and framworks may be used to address areas of change in the pre-implementation phase. For instance, the Transtheoretical model (TTM) has been used successfully to provide a systematic approach to organizational readiness (28) (see further details under "Addressing hesitancy amongst employees").

Change management – a crucial element when implementing change.

Implementing an innovative technology, such as ML, into a workplace is an example of a change process, as it requires employees to change how they work, and effective change management is essential in such processes. Change management is crucial for organizations to navigate transitions effectively, minimize disruption, and maximize the likelihood of successful outcomes (29). Change management focuses on figuring out how to get through the change process efficiently, rather than deciding on specific goals or strategies to reach the goal. Change management is more about helping the people inside an organization adapt to changes happening within it (30).

In change processes, various levels of management play different but equally crucial roles. Top management is responsible for setting the direction and clearly communicating goals for the change process. Their clear leadership is crucial for creating understanding and loyalty to the change among employees. At the same time, top management must establish structures and ground rules that support the orderly implementation of the changes (31).

Middle managers, on the other hand, are increasingly recognized as critical players in change processes. They function as the link between strategic goals and day-to-day operations and must balance the need for change with the need for stability. Middle managers have a unique opportunity to translate the organization's strategy into action and must be able to effectively manage employee reactions to change. Their role extends from implementing the changes to actively participating in the project work that supports the change process (31).

The project manager has an operational role in the change process, with responsibility for the practical management, planning and implementation of the changes. They ensure that the changes are in line with the organization's strategic goals and that the processes supporting the change are effective (31).

In our process, the ML team members have taken on the role of both middle manager and project manager as it is the ML team that has been responsible for the practical management, planning and implementation of the changes, as well as directly participating in the different projects where ML has been implemented. Throughout this document, and the different implementation phases, we highlight the key role and tasks of top managers and middle managers in supporting and sustaining the change process and refer to them collectively as “leadership” throughout.

Securing support and buy-in from leadership in all phases of the change process is crucial for the successful implementation of ML in evidence synthesis within an institute or group. This support is essential for several reasons:

- It ensures the allocation of necessary financial and human resources, guaranteeing funding, skilled personnel, and infrastructure to facilitate implementation.
- Leadership involvement aligns the integration of ML with the institute's strategic objectives, preventing divergence of efforts and resources.
- Leadership buy-in signals to the group and the broader organization that the implementation is endorsed at the highest levels. This endorsement can help mitigate resistance and encourage a more positive reception of the workflow- and process changes that comes with using ML in the evidence synthesis process.
- Leadership buy-in from the start signals a long-term commitment to ML integration, providing stability and continuity for the ongoing implementation process.

Defining the implementation objectives

Once the decision has been made to implement ML, planning can begin around the implementation process. One of the first steps is to define the implementation objectives.

The primary goal is to implement the use of ML into evidence synthesis products conducted within an institute or group where the majority of employees are ML naïve, i.e., have little or no experience with using ML in their work. To achieve this overarching goal, consider establishing learning objectives using the SMART goal method and leveraging Blooms Taxonomy.

The SMART method ensures that learning objectives are Specific, Measurable, Attainable, Relevant, and Timely, offering a structured approach to goal-setting that enhances the effectiveness of the implementation process (32). Blooms Taxonomy aids in formulating measurable learning outcomes based on the desired level of skill or learning amongst the employees (33). The taxonomy comprises Lower Order Thinking Skills (LOTS) and Higher Order Thinking Skills (HOTS). LOTS cover skills related to remembering, understanding, and applying concepts, while HOTS involve skills associated with analysing, evaluating, and creating. This division allows for addressing a spectrum of cognitive abilities in the implementation of ML. Proficiency in both lower and higher order thinking skills equips employees to navigate the complexities of ML in evidence synthesis, contributing to a more efficient and effective production of evidence synthesis products.

For LOTS, learning objectives for employees could be:

- Understand basic conceptual knowledge of the use of ML in evidence synthesis processes (e.g., what is a ranking algorithm?).
- Comprehend the basics of how relevant ML functions work (e.g., ranking algorithms, classification algorithms).
- Understand the distinctions between different ML functions (e.g., difference between supervised and unsupervised ML).
- Understand the fundamental principles of a graph neural network.
- Apply ML functions in the evidence synthesis process.

For HOTS, learning objectives for employees could be:

- Determine the appropriate ML functions for different evidence synthesis projects.
- Establish a routine for documenting the use of ML in evidence syntheses.
- Evaluate the effectiveness of ML functions for specific projects.
- Evaluate the prerequisites, possibilities, and pitfalls of using ML in the evidence synthesis process.

Putting together the implementation team

After the decision to implement ML has been made, a decision regarding how ML should be implemented has to be made. Based on our experiences we suggest putting together a specialized ML team that will be responsible for the implementation efforts. We have had great success by structuring the implementation in this way. This team will become the in-house ML experts and lead the implementation process. It is crucial that leadership define clear expectations, time allocation and guidance for the team. This will form a clear mandate for the team's work, to avoid uncertainties and unclear directives around the team's scope, role, and objectives.

ML team structure

The team should consist of a team leader and team members. If other activities beyond implementation are included in the team mandate, we suggest naming an overall team lead, as well as an implementation lead. The total number of team members will depend on many factors. Based on our experience, key factors involve:

1. The size of the institute or group that comprise the implementation unit, i.e., the employees that will receive the implementation efforts.
2. The innovation maturity and implementation readiness of the employees in the implementation unit, as this will guide the potential scale of the implementation efforts.
3. Each team members' availability (the amount of time available to dedicate to the ML team). To be able to have good progression with the implementation, we recommend a minimum of 40% protected time for team leader(s) and minimum 20% protected time for other team members during pre-implementation and implementation. This is based on our own experiences in our organizational environment, and other groups or organizations might need more or less time, depending on the employees' capacities and scale of planned implementation.

ML team members having protected time for the ML implementation activities is crucial for a successful implementation. Providing protected time allows team members to dedicate focused hours to learning and mastering ML skills. Implementing ML in evidence synthesis is a complex task that involves understanding both the theoretical and technical aspects of ML functions, how existing workflows will change, as well as gaining knowledge on how to effectively implement ML in the evidence synthesis process. Also, ML is a rapidly evolving field, necessitating ongoing learning to stay updated on advancements and adapt to evolving best practices. Protected time supports attending training, staying informed, and continuous skill development. Lastly, providing protected time fosters innovation by allowing team members to explore creative approaches, experiment with methodologies, and tailor solutions to the specific needs of the research group. It also enhances productivity and reduces burnout, fostering a culture of learning and problem-solving within the team.

In Appendix 1 we have provided a concrete suggestion of the different team roles that have comprised our ML team throughout our whole implementation process, as well as key competency requirements.

Capacity building within the team

One effective method, especially for educating the ML team and other adoptees, is the "train the trainer" approach. The ”train the trainer” approach is a methodology for preparing individuals to pass the methods and expertise they have learned on to others, who then becomes trainers themselves (34). The decision to adopt this approach stems from our commitment to empowering researchers within our institute to be independent in utilizing ML, hence, to ensure that the implementation of ML was not solely reliant on the capacity of a dedicated ML team or function experts within the team.

To kickstart this approach, we recommend assigning each team member responsibility for a specific function earmarked for implementation within the institute or group. Each team member then becomes a specialist in their designated function and is tasked with creating or curating comprehensive training materials for “their” function. These materials serve a dual purpose, educating both the ML team members as well as employees outside of the team. Once team members have developed a profound understanding and acquired skills in using their designated function, they can employ the same methods to teach other employees within the institute or group.

This approach offers several advantages to organizations. Scalability is a key benefit, as a smaller group of experts can efficiently extend training efforts to reach a larger audience. The approach also facilitates expertise distribution by designating specialists in various areas. Flexibility is another notable advantage, as trained trainers can adapt their materials to the specific needs and skill levels of their audience. Furthermore, this approach promotes sustainability in training initiatives, establishing a self-sustaining system that not only disseminates knowledge effectively in the present but also identifies and trains new instructors as the need arises. This longevity fosters a culture of continuous learning and development within the organization.

Communication plan

In the pre-implementation phase, the team needs to work together with leadership to create a communication strategy to be used throughout the ML implementation. We will discuss communication strategies for specific function/tool implementation in the next section of the report. There is a lot of existing literature on the development of communication change strategies (e.g.: (35-37)). We will not enter into detail here but provide an explanation of how, from our experience, the process could be shaped.

Mapping users

As part of the implementation process and to effectively develop a communication plan, it is important to map the users who will receive or be involved in the implementation, by doing a stakeholder analysis (38). In our case the users are the employees that will receive the implementation efforts for them to start using ML in their work. This is something we did not do initially but did at later stages in conjunction with developing e-learning courses. In hindsight, we acknowledge that mapping users during the pre-implementation phase would have benefited the implementation process and would have resulted in a more tailored implementation approach.

What is important when mapping your users is considering the unique situation at your workplace. When mapping users it is important to get an overview of their characteristics which will influence the implementation approach(es). To map users, perform a user needs analysis, which involves gaining an overview of:

- employees' "wants and needs", goals and aspirations when it comes to ML,
- current ML and technical expertise levels,
- learning preferences,
- attitudes towards using ML in their work.

These user profiles will then serve as a basis for your communication strategy development.

When we mapped our users during later phases of implementation, we used our experience so far from the ML implementation efforts as a foundation. In our case, our employee pool is quite heterogenous when it comes to ML experience level, methods expertise, learning preferences and openness to innovation, which is reflected in the user profiles in Appendix 2, which represent the results of our user mapping.

Team role

The implementation team is responsible for drafting the communication plan and strategies for implementation. The team should be aware of any pro-innovation bias and reflect over how their perceptions, attitudes and assumptions may impact the communication strategy (25). The team should also try to align the communication plan and strategies for implementation to focus on compatibility with the values of the group with which they are working. For example, transparency, trialability, quality and observability (25).

The communication plan should clearly state why the new technology is being implemented, the messaging that will be delivered, communication channels and timing, a plan for dealing with hesitancy, difficult questions and debate (25). The plan should also include a detailed time plan for communication with the message, channel, sender, and responsible party for implementation included. The message that will be communicated out to employees should be short and precise. Figure 4 contains an example of the outline of a communication change message used in our division. Once the plan and strategy are developed, the team will present them to leadership for feedback, discussion, and approval.

Figure 4: Example of points outlined in a communication strategy for ML implementation.

Leadership role

The role of leadership is to provide feedback on the communication plan and strategy. Once they approve the strategy they should be actively involved in implementation and fulfil the tasks allocated to them within the communication plan. In order for adoption to be even across employee groups (if more than one leader) all leaders need to deliver the same message consistently and with the same frequency.

Addressing hesitancy amongst employees

Addressing hesitancy among employees during the implementation of innovative technologies, such as ML functions, should be a continuous process that involves several key communication strategies. When we first started implementing ML in our group, this was something we did not do but should have done. In the following section we will present some recommendations on how this can be done, based on our experiences so far.

Firstly, transparent communication should be emphasized, stressing the reasons behind ML implementation and the overarching goal of enhancing evidence synthesis workflows. Acknowledging the diversity in individuals' acceptance of new technologies is key, and tailoring ML introduction to meet specific needs and concerns is advised. Also, it is crucial to dispel misconceptions and highlight the positive outcomes resulting from integrating ML functions while being open to critique and critical questions. We therefore advocate for an "open door" policy, particularly during initial implementation phases, as an open dialogue will create an environment where people feel comfortable expressing concerns and asking questions. Furthermore, implementing pilot projects for small groups to test and provide feedback on ML functions before full-scale integration can be considered as a method to facilitate gradual acceptance and adjustment, ensuring a smoother adoption process, as well as aid in building trust and credibility in the ML functions.

As it is likely that the employee base consists of people in different stages of readiness or willingness to adopt a new technology, you can use a behaviour change model, for example the Transtheoretical model, also called the 'Stages of change model' as basis for how you communicate and implement change (39). This model fits well with the EPIS framework, as EPIS provides the overarching framework for the whole implementation process, while a behaviour change model specifically addresses how implementation strategies can be best adapted to your user's readiness for change. The Transtheoretical model suggests that behaviour change involves progress through six stages of change: precontemplation, contemplation, preparation, action, maintenance, and termination (39). In Appendix 3 we have provided some concrete suggestions of how each stage can be addressed in relation to communication and implementation.

Developing an implementation plan

There are three areas that should be considered when developing an implementation plan: user support, teaching and feedback. Each of these areas needs to be considered in order to have a successful implementation. The implementation plan should be reflected in the communication plan.

User support plan

The implementation of an innovative technology or innovation may involve a steep learning curve for the users, i.e., the employees. To secure a successful implementation, a plan needs to be developed that focuses on how the ML team will support end users through the implementation phase, addressing their questions and concerns as well as answering technical questions about how the software works and if mistakes have been made while setting up an ML function. The user profiles developed as part of the communication plan can also help to highlight what type of support end users will prefer.

In collaboration with the evidence synthesis teams, we have found that the following approaches by the ML team, especially in the initial phases of the implementation, benefit the implementation efforts when providing support to project teams:

- Be flexible to support individuals who need/want one-to-one support. These individuals will create further interest in ML, and help to learn others and work independently.

- Be open and empatic to critical questions, the people you teach should be encouraged to ask questions, also the critical ones. Often these critical questions reveal weaknesses in team knowledge or holes in the teaching material that need to be addressed.
- Be very transparent with the reasons for why the specific ML functions are used and clearly communicate the results of any evaluations behind them.

Training a super user

To address technical questions related to the software that inhabits the ML functionality used, we recommend training at least one super user. The super user(s) will have responsibility for addressing more technical questions regarding the tool(s) used by the evidence synthesis group, and not just use related to ML functions. All ML team members should have a good knowledge of the technical aspects of the software so that only more challenging support problems move up to the super user level. If the data management tool used provides user support, consider referring employees to them for technical questions, particularly if group resources are scarce.

Team help-requests and one-to-one support.

Evidence synthesis teams need to be able to reach out and request support when they need it. This should begin during the protocol phase when teams are planning what type of ML they will use in their project. We have found that assigning an ML contact to each team has been a successful approach (see Appendix 1 for detailed information on the ML contact). The ML contact is a member of the ML team and is responsible for providing one-to-one ML support to the evidence synthesis team for the duration of the project. Thought should be given to how teams can request help and how ML contacts will be assigned to a project. We have found the use of an online request form coupled with a low threshold for requesting help over email to be successful. ML contacts are assigned at weekly team meetings based on competency and availability.

Training plan

Based on the preferences identified in the user profiles (see Appendix 2) as well as the capacity of the team, a training plan should be developed. The training plan needs to cover conceptual knowledge, technical how-to and understanding when the ML tool or function should be used. During the pre-implementation phase the team will not be developing the learning content, but deciding on the learning channels, frequency, and intensity of training activities.

A possible way of thinking about and structuring the training plan is to focus on the three types of knowledge as described by Rogers (25;40): Awareness knowledge, How-to knowledge, and Principles knowledge. The objectives would reflect:

- Awareness knowledge - the knowledge of the innovation’s existence
- How-to knowledge - the knowledge of how to use the innovation appropriately.
- Principles knowledge - how and why an innovation works.

Feedback plan

Establishing a feedback mechanism is crucial for evaluating the implementation efforts. The collection and utilization of feedback plays a pivotal role in enhancing the innovation and adapting its implementation over time.

Incorporating various feedback mechanisms and structures is essential for a comprehensive assessment. Mechanisms can include employee data collection through tools like questionnaires, piloting of training materials and qualitative interviews. These tools not only capture real-time information but also allow for iterative adjustments based on evolving needs.

Using questionnaires to gather feedback at multiple stages of the implementation process, can be an easy and resource efficient feedback mechanism. Here employees can report issues, provide suggestions, and express concerns, allowing for adjustments throughout the implementation process. Additionally, qualitative interviews, although more resource intensive, provide more nuanced insights into user experiences, attitudes, and perceptions, complementing the quantitative data gathered through questionnaires, which ensures a more nuanced understanding of the implementation efforts, allowing for adjustments throughout the implementation process. Piloting training materials with a small user group provides similar insights, however, these will be specific for the materials piloted.

Another valuable feedback mechanism is the help request system (see 'Team help-requests' above), by which organizations can gain insights into the specific support needs of employees and the project teams' level of experience with ML, both on an individual level as well as on a project level. This data is invaluable for tailoring assistance to each project team, as well to keep track of how many project groups require help over time. For example, you would expect that fewer groups would request and need face to face support over time, as the ML knowledge level increases amongst the employees and they become more autonomous in using ML in their work. Alternatively, the amount of help requests might not decrease substantially, but the nature of the help requests change, by requesting help only with the more advanced ML tools or functions available to the employees.

These feedback mechanisms serve multiple purposes, such as identifying areas that work well and areas that require improvement. The process is iterative, where input from one source, like qualitative interviews, informs adjustments to another, like questionnaires. This iterative approach enhances the precision of the feedback mechanism, ensuring that relevant areas for improvement are accurately identified and addressed.

Looking ahead, adopting innovative feedback mechanisms for the future is crucial. This may involve exploring emerging technologies and employing more advanced analytics to extract deeper insights from the feedback data. The goal should be to continuously evolve the feedback process, making it more efficient, user-centric, and aligned with the evolving landscape of ML in evidence synthesis processes as well as with the organizational goals.

Planning for workflow changes

In recent years, the implementation of ML has transformed the way research is conducted in HTV. This technological leap has not only accelerated the pace of deliveries but has also brought about significant changes in the process of conducting a systematic review. Here, we present the key aspects of the workflow processes that have changed in order to use ML efficiently.

More intensive teamwork

The integration of ML function has led to a shift from individual efforts to more intensive teamwork. Larger researcher teams, engaged in multiple systematic reviews with extended deadlines, have transitioned to smaller, tightly knit teams with shorter timelines. These teams now employ agile project management methods, emphasizing increased collaboration, agility, and fewer projects per team member. Agile project methods have introduced flexibility and adaptability into our workflow. Embracing iterative cycles, regular feedback loops, and swift responsiveness to emerging insights or challenges have become standard practices.

Simultaneous work across review steps

Traditional sequential steps in the evidence synthesis process have given way to a more simultaneous approach. Teams no longer wait for one step to be completed before proceeding to the next. For example, during title and abstract screening, team members may concurrently begin reading the full text of included titles and abstracts. Other simultaneous processes, such as data extraction and risk of bias assessments during full-text assessment, enhance efficiency and provide a comprehensive overview of the included studies early in the process.

Working more closely together during screening and study selection

Teams now collaborate more closely during the study selection phase. Using ML correctly allows for relevant references to be screened much earlier on in the screening process, in contrast to throughout the entire screening process which is the case for screening at random. Hence, it is even more important to maintain consistency in screening practices when utilizing ML, as the algorithm's effectiveness relies heavily on the quality of its input.

To ensure uniformity in screening, an initial calibration meeting is conducted, aligning all team members with the inclusion and exclusion criteria. The emphasis is on simultaneous screening practices, with regular meetings to ensure ongoing consistency in screening practices and promptly resolve any disagreements that may arise during the screening process.

Changes to study selection timeline

Incorporating ML functions shorten the study selection timeline from months to days or weeks, reallocating time for more intellectually demanding tasks like data extraction, risk of bias assessment, analysis, and writing. Consequently, the finished evidence synthesis product can often be delivered within a shorter time frame.

Communication with commissioners

Using ML transforms communication dynamics with commissioners, offering early insights into anticipated research findings. By being presented with the majority of the relevant references in early phases of the screening process, the review team can provide realistic expectations towards the commissioner and establish an accurate and achievable timeline. This more streamlined workflows facilitate efficient and transparent communication, enabling prompt updates and increased responsiveness to commissioner needs or modifications.

Decide on key performance indicators (KPIs)

Measuring the success of the ML implementation in the evidence synthesis process requires a set of key performance indicators (KPIs). These KPIs should provide a clear and tangible assessment of the impact of using ML on both efficiency and outcomes. Regular monitoring and analysis of KPIs will provide actionable insights for ongoing optimization and improvement of the implementation efforts. Utilizing KPIs involves establishing performance benchmarks (desired levels) and monitoring advancement toward these benchmarks (41). The KPIs should be decided on during the pre-implementation process and should be based on the objectives and goals of the implementation. For suggestions of specific KPIs, see Appendix 4.

We recommend developing a measure to assess KPIs early in the pre-implementation phase. Ideally, the measure should be able to assess domains perceived as important by the different levels of the inner context of the implementation (e.g., the organization's strategic goals, leadership aims and team member preferences). A set of qualitative interviews would be useful to determine what is perceived as important by different groups. However, if there is a lack of time and resources to carry out interviews, selection of KPIs could be based on relevant theories and research literature.

Another way to measure if KPIs for the implementation process have been reached is to develop an employee survey that can be repeated during different stages of the implementation process. One way of developing a survey, based on theory, research and feedback from union representatives, is described in Appendix 5.

Take home message for the pre-implementation phase.

One major lesson we learnt from initiating the ML efforts at NIPH in 2020 is that we largely ignored conducting a pre-implementation phase before implementing the use of ML in the evidence synthesis process amongst our employees. As the pre-implementation phase is a crucial stage in the process of integrating ML into evidence synthesis processes (24), we highlight below several important aspects that should be considered during this phase to ensure a successful and effective implementation, based on our experiences so far. By carefully addressing these aspects during the pre-implementation phase, organizations can set a strong foundation for a successful integration of ML into evidence synthesis processes. This phase lays the groundwork for a well-planned and informed implementation, increasing the likelihood of positive outcomes.

Defining the implementation objectives

- Align the ML implementation strategy with the overarching goals and strategic objectives of the evidence synthesis group or institution. By ensuring that ML efforts contribute to the group's mission and enhance the overall effectiveness of evidence synthesis, it will contribute to earning employees and leaderships buy-in towards the implementation.
- Conduct a risk assessment to identify potential challenges and obstacles that may arise during implementation. This would also include developing a contingency plan and mitigation strategies to address identified risks.

Putting together the implementation team

- Evaluate the availability of resources, including financial, technical, and human resources, required for ML implementation. Determine the budget, infrastructure, and personnel needed to support the pre-implementation and subsequent phases.
- Define the composition of the ML implementation team, including both their current skills, interest areas and personal suitability for specific roles in the team.
- Assess the current skill set within the team and identify any training needs for building expertise in ML methodologies.

Communication plan

- Conduct a thorough needs assessment to map the target users (i.e., employees) and stakeholders (i.e., other groups or persons involved in the evidence synthesis process, like commissioners, reference- and user groups as well as the developers of the evidence synthesis tool used) who will be affected by the implementation.
- Employees should be mapped and recognized by potential user group (for example researchers, librarians, leaders) and technological literacy or openness to change/innovativeness. Part of this process will be to identify key employees who should be targeted in each implementation phase.
- Once the target users have been identified and mapped it is helpful to create a user profile for each group that investigates and maps technological considerations, barriers to learning, digital learning experience and learning preferences. For an example of user profiles see Appendix 2.
- Clearly communicate the potential benefits and long-term vision of ML implementation to gather support from employees. Before we implemented ML in our group, we did do in-house evaluations with the aim of building trust and highlighting benefits of using ML, but in hindsight this work was not communicated clearly enough.
- Continuously address concerns and expectations and establish channels for ongoing communication throughout the implementation and sustainment process.

Developing an implementation plan

Another thing that we did not do, but should have done, was to use a theoretical basis for our implementation. Implementation theories, models and frameworks are especially important tools that will provide you with a roadmap towards systematic, transparent, and effective ML integration in the evidence synthesis process. It enhances planning, execution, and evaluation across different phases of implementation, contributing to the overall success and sustainability of the implementation effort. In the introduction we introduced some frameworks and models that can be used, but please be aware that this is not an exhaustive list, just a brief selection. To read more about this field, we advise you to read some of these key papers (19;42-44).

Implementation phase

In this section we will discuss the implementation of a single ML function into an evidence synthesis group. See Figure 5 for an overview of the activities to be carried out during the implementation phase.

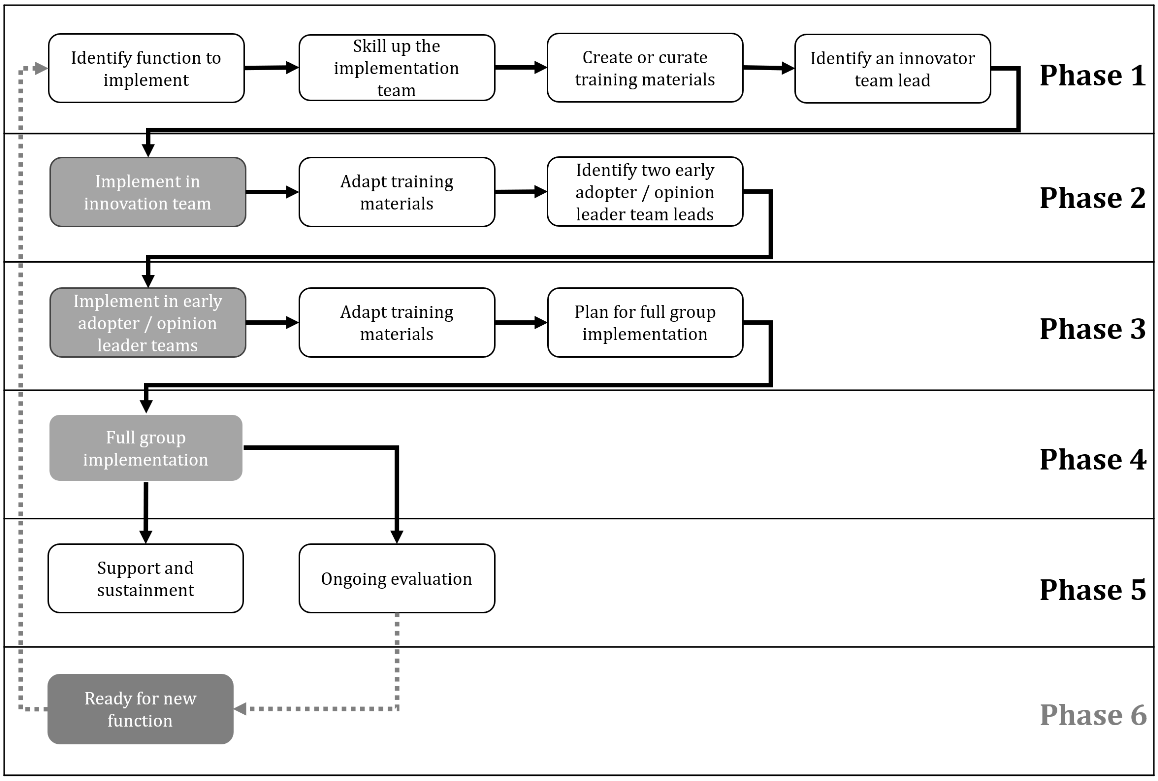

Figure 5: An overview of the activities to be conducted during the implementation phase.

The following chapter will emphasize the Implementation stage of the EPIS framework (see Figure 6). Most sections will concern the inner context, primarily how to work with team leaders and team members on skilling up the team and building their training materials. Other organizational levels may also be involved at this stage, including a communication department which may assist with utilizing the communication plan. Innovation factors, such as innovation characteristics, the innovation developers and innovation fit, are all relevant to the current chapter, especially in the section on identifying the function to implement.

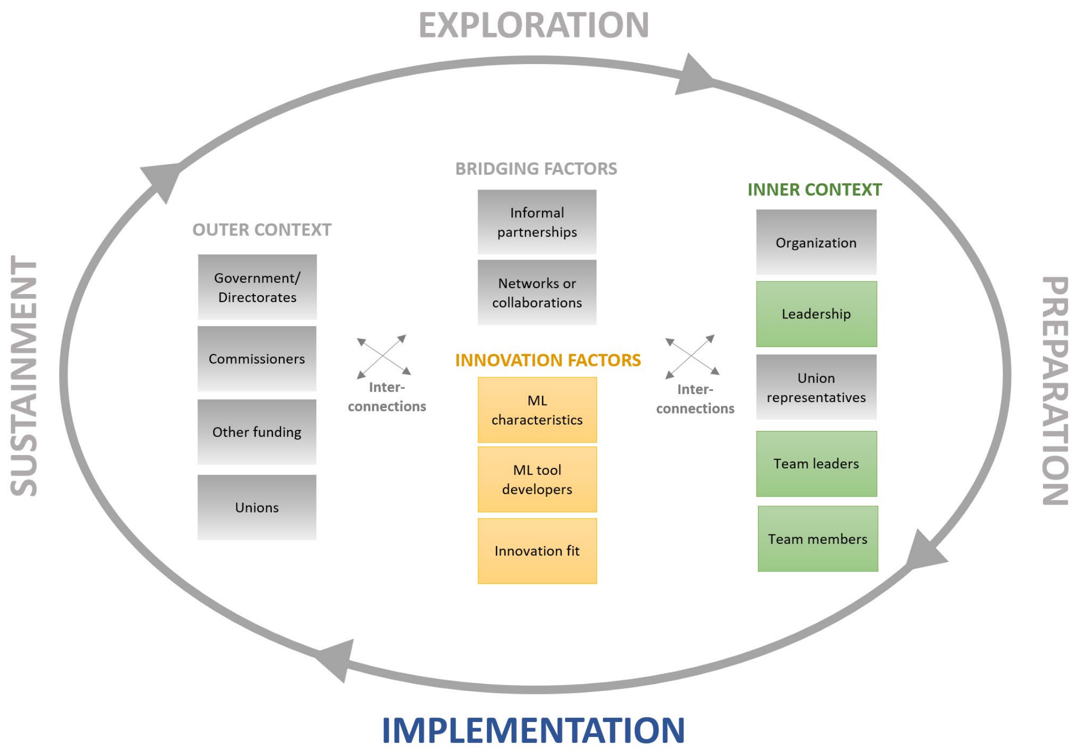

Figure 6: The EPIS framework with emphasis on the Implementation stage

Application of the EPIS framework

Once the Pre-implementation stage is over, the work of putting the innovation into practice (i.e., the Active Implementation stage) can begin. During this stage, according to the EPIS framework, it is critical to incorporate ongoing monitoring of the implementation process and to adjust the implementation strategy accordingly (20). Ongoing fidelity monitoring has previously shown to be successful in enhancing staff retention, when implemented with evidence-based practices (45).

Other important components during active implementation include factors like recruitment and selection, preservice and in-service training, staff performance evaluation, and consultation and coaching. These are all referred to as “implementation drivers” and are assumed to lead to successful implementation (46). A common aim of many of these drivers is that they serve to enhance competency. For instance, preservice training may be an efficient way to increase knowledge of the innovation, whereas ongoing coaching can help professionals learn skills on the job. Drivers like staff performance evaluation, on the other hand, can help maintain necessary skills by ongoing monitoring (see above). Finally, drivers like staff selection can be important when certain professional characteristics are difficult to teach (e.g., willingness to learn, common sense etc.).

Communication plan

Using the overarching communication plan developed during the pre-implementation phase as a guide, the implementation team needs to develop a communication plan specifically related to the implementation of each function. This plan should include:

- The communication activities to be carried out
- A timeline for when messages will be sent
- The content of the messages
- Who the sender of the message will be (team or leadership)
- Who has responsibility to send the message

The communication plan around function implementation should contain the same core message as the overarching plan. In addition, the communication plan should include announcing the implementation of the new function, when invitations to various learning events need to be sent out, reminders for registration and follow-up communication around questions and support needs.

Identify the ML function to implement.

The first step in the process is to determine which ML function the group would get most benefit from. From our experience, implementing the most straightforward and simplest function first is ideal, like a ranking algorithm. A ranking algorithm is a type of supervised machine learning which reorders the references to be screened based on the decisions taken by the review team. It pushes references that are similar to the references included by the human reviewers to the front of the queue. Introducing this function first will provide a "soft" introduction to ML functions, and it will pose the smallest risk of being perceived as overwhelming to the employees. Then, when the uptake of that function is well integrated into the employees work processes, a new function can be implemented. Also consider whether the whole employee pool should receive implementation of all functions. E.g., if an organization has a dedicated librarian team it might be more relevant to implement an ML function relevant for their work first, e.g., OpenAlex or another literature search function.

If there is uncertainty around which function that should be implemented and in which order, the implementation team should conduct a needs assessment and determine what the group where the function will be implemented feel is the most important to streamline their work processes and meet their current needs.

Skill up the implementation team

The team implementing the function needs to become the in-house experts around how the function works both conceptually and technically and have a level of knowledge where they feel comfortable responding to questions, training others how to use it and providing hands on support and assistance. We have used a train-the-trainer approach to upskilling our team on a new function. The tool is broken down into its components and each team member is assigned a component to research, trial and train up the other team members. When necessary, we have called in external supports for technical training or conceptual understanding for more complex tools such as custom classifiers.

Create or curate training materials.

When developing training materials, a number of elements need to be taken into consideration. The team should refer to their user profiles to help frame the learning objectives and decide on the implementation pathway. When introducing a new ML function, three aspects need to be covered; conceptual knowledge of how the function works, technical knowledge on how to use the function or software and an understanding of when and why to use the function in an evidence synthesis process. For examples of training activities we have conducted, see Appendix 6.

The team should assess if there are existing training materials that can be used that fit their needs. If materials exist, the team needs to assess if they need be tailored or curated to match the in-house context. For example, we mainly use EPPI Reviewer, a data management tool (47) in our work when screening references, and they have an extensive portfolio of guidance documents and videos explaining how to use the tool.

If no existing materials are found, the team will need to build their own training materials. The training materials should be piloted first on the ML team members. Further piloting is described in the next section. A suggestion for a detailed work plan is presented in Appendix 7.

We have developed training material for our most used ML functions, which includes plain language written PDF's on both the conceptual part and technical part of an ML function. In addition, we have developed e-learning courses designed to be interactive and short covering the conceptual parts of the e-learning functions. We have gotten great feedback on these e-learning courses. Another format of training material we have found valuable is recordings of previous digital meetings where one or more members of the ML team has held a workshop covering the conceptual and/or technical aspects of how to use an ML function.

The implementation process.

The process for implementing a new function is described in Figure 5. It is an iterative process where the implementation of the new function happens first in a team with an innovator team lead. Innovators are open to new concepts and eager to be the first to try new functions or tools. However, they can be on the outside of the social circles within a working environment and do not have the influence of an opinion leader or early adopter (25). The team receives training in the new function and support to implement it in a project. During this process they provide feedback on the training materials and the support framework. After the implementation is completed in the innovator team, the ML team adapts the training materials and support process based on the feedback.

The process is then repeated with two further teams that have early adopters or opinion leaders as team leads or prominent members. It is important that this phase is conducted with early adopters and/or opinion leaders in order for others to gain acceptance of the new function. Early adopters and opinion leaders have social influence within the research group. Early adopters are looked to for what the next cool or influential thing should be. Others follow their lead and often choose to use the same functions or processes they do. Opinion leaders are respected in the work environment. If they adopt a new function or technology, it often leads others to believe that adopting the same technology is a good choice (25). After implementation with and feedback from these two teams, the ML team makes the final adjustments to the teaching materials and support materials. Once these are completed, the plan for full group implementation can be created and followed.

The full group implementation plan should be linked to the communication plan and contain the specifics of which training activities will be conducted when and by whom.

After full implementation, a system of sustainment, support and feedback should be put in place. This as well as deciding when a new function can be implemented are discussed in the next section of the report.

Take home message for the implementation phase.

A key takeaway from our experiences is that our team faced challenges in effectively communicating evaluation results to employees within HTV during the implementation phase. Subsequent employee satisfaction evaluations revealed that some perceived the use of ML functions as "forced" upon them or were unsure if they could trust them as well as lacking sufficient information on the benefits of integrating ML functions into their specific workflows. In hindsight, it becomes evident that we could have introduced the use of ML in a more tailored manner, one function at a time, acknowledging the diverse ways individuals accept and adopt new technologies. Improved communication strategies, especially in conveying the practical implications of evaluation results to employees, would have been beneficial.

- Conduct a stepwise implementation of ML focusing on one ML function at the time. Introducing too many ML components too fast can overwhelm the employees and create starting point for implementation. We recommend that another ML function is not introduced before the majority of employees have mastered the already implemented ML function(s).

Other tips

- Consider implementing small-scale pilot projects to evaluate ML applications in a controlled environment. Gather feedback, assess the feasibility, and identify any adjustments needed before full-scale implementation.
- Plan for training sessions and capacity-building activities to equip team members with the necessary skills for ML adoption. Consider both technical training for ML functions and methods, as well as broader education on the implications and benefits of using ML in the evidence synthesis process.
- When planning training material and sessions, consider ethical implications associated with ML implementation, including bias in algorithms, transparency, and fairness.

Sustainment and evaluation phase

The following chapter will emphasize the Sustainment stage of the EPIS framework (see Figure 7). We will discuss how to address the inner context in many of the sections below, with leadership being particularly important to the section on support and sustainment of the new function, whereas team leaders and team members are the primary focus in the sections on communication plans and assessing when to implement a new function. Further, the section on survey development will highlight how union representatives may be essential in guiding the development and application of an ML attitude survey. As for innovation factors, the section on ongoing evaluation is particularly relevant to innovation characteristics, whereas the section on monitoring of KPIs is relevant to the innovation fit.

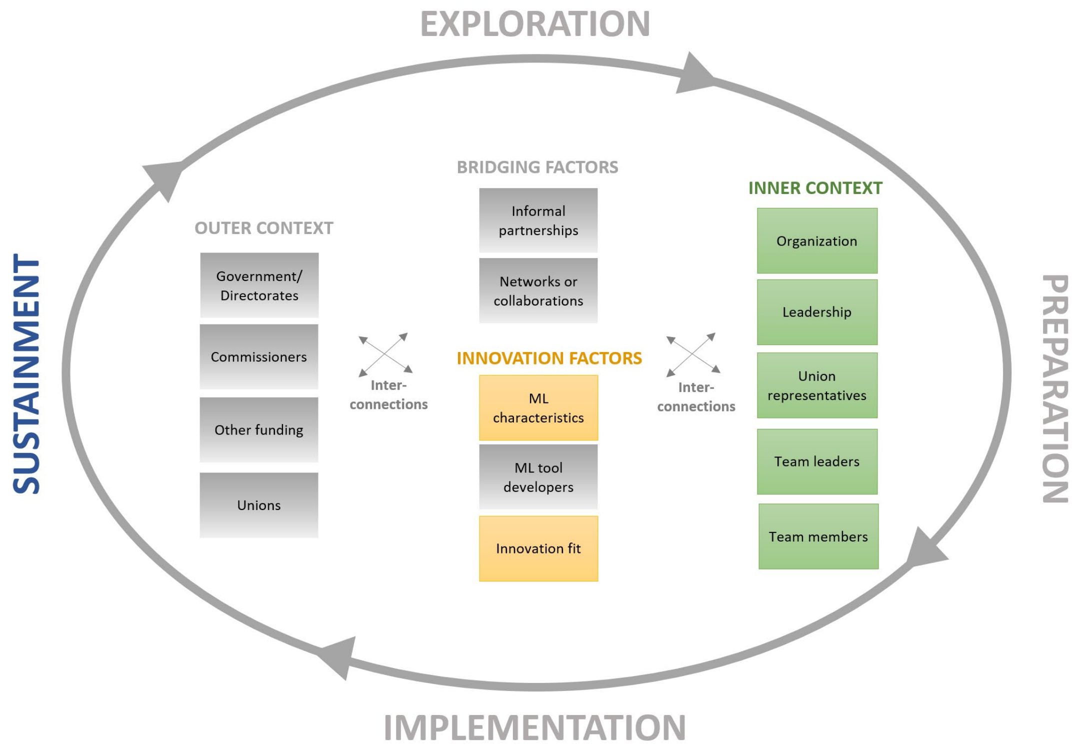

Figure 7: The EPIS framework with emphasis on the Sustainment stage

Application of the EPIS framework

The Sustainment phase in the EPIS framework represents a critical period following the initial implementation of an innovation. This phase focuses on ensuring the long-term viability, effectiveness, and integration of the intervention within the organization.

Sustainment includes key activities like ensuring stable funding and ongoing monitoring, engaging with external structures (e.g., policymakers, funding organizations), and addressing internal structures (e.g., leadership support, innovation fidelity, continued staffing) (48). Successful sustainment efforts are characterized by the establishment of sustainable infrastructure, ongoing training and capacity-building initiatives, and the fostering of a culture of continuous quality improvement within the implementing organization or system.

The implementation efforts involved in sustainment of an innovation could also benefit theoretical perspectives from fields relevant to the innovation. In terms of implementing an ML innovation, information systems theories like the Technology Acceptance Model (TAM) (49) may help inform why practitioners continue to use, or fail to use, the innovation over time. By assessing factors such as perceived usefulness and perceived ease of use, TAM can help identify potential barriers to sustained usage of the technology and inform strategies to address these barriers. For example, if end-users perceive the technology as difficult to use or lacking in usefulness, interventions can be designed to provide additional training, support, or customization to enhance usability and functionality.

Communication plan

In the sustainment/maintenance phase, the communication strategy shifts towards celebrating successes and fostering a community of practice. Regular sharing of success stories reinforces the sustained positive impact of ML on evidence synthesis processes, motivating employees and showcasing the benefits, which boost morale and ensure maintained enthusiasm among employees. Establishing a formal community of practice facilitates ongoing collaboration, knowledge sharing, and the exchange of best practices among employees, enabling employees to learn from each other, share insights, and address challenges collectively.

Continuous learning and improvement are encouraged through ongoing access to training resources and materials, supporting employees in staying updated with ML advancements. The team should plan when reminders will be sent out, in which form and by whom. The reminders should be sent by both the ML team and leadership to maintain the visibility of leadership support for use of ML.

Recognizing the possibility of relapse, the strategy should include a focus on providing ongoing support, refresher training sessions, and access to updated resources to re-engage individuals who may have reverted to previous practices. There should also be a low threshold and open-door practice for employees to ask questions and voice their potential hesitancies and concerns.

Support and sustainment of the implemented function

Below we describe the different forms of user support provided by the ML team to employees, either personally or based on training material and -resources developed by the team. With all types of implementation efforts, continuous endorsement from leadership on the importance to participate in training sessions and engage with other learning resources is central. This support creates a culture that encourages continuous learning and professional development among the employees.

Types of support provided to project teams can include one-to-one support to project teams, easy access, e.g., via an internal website, to in-house developed training materials and other relevant resources. For suggestions of types of support, see Appendix 8.

Assessing when to implement a new function.

Assessing the readiness of a group for the implementation of a new ML function should be linked to the initial implementation objectives and should encompass direct involvement of employees in the decision-making process to enhances engagement and ensure that their perspectives are considered. Here are some specific suggestions for assessments, and combining these methods facilitates a holistic evaluation of group readiness for ML implementation:

Monitoring of KPIs

Regularly monitoring KPIs allows organizations to objectively gauge employee readiness and assess the impact of ML implementation on evidence synthesis processes. You would then need to set a target level, particularly for the hard outcomes, which will decide when the employee group is ready for implementation of a new function. One example could be that 90% of employees are confident in using the existing ML function independently.

Conducting a small pilot test of the new function in one or two projects allows for real-world feedback assessment. Team discussions based on this feedback then inform the decision on whether to implement the new function across the entire group, while at the same time contributing to the creation of tailored training materials.

Ongoing evaluation

The ML functionalities integrated into the data management tool we use are fully developed, accompanied by extensive documentation. Still, for us, conducting in-house evaluations of, for example, a new ML function, was seen as an essential and strategic first step, to be able to document benefits of using the functions in a transparent and concrete way, as this is the best way to increase trust and buy-in among colleagues and leadership. Additionally, these evaluations provided a stronger foundation to evaluate functions' usefulness to our workflows. Some of the evaluations were integrated into ongoing projects or were conducted retrospectively on already completed reviews.

Considering the integration of ML into your workflow to enhance the efficiency of evidence synthesis, while maintaining methodological rigor, involves a deliberate approach, where these evaluations can be very beneficial. In Appendix 9 we provide a general approach to what we have found have been important to consider when doing in-house evaluations. For particularly relevant or comprehensive evaluations, you can consider publishing an evaluation protocol of your planned evaluation to reach a larger audience and to make it available to the international research community. Our experience indicates that striving for publication enhances the quality of our evaluations. A suggested evaluation template is also presented in Appendix 10.

Survey development

We suggest developing a survey to assess employees' overall acceptance and willingness to adopt new ML functions. Our recommended approach consists of combining evidence synthesis methods and theoretical understanding to support the survey development process. In our work, we utilized the Technology Acceptance Model (49) as a theoretical framework due to its widespread use in explaining technology acceptance. Moreover, reviewing empirical literature on attitudes towards ML and technology adoption would supplement these theoretical insights. Finally, the involvement of union representatives in the survey development process will ensure relevance and practicality within the specific workplace context, both through feedback on survey items and process optimization. See Appendix 5 for further explanation of how a combined approach of theoretical perspectives (e.g., using TAM), review methods and union representative feedback can be used in survey development.

Take home message for the sustainment and evaluation phase.

During the sustainment and evaluation phase there are two key goals: 1) to provide support of the implemented ML function(s) to sustain and improve employees use of the function(s), and 2) to decide on when is the right time to implement a new ML function. To do this we recommend doing the following:

- Provide ongoing training initiatives to support employees in their use of ML functions. Offer easy access to training resources and materials and send out regular reminders to where employees can find these resources.
- Ensure continuous endorsement from leadership to encourage participation in training sessions and engagement with ML resources. Leadership support creates a culture that values continuous learning and professional development among employees.
- Shift communication strategies towards celebrating successes and fostering a community of practice. Any results from in-house evaluations should be highlighted to clearly communicate the benefits of using ML functions in the evidence synthesis process.
- Provide ongoing support, refresher training sessions, and access to updated resources to re-engage individuals who may have reverted to previous practices and to maintain employees ML knowledge.
- Foster an open-door policy for employees to voice their concerns and questions.
- Assess the readiness of employees for the implementation of new ML functions by involving them in the decision-making process. Use team discussions based on feedback from employees to inform the decision on whether to implement new functions across the entire group and create tailored training materials.
- Monitor Key Performance Indicators (KPIs) to objectively gauge employee readiness and assess the impact of ML implementation on evidence synthesis processes.

- Conduct in-house evaluations of ML functions to document benefits transparently and increase trust and buy-in among colleagues and leadership. Integrate evaluations into ongoing projects or conduct retrospective evaluations on completed reviews to assess functions' usefulness to workflows.
- Develop surveys to assess employees' acceptance and willingness to adopt new ML functions. Utilize theoretical frameworks like the Technology Acceptance Model and involve union representatives in the survey development process to ensure relevance and practicality within the specific workplace context.

Conclusion

In conclusion, this implementation guidance document serves as a toolbox of implementation recommendations derived from our experiences in integrating ML into evidence synthesis processes within HTV at NIPH. Rooted in reflections on both successful implementation efforts and valuable lessons learned, our suggestions aim to provide a practical guidance for institutions or groups seeking to implement ML into evidence synthesis processes within ML-naïve environments.

This implementation guide underscores the importance of a robust pre-implementation phase, emphasizing critical aspects such as defining implementation objectives, assembling a dedicated implementation team, and developing a comprehensive communication plan and an implementation plan. Furthermore, we advocate for the incorporation of theoretical frameworks and models, as well as evaluation mechanisms throughout the implementation process to foster systematic, transparent, and effective implementation of ML into the evidence synthesis process.

During the implementation phase we highlight the significance of tailored communication strategies and stepwise implementation approaches to address diverse user needs and mitigate hesitancies. Also, conducting small-scale pilot projects to evaluate ML functions before implementation in the whole group, as well as larger in-house evaluations, can be very important approaches in encouraging both ML adoption and sustainment.

During the sustainment and evaluation phase, we underscore the importance of continuous training and support initiatives to ensure sustainable implementation efforts. Additionally, we recommend involving employees in decision-making processes, monitoring KPIs, and utilizing surveys to gauge readiness for implementation of new ML functions, enhance acceptance, and inform iterative improvements of the implementation process. Finally, leadership support and buy-in during the whole implementation process and the associated change processes is crucial for successful implementation.

By following the advice provided in this implementation guidance document, you should hopefully have the necessary resources to navigate the ML implementation process with confidence and success.

References

1. Folkehelseinstituttet. Strategi mot 2024 - Om Folkehelseinstituttets strategi – ti satsinger. 2023. Available from: https://www.fhi.no/om/folkehelseinstituttets-strategi/om-strategien/innledning-til-de-ti-satsingene/?term=
2. Borge T, Ames H, Jardim P, Bergsund H, Larsen M, Rose C. Implementation of Machine Learning in Cluster for Reviews and Health Technology Assessments: Results for ML 3.0. Oslo: Norwegian Institute of Public Health; 2024. Available from: https://doi.org/10.13140/RG.2.2.13135.91049
3. Borge T, Ames H, Jardim P, Meneses-Echavez J, Himmels J, Rose C, et al. Implementation of machine learning in evidence syntheses in the Cluster for Reviews and Health Technology Assessments: Final report 2021-2022. Oslo: Norwegian Institute of Public Health; 2023. Available from: https://doi.org/10.13140/RG.2.2.32341.23521
4. Muller A, Ames H, Himmels J, Jardim P, Nguyen L, Rose C, et al. Implementation of machine learning in evidence syntheses in the Cluster for Reviews and Health Technology Assessments: Final report 2020-2021 Oslo: Norwegian Institute of Public Health; 2021. Available from: https://doi.org/10.13140/RG.2.2.32822.42562
5. Borge T, Ames H, Jardim P, Bergsund H, Larsen M, Rose C. Implementation of Machine Learning in Division for Health Services: Strategy Proposal from 2024. Oslo: Norwegian Institutet of Public Health; 2024. Available from: https://doi.org/10.13140/RG.2.2.26557.68322
6. Borge T, Bergsund H, Himmels J, Meneses-Echavez J, Rose C, Muller A. Strategy proposal for further implementation of machine learning in the Cluster of Reviews and Health Technology Assessments. Oslo: Norwegian Institute of Public Health; 2023. Available from: https://doi.org/10.13140/RG.2.2.23220.94086
7. Muller A, Ames H, Himmels J, Jardim P, Nguyen L, Rose C, et al. Aims and strategy for the implementation of machine learning in evidence synthesis in the Cluster for Reviews and Health Technology Assessments for 2021-2022. Oslo: Norwegian Institute of Public Health; 2021. Available from: https://doi.org/10.13140/RG.2.2.16045.20963
8. Borge TC, Muller AE. Overdosevarslingssystemer – en kartleggingsoversikt med maskinlæring. Nordisk Alkohol Nark 2023;40(5):443-62. DOI: https://doi.org/10.1177/14550725221143180
9. Jardim PSJ, Rose CJ, Ames HM, Echavez JFM, 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 Medical Research Methodology 2022;22(1):167. DOI: https://doi.org/10.1186/s12874-022-01649-y
10. Muller AE, Ames HMR, 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. DOI: https://doi.org/10.1002/jrsm.1541

11. Muller A, H. M. R. Ames, T. Borge, C. Hestevik, J. F. Meneses-Echavez JF, and C. J. Rose. A protocol to evaluate unsupervised text clustering to screen and categorize studies in systematic reviews [Preprint]. ResearchSquare 2022. DOI: https://doi.org/10.21203/rs.3.rs-1644531/v1
12. Muller AE, Berg RC, Meneses-Echavez JF, Ames HMR, Borge TC, Jardim PSJ, et al. The effect of machine learning tools for evidence synthesis on resource use and time-to-completion: protocol for a retrospective pilot study. Systematic Reviews 2023;12(1):7. DOI: https://doi.org/10.1186/s13643-023-02171-y
13. Rose CR, M. Bidonde, J. Glanville, J. Berg, R. Cooper, C. Muller, A. Bergsund, H. Meneses Echávez, J. Potrebny, T. Using a large language model (ChatGPT) to assess risk of bias in randomized controlled trials of medical interventions: protocol for a pilot study of interrater agreement with human reviewers [Preprint]. ResearchSquare 2023. DOI: https://doi.org/10.21203/rs.3.rs-3288515/v1
14. Ames HN, Noyes, J., and A. Booth (2023). Chapter 6: Selecting studies and sampling. Draft version. . In: Jane Noyes (Senior Editor) AHSE, editor. Cochrane-Campbell Handbook for Qualitative Evidence Synthesis, Version 10 2023. Available from: https://training.cochrane.org/cochrane-campbell-handbook-qualitative-evidence-synthesis
15. Clark J, McFarlane C, Cleo G, Ishikawa Ramos C, Marshall S. The Impact of Systematic Review Automation Tools on Methodological Quality and Time Taken to Complete Systematic Review Tasks: Case Study. JMIR Med Educ 2021;7(2):e24418. DOI: 10.2196/24418
16. Guo E, Gupta M, Deng J, Park Y-J, Paget M, Naugler C. Automated Paper Screening for Clinical Reviews Using Large Language Models: Data Analysis Study. J Med Internet Res 2024;26:e48996. DOI: 10.2196/48996
17. Yao X, Kumar MV, Su E, Flores Miranda A, Saha A, Sussman J. Evaluating the efficacy of artificial intelligence tools for the automation of systematic reviews in cancer research: A systematic review. Cancer Epidemiology 2024;88:102511. DOI: https://doi.org/10.1016/j.canep.2023.102511
18. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging Research and Practice: Models for Dissemination and Implementation Research. American Journal of Preventive Medicine 2012;43(3):337-50. DOI: https://doi.org/10.1016/j.amepre.2012.05.024
19. Nilsen P. Making sense of implementation theories, models and frameworks. Implementation Science 2015;10(1):53. DOI: 10.1186/s13012-015-0242-0
20. Moullin JC, Dickson KS, Stadnick NA, Rabin B, Aarons GA. Systematic review of the Exploration, Preparation, Implementation, Sustainment (EPIS) framework. Implementation Science 2019;14(1):1. DOI: 10.1186/s13012-018-0842-6
21. Aarons GA, Hurlburt M, Horwitz SM. Advancing a Conceptual Model of Evidence-Based Practice Implementation in Public Service Sectors. Administration and Policy in Mental Health and Mental Health Services Research 2011;38(1):4-23. DOI: 10.1007/s10488-010-0327-7
22. Lyon AR, Cook CR, Duong MT, Nicodimos S, Pullmann MD, Brewer SK, et al. The influence of a blended, theoretically-informed pre-implementation strategy on school-based clinician implementation of an evidence-based trauma intervention. Implementation Science 2019;14(1):54. DOI: 10.1186/s13012-019-0905-3
23. Zhang Y, Cook CR, Azad GF, Larson M, Merle JL, Thayer J, et al. A Pre-Implementation Enhancement Strategy to Increase the Yield of Training and Consultation for School-Based Behavioral Preventive Practices: a Triple-Blind Randomized Controlled Trial. Prevention Science 2023;24(3):552-66. DOI: 10.1007/s11121-022-01464-3
24. Alley ZM, Chapman JE, Schaper H, Saldana L. The relative value of Pre-Implementation stages for successful implementation of evidence-informed

programs. Implementation Science 2023;18(1):30. DOI: 10.1186/s13012-023-01285-0
25. Rogers EM, editor. Rogers EM. Diffusion of innovations. 5th ed. New York: Free Press; 2003.
26. Weiner BJ. A theory of organizational readiness for change. Implementation Science 2009;4(1):67. DOI: 10.1186/1748-5908-4-67
27. Taylor A, Cocklin C, Brown R, Wilson-Evered E. An investigation of champion-driven leadership processes. The Leadership Quarterly 2011;22(2):412-33. DOI: https://doi.org/10.1016/j.leaqua.2011.02.014
28. Vax S, Farkas M, Russinova Z, Mueser KT, Drainoni M-L. Enhancing organizational readiness for implementation: constructing a typology of readiness-development strategies using a modified Delphi process. Implementation Science 2021;16(1):61. DOI: 10.1186/s13012-021-01132-0
29. By RT. Organisational change management: A critical review. Journal of Change Management 2005;5(4):369-80. DOI: 10.1080/14697010500359250
30. Lauer T. Change Management: Fundamentals and Success Factors. 1st Edition 2021 ed. Berlin, Heidelberg: Berlin, Heidelberg: Springer Berlin / Heidelberg; 2020.
31. Statens arbeidsgiverportal [Norwegian Government Employer Portal]. Change Management [In Norwegian][cited 14.03.2024]. Available from: https://arbeidsgiver.dfo.no/omstilling-og-endring/endringsledelse
32. Doran GT. There's a S.M.A.R.T Way to Write Management's Goals and Objectives. Management review (Saranac Lake, New York) 1981;70:35-6.
33. Bloom BS. Taxonomy of educational objectives : the classification of educational goals. New York: McKay; 1956.
34. National Center for Chronic Disease Prevention and Health Promotion Division of Population Health. Understanding the Training of Trainers Model [cited 1.3.2024]. Available from: https://www.cdc.gov/healthyschools/professional_development/documents/17_279600_TrainersModel-FactSheet_v3_508Final.pdf
35. Ettlie JE, Subramaniam M. Changing Strategies and Tactics for New Product Development. Journal of Product Innovation Management 2004;21(2):95-109. DOI: https://doi.org/10.1111/j.0737-6782.2004.00060.x
36. Lewis LK. Organizational change : creating change through strategic communication. 2nd ed. Hoboken, New Jersey: Wiley Blackwell; 2019.
37. Lewis LK, Seibold DR. Reconceptualizing Organizational Change Implementation as a Communication Problem: A Review of Literature and Research Agenda. Annals of the International Communication Association 1998;21(1):93-152. DOI: 10.1080/23808985.1998.11678949
38. Brugha R, Varvasovszky Z. Stakeholder analysis: a review. Health Policy and Planning 2000;15(3):239-46. DOI: 10.1093/heapol/15.3.239
39. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. DOI: 10.4278/0890-1171-12.1.38
40. Sahin I. Detailed review of Rogers' diffusion of innovations theory and educational technology-related studies based on Rogers' theory. The Turkish Online Journal of Educational Technology 2006;5:14-23.
41. KPI.org. What is a Key Performance Indicator (KPI)?[cited 08.02.2024]. Available from: https://www.kpi.org/kpi-basics/
42. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. BMC Psychology 2015;3(1):32. DOI: 10.1186/s40359-015-0089-9
43. Bauer MS, Kirchner J. Implementation science: What is it and why should I care? Psychiatry Research 2020;283:112376. DOI: https://doi.org/10.1016/j.psychres.2019.04.025
44. Eccles MP, Mittman BS. Welcome to Implementation Science. Implementation Science 2006;1(1):1. DOI: 10.1186/1748-5908-1-1

45. Aarons GA, Sommerfeld DH, Hecht DB, Silovsky JF, Chaffin MJ. The impact of evidence-based practice implementation and fidelity monitoring on staff turnover: evidence for a protective effect. J Consult Clin Psychol 2009;77(2):270-80. DOI: 10.1037/a0013223
46. Fixsen DL, Blase KA, Naoom SF, Wallace F. Core Implementation Components. Research on Social Work Practice 2009;19(5):531-40. DOI: 10.1177/1049731509335549
47. Thomas J, Graziosi S, Brunton J, Ghouze Z, O'Driscoll P, Bond M. EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis. London: UCL Social Research Institute: EPPI-Centre Software.; 2020.
48. McGuier EA, Kolko DJ, Stadnick NA, Brookman-Frazee L, Wolk CB, Yuan CT, et al. Advancing research on teams and team effectiveness in implementation science: An application of the Exploration, Preparation, Implementation, Sustainment (EPIS) framework. Implementation Research and Practice 2023;4:26334895231190855. DOI: 10.1177/26334895231190855
49. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-40. DOI: 10.2307/249008
50. Scherer R, Siddiq F, Tondeur J. The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. Computers & Education 2019;128:13-35. DOI: https://doi.org/10.1016/j.compedu.2018.09.009
51. Sohn K, Kwon O. Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products. Telematics and Informatics 2020;47:101324. DOI: https://doi.org/10.1016/j.tele.2019.101324
52. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Universal Access in the Information Society 2015;14(1):81-95. DOI: 10.1007/s10209-014-0348-1
53. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017;19(11):e367. DOI: 10.2196/jmir.8775
54. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425-78. DOI: 10.2307/30036540
55. Hoff KG, Helbæk M. Økonomistyring 1: Grunnleggende bedriftsøkonomisk analyse: Universitetsforl.; 2016.
56. Thomas J, McDonald S, Noel-Storr A, Shemilt I, Elliott J, Mavergames C, et al. Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. Journal of Clinical Epidemiology 2021;133:140-51.
57. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. Research synthesis methods 2018;9(4):602-14.
58. EPPI Centre. Automation tools in EPPI-Reviewer[cited 12.1.2024]. Available from: https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3772
59. Tidemann A, Elster AC. maskinlæring[cited 4 Oktober]. Available from: https://snl.no/maskinl%C3%A6ring
60. OpenAlex. OpenAlex documentation[cited 17.03]. Available from: https://docs.openalex.org/

Appendix 1: Description of the ML team at NIPH

The ML team's activities since initiation have covered:

- **Horizon scanning and innovation**: Identify new ML features or new applications of features, and possibly prioritise for evaluation.
- **Evaluations**: Plan and conduct evaluations of selected ML features with respect to acceptance and workflow changes, and prioritise the most effective ones for implementation. Identify needs for workflow changes that can be achieved with new ML features.
- **Implementation and capacity building**: Improve existing training materials. Improve HTV project managers' capacity to implement ML features. Increase HTV employees' knowledge, acceptance and expertise in ML, via training materials, tutorials and seminars addressing basic concepts of ML
- **Dissemination**: Communicate the results of the team's work to within HTV, as well as outside NIPH. Team members attending conferences and external meetings not only facilitated exposure to diverse perspectives but has also fostered networking opportunities, enabling the team to stay abreast of advancements in the field. This has also been the one activity that has been most effective in communicating our work out to other evidence synthesos groups, oth nationally and internationally
- **Collaboration**: liaise with librarians' ML activities as well as collaboration with other groups or teams outside the institute that conduct relevant ML or automation activities

The team has since initiation had the following goals:

- to test and document pros and cons of using ML in various phases of the evidence syntheses process.
- to build employees’ competence in using ML
- to contribute to ML being used in HTV's evidence synthesis products,
- to identify and evaluate new ML functions.
- to focus on the further implementation of ML and capacity building activities

ML team roles – define during pre-implementation phase.

Implementation-lead: Responsible for coordinating implementation activities. Collaborates with ML team lead in facilitating other key activities, when appropriate.

ML team lead (if the scope of the team goes beyond implementation alone)**:** Supports implementation-lead to match implementation activities with e.g., evaluations.

Responsible for coordinating other team key activities, like evaluations and horizon scanning.

ML Mentor: The advisor should have advanced knowledge of ML and contribute to strategy planning, discussions, evaluations, and questions from the team. We advise the team lead(s) and mentor to have regular discussion meetings, especially at the beginning of the implementation phase.

ML team members:

A key responsibility for other team members will be to be an ML contact assigned to a project team conducting an evidence synthesis. This should be a rotating role depending on availability and interest, and all team members can be an ML contact. The ML contacts main task is to provide ML training to the evidence synthesis team, which involves advising the team on the use of recommended ML and corresponding workflow optimization. The ML contact coordinates with the ML lead to approve ML language in protocol and final report.

In the case when ML is not being used or is not being used in the recommended manner (which typically means without workflow changes), the ML contact must understand why, as soon as possible, for example by meeting with the project leader. This process should begin with discussion within the ML team for advice. If the situation is not resolved, the ML contact must clearly communicate concerns in writing to the project leader, including a brief analysis of the situation, consequences of not using recommended ML and further recommendations for how to move forward.

Competency requirements

The ML team members should possess several key competency requirements crucial for successful implementation:

- A basic knowledge of implementation theory will provide a foundational understanding of the processes and strategies involved in effectively integrating ML into evidence synthesis practices.
- The willingness to quickly learn a new field is vital in the rapidly evolving landscape of ML. The ability to adapt and acquire knowledge swiftly ensures that team members stay abreast of the latest advancements, fostering innovation and maintaining the team's competitiveness in the field.
- The capacity to absorb and comprehend technical information, coupled with the ability to apply it in real-life scenarios. This competency enables team members to translate theoretical knowledge into practical solutions, ensuring the successful implementation of ML functionalities in evidence synthesis processes.
- Possessing the ability to work effectively with individuals who may be hesitant or sceptical about adopting ML is crucial for fostering a positive and collaborative team environment. This quality promotes open communication and helps in overcoming resistance, facilitating a smoother implementation process.
- Possessing good training and facilitation skills is essential in conveying complex ML concepts to diverse audiences.

- Team members should have skills enabling them to create structured and effective training materials, ensuring that the learning resources support the successful implementation of ML practices.

We also recommend that the team composition is interdciplinary in nature. Our experience is that the interdisciplinary nature of the team has been a key success factor, with the inclusion of a librarian proving particularly valuable. Consistent weekly meetings have played a pivotal role in maintaining team cohesion, preventing potential drift. Furthermore, the team's intrinsic motivation for ML has been a driving force, propelling the group forward.

The ML team in the sustainment /evaluation phase

Onboarding new members

Over time, the ML team members will change as the team expands or as members are moved to other tasks. The onboarding of new members should be formally planned for. The following is a description of our new member onboarding process.

New members follow a syllabus of 8-10 required resources and numerous optional resources, organized by topic. Progressing through this syllabus will be the first task of the new member, and they will work with the ML lead to determine the pace of this learning and how learning will be monitored/ensured. There is a lot to learn, so the new member needs to feel that they have the time to learn, access to the materials they need, and the support to embed this learning. To as great an extent possible, expert ML members will monitor learning of specific topics, such as by calling the new member in to a topic-specific meeting. For example, the new member can focus on a different ML function or tool each week. In this case, the relevant function expert becomes their mentor for that week. It is the new member and the expert member's responsibility to have read the relevant syllabus material thoroughly; it is not the expert member's responsibility to teach, rather to be able to answer questions. This is to ensure a breadth of expertise within the team.

Each member that joins the team receives training on established functions for them to be able to support future project teams. This training follows these four steps:

1. The new team member becomes familiar with the training materials/software on their own, including practicing independently at least twice and according to any assigned tasks in Teams, and prepares any questions they may have.
2. An expert ML team member uses the training materials to train the new member, potentially within an ongoing project.
3. The new member practices until they demonstrate to the team lead that they can train someone else.
4. Quality assurance: The ML lead or co-lead must approve the new member's training capability because the new member should be able to train others.

Appendix 2: User profiles

In developing the different user profiles, we used our experience from the implementation of ML at our institute, which we initiated in 2020. The user profiles reflect the heterogeneity of our target audience for our e-learning course, both when it comes to previous experience with ML, their openness to change as well as their learning preferences.

These user profiles were developed later in the implementation process so are focused on level of familiarity with ML. When developing profiles during the pre-implementation phase you may want to focus on technical competency, openness to change and learning preferences.

The ML naïve user

Name: **Ann Sophie Nguyen**

Short bio: A newly employed PHD graduate working with evidence synthesis. Limited experience with machine learning functions (ML).

Demographics:
Age: 42
Job: Researcher
Workplace: Oslo, Norway
Time with NIPH: 1.5 years

Goals and aspirations

- Wants to be able to produce more research in a shorter time
- Eager to stop doing repetitive tasks that could be done by a machine
- Wants to gain foundational knowledge and skills in the use of ML

Technological considerations

Uses a laptop to complete most of her working tasks. She is very familiar with Microsoft office tools and uses them on a regular basis. Ann Sophie is only familiar with basic screening tools suitable for new beginners and is not familiar with software programs that use machine learning.

NIPH

Barriers to learning:

- Technical competence
- Time available for learning
- Current user guides to machine learning are overwhelming

Digital learning experience:

- No experience with digital learning resources
- Positive expectations

Learning preferences

Needs a learner-centred approach with heavy support and guidance at the early stages. Goals and communication needs to be clear and in plain language. Prefers interactive learning with practical exercises

Photo: Stockphoto

The ML familiar user

Name: Kevin Smith

Short bio: Masters' degree in psychology, interested in methods innovation. He has some basic understanding of ML but needs more training to increase confidence in his own capabilities and to become independent in the use of ML.

Demographics:

Age: 37
Job: Researcher
Workplace: Oslo, Norway
Time at NIPH: 4 years

Photo: Colourbox

NIPH

Goals and aspirations

- Self-motivated to work more efficiently
- To increase skills in the use of ML functions to become more independent
- To increase confidence in his ML skills
- Long term goal is to become independent in use of ML functions
- Working more efficiently

Technological considerations

He is very familiar with basic Microsoft office tools and use these on a daily basis in his work. When introduced to a new tool, he usually needs extensive one-to-one support to get started, and often use quite some time to get familiar with the tool. Due to fire wall restrictions, there are a limited number of desktop tools that he can download on his work computer.

Barriers to learning:

- Struggling with software interface
- Lacks trust in own ML competence
- Some conceptual misunderstandings of how the functions works

Digital learning experience:

- Enjoys the flexibility of digital learning,
- Prefers a mixed learning experience and learning environments with an instructor for closer follow-up

Learning preferences

- Prefers self-paced e-learning courses that allow him to review content as needed
- Prefers courses where you can receive peer support and feedback

Photo: Colourbox

The ML advanced

Name: **Laura Di Guilimi**

Short bio: PhD in social welfare and has extensive experience with ML, is confident in her own abilities and is able to use it independently. She believes that ML can be a powerful tool when used appropriately.

Demographics:
Age: 36
Job: Researcher
Workplace: Oslo, Norway
Time at NIPH: 8 years

Goals and aspirations

- She is highly motivated and is eager to use ML in new and exciting ways in the evidence synthesis process.
- To advance her knowledge and expertise in ML
- To stay up-to-date with the latest best practices and trends in the use of ML in evidence synthesis
- To contribute to the field through research and innovation

Technological considerations

Laura uses many different digital tools on a daily basis and is also very active on social media. She is confident in her own technological abilities and learns to work with a new tool quickly. Due to fire wall restrictions, there are a limited number of desktop tools that she can download on her work computer.

NIPH

Barriers to learning:

She may be critical of potential biases involved in using ML

Digital learning experience:

She enjoys the flexibility digital e-learning provides

Learning preferences

She prefers e-learning courses that:
- provide in-depth insights into specific techniques or applications.
- offer personalized content and opportunities for peer-to-peer collaboration.

Photo: Colourbox

The Sceptic

Name: **Solveig Sinsen**

Short bio: An older researcher who has been working with the gold standard of evidence synthesis for over 20 years. Well respected in her institution, with a long list of high-quality publications

Demographics:
A researcher who lacks trust in process changes and the use of ML. She prefers things to remain the same. She is skeptical of the effectiveness and reliability of new technologies.

Goals and aspirations

- To maintain high quality evidence syntehsis using the gold standard
- To gain a better understanding of the strengths and weaknesses of machine learning functions for evidence synthesis, and to make informed decisions about whether or not to use them.
- Feels pressure to use the new tools as they are recommended by leadership so wants to understand them and make her own decision

Technological considerations

- Solveig uses some digital technology as part of her daily work but is not an advanced user.
- It takes time for her to feel comfortable using a new program and she requires one-to-one support to get started and further use.

NIPH

Barriers to learning:

- Possibly close to retirement
- A traditionalist with a preference for maintaining the status quo

Digital learning experience:

- Preference for traditional methods of learning - face to face
- Finds digital learning to be non-interactive and boring

Learning preferences

- Solveig has a preference for interactive learning that is based in two way communication where she receives a lot of feedback.
- Learning need to be simple to use and personalized to her level.

Photo: Colourbox

Appendix 3: Addressing hesitancy using the transtheoretical model.

Effectively communicating with employees hesitant to adopt new technology, such as ML functions, demands a thoughtful and empathetic approach. Using the Transtheoretical model (39), here are some suggestions for how you can communicate with hesitant employees, depending on where they are in the stages of change.

In the **precontemplation** stage, where individuals may not be ready for change, communication strategies involve raising awareness through e.g., newsletters, emails, or informative sessions about the potential benefits of integrating ML into the evidence synthesis process. Moving to the **contemplation** stage, where employees are considering change, it becomes crucial to actively listen to their concerns, empathize with their reservations, and provide clear, concise, and accessible information about the ML functions. Communication can focus on sharing case studies or success stories within the organization, emphasizing the positive impact of ML on workflow efficiency. In the **preparation** stage, as employees get ready for change, effective communication includes highlighting available resources, training materials, and support for those willing to use ML functions in the evidence synthesis process. As employees transition to the **action** stage, making the change, communication efforts should recognize and celebrate early adopters, showcasing their achievements with using ML in the evidence synthesis process. Furthermore, tailoring communication to address specific concerns, offering benefits-oriented messaging, and assuring a low threshold for addressing questions and concerns are essential strategies to guide employees through the **preparation** and **action** stages. In the **maintenance** stage, focused on sustaining the change, communication strategies entail regularly sharing success stories, fostering a community of practice, and acknowledging small wins to reinforce the positive impact of using ML in the evidence synthesis processes. The **termination** stage represents a state where employees have fully integrated the use of ML into their evidence synthesis practices and will not return to their previous practices. At this stage, using ML functions should be considered a standard part of evidence synthesis workflows within the organization or group. Communication strategies at this stage might involve celebrating long-term success, acknowledging the established use of ML, and encouraging continuous improvement through on-going learning and adaptation to new advancements within the ML technology field. Finally, the model acknowledges that **relapse** can occur during any stage. To prevent a return to previous practices, communication strategies involves providing ongoing support and reminders of the benefits of using ML functions.

Appendix 4: Suggestions for KPIs

Below are some suggestions for specific KPIs. These KPIs can serve as valuable benchmarks for assessing the success of ML implementation in the evidence synthesis process. We suggest dividing the KPIs on two levels: KPIs specific to the implementation process and more overarching KPIs at the institute level.

For the implementation process we suggest the following possible KPIs:

- Increase in the number of employees that use at least one ML function when conducting a review. This can be measured by looking at the reporting of the use of ML in the ML Appendix section of each report. Our reporting template can be found in Appendix 10 in this report.
- That the ML functions that have been implemented are being used correctly. This can be measured by investigating the text in the Appendix reporting ML use in each evidence synthesis product, through one-to-one user support and through evaluating random projects underway via check-in meetings.
- Regularly measured satisfaction scores or feedback from employees regarding their experience with ML functions. This provides insights into the user-friendliness and overall satisfaction with the implemented ML function(s) and should inform further implementation processes.

At the organizational level, we recommend the following KPIs:

- Reduction in the time required to complete evidence synthesis projects. This can be measured by comparing commission timelines from previous years, starting from protocol approval or literature search completion to completion of projects.
- Evaluate productivity by comparing ratio of publications to employees before and after implementation.
- Evaluate the institute's adaptability to workflow changes related to ML implementation, by assessing the adoption of agile project approaches in teams.
- Increase in the number of publications produced after ML implementation. This correlates ML impact with the team's productivity and output.
- Percentage reduction in time spent on specific evidence synthesis tasks (e.g., screening of references, data extraction) with the introduction of ML. This can be done by logging time spent on the different phases of the review. This will demonstrate the efficiency gains achieved through ML, providing a concrete measure of time saved.

- Reduction in human resources required for evidence synthesis tasks, after ML implementation, i.e., fewer employees needed per project. This will provide an indication of the optimization of human resources achieved through ML integration.

Appendix 5: Survey development

The following is an example of how a blend of evidence synthesis methods and theoretical understanding may be utilized to develop a survey to assess attitudes towards AI tools in the workplace.

Selecting a theory

In order to better understand how a novel technology is received and approached in the workplace, it is useful to apply an appropriate theoretical framework. For the following example, we selected the Technology Acceptance Model (TAM) which has previously been used to explain what causes people to reject or accept information technology (49).

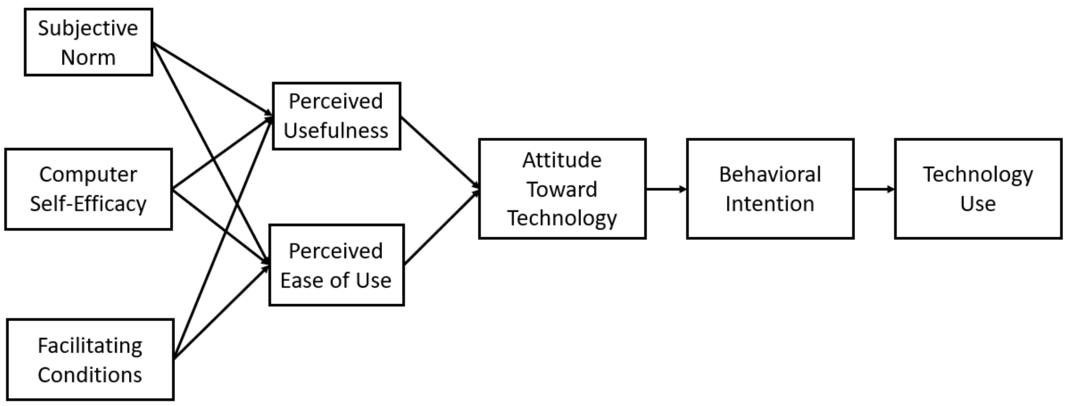

Figure 8. The TAM-2 framework (adapted from the Figure by Scherer et al (50)).

The TAM is one of the most widely used models to explain use of innovative technology and has been applied in a wide range of fields, such as intelligent healthcare systems, learning environments and advertising systems (51). It also has empirical support from several studies, demonstrating its usability for explaining technology acceptance and use (50). The core variables are Perceived Usefulness, Perceived Ease of Use and Attitudes Toward Technology (see Figure 8). Perceived Usefulness refers to the belief that using the new system or program would enhance one's job performance and Perceived Ease of Use refers to the belief that using the new system of program is free from difficulty or great effort. These variables are expected to directly explain one's evaluation of the new system or program, i.e., their Attitudes Toward Technology. It is further assumed that a positive attitude will be directly linked to a person's intention to use that technology (Behavioural Intention), which in turn will predict actual use of said technology (Technology Use). In addition, three external variables are often used explain variance in the motivation variables: Subjective Norm, Computer Self-Efficacy and Facilitating Conditions (52).

The variables outlined in TAM may be applied to design a survey to assess workers' attitudes towards ML. Given that a great number of studies support the applicability of TAM in predicting intentions to use technology, it may be highly useful in identifying areas which require more attention and enhanced implementation efforts. However, other theories may be equally suitable to fulfil the same purpose. In such cases, we recommend selecting theories that a) address technology specifically, and b) have a solid empirical basis in predicting technology use or acceptance. Theories that may be considered include Non-adoption, Abandonment, Scale-up, Spread and Sustainability (NASSS) (53) and the Unified Theory of Acceptance and Use of Technology (UTAUT) (54).

Review of empirical literature

In addition to letting a theoretical framework guide the survey development process, we suggest supplementing this knowledge with empirical literature. The field of technology adoption is constantly evolving, and so recent findings are useful in strengthening the relevancy of the survey questions. Ideally, the studies should be as close to the context in which the final survey is going to be applied as possible. For instance, if the survey is going to be used to assess information specialists' attitudes towards a novel ML-based search function, the selected studies should preferably cover similar professional groups and technologies. However, this may result in a very limited selection of studies, as this field is not rife with empirical research. In such cases, we recommend searching for studies that are as close in context as possible, while simultaneously ensuring a sufficient level of relevant research to inform the survey development. For instance, searching for studies on ML attitudes among scientific professions in general may provide more relevant studies.

We recommend exploring both qualitative and quantitative studies that address attitudes towards introducing new technology. This will allow for more studies to be included and thus providing a richer foundation for the survey. More importantly, however, they may be able to support the survey development process in different ways. Qualitative studies can provide an overview of issues that professionals regard as salient. These issues may overlap with existing domains from the theoretical framework, or they may offer insight into new areas that may be useful to explore in a survey. Quantitative studies, on the other hand, may be less useful in identifying new domains, but can be helpful in identifying other questionnaires or items that have proven to have high psychometric quality. Assuming that these questionnaires are open and free to adapt, they may be used, either in part or in full, in the survey.

Involving union representatives

After the selection of a theoretical framework and selection of studies to serve as the foundation for the survey development, we recommend testing the early version of the survey with colleagues who are similar to the ones who are going to respond to the final survey. Our recommendation is to use union representatives, as these may not only provide feedback on the survey items, but also recommend how the survey process should be carried out in a way that is ideally suited to that particular workplace (e.g., what would be an ideal length of the survey to ensure that respondents have time to complete it?).

Appendix 6: Examples of training activities

We developed stand-alone training materials for project leaders and/or project members. These materials encouraged employees to begin ML implementation independently of the ML team. The training materials were successful in supporting project leaders to implement ML functions more independently, and thereby reducing technical assistance needs from the ML team. This change in procedure changed how ML assistance was given. Depending on the nature of the request, there were two potential assistance methods: a) If the issue could be addressed through existing training materials, a member of the ML team sent these materials to the project team lead and concluded the process, or b) If technical assistance was required, an ML contact were assigned to the project team in the ML team meeting where the help request was discussed.

Approaches for doing a large-scale training for employees: ML week and e-learning courses.

For large scale training we have used two different approaches: synchronous learning and blended learning.

In the synchronous approach we delivered training sessions covering both conceptual understanding and practical application of the ML functions we use in our group. This was in the shape of an "ML week", given about once a year to all employees, consisting of both traditional "classroom lectures" with discussions as well as workshops. This was catered to both ML naïve and new employees as well as employees that are comfortable with the use of ML. The ML week required pre-registration and for some sessions there where some practical exercises that had to be completed ahead of the sessions. The training materials used during the ML weeks were tailored to meet the specific needs of the audience and tailored to different difficulty levels. Sessions that required employees to be intermediate or advanced where clearly announced as this. Prior to the practical sessions, learners were provided with a few resources to prepare. In the lessons the learners first got a conceptual understanding of the algorithms to the ML function follow-up by a practical session where the employees learned how to use it with guidance from an instructor. In the practical sessions the training was simplified with step-by-step approach and also used examples from concrete projects to engage the employees and also create relatability with the conceptual training. This allowed the instructor to break down complex concepts, making it easier for employees to grasp and apply the information. The importance of participation in the ML week was emphasized by leadership which might be one of the reasons for the high participation rate.

In the blended approach we made short e-learning courses on our most used functions, focusing on how the functions work to give the employees an increased understanding of the theoretical background of the functions. The employees were required to complete the e-learning before they could participate in later in-person training sessions. The e-learning is not software specific and can therefore be used independently of review tools. To make the e-learning as accessible as possible, plain language was used. The e-learning received great feedback from employees.

The e-learning was sent out a month before the in-person training and the participate could complete the training in their own pace at the time they wanted. The e-learning consisted of short modules (5-12 minutes) with theoretical information about the functions and interactive learning exercises to test the learner's knowledge and create engagement. The modules also had clear learning objectives and had a mix of text, videos, and illustrations. The following in-person sessions did not contain any training on conceptual knowledge of the function; it was focused on how to use the ML functions in practice.

Approaches for training the "ML sceptics" or those who are hesitant

Individuals who initially hesitated to adopt ML or held scepticism toward its usage underwent the same training as other employees. Moreover, a subgroup received additional one-on-one instruction from a member of the ML team. During the ML week, dedicated sessions were conducted to address open questions, where we clarified and expanded on concepts that was understood as potential causes of the scepticism.

Appendix 7: Working plan for developing or curating training materials.

Step	Procedure	Quality assurance, before subsequent step
1.	The ML lead identifies a ML member to "own" this process, per function. This ML member will likely be one involved in the evaluation of this function	
2.	The team discusses how to add the function to the flow chart overview	
3.	The ML team member in coordination with the implementation lead, pulls together all existing training material (internal and external) and determines if in-house training materials need to be created, if existing training materials are sufficient, or if existing materials can be tailored to in-house needs.	Implementation lead involved in determining adequacy of existing material
4.	If new materials are needed, a preliminary presentation is developed including a description of the underlying ML functions, using their notes from the evaluation stage, the ML team's literature collection, and any other resources necessary. Training materials must first describe "the conceptual", and then explain "the practical". As much as possible, slides are re-used from previous functions to maintain continuity.	ML lead reviews the conceptual part of training materials
5.	The draft training materials are piloted within the ML team and feedback is given.	ML implementation lead provides ultimate guidance on direction of draft training material
6.	The ML team member and implementation lead adjust training materials based on the ML team feedback. If adjustments were minor, training materials are sent to quality assurance, then piloted within an ongoing project. If the adjustments were major then the training materials are piloted to the ML team again, then sent to quality assurance.	ML lead reviews the conceptual part of training materials. EPPI superuser reviews practical part.
7.	The training is piloted in a hands-on support manner within an ongoing project. Feedback is given during and after.	
8.	The training materials are adjusted based on the feedback. If minor, move to step 8, if major, repeat step 6 with a new project team.	
9.	The training materials are handed off to an ongoing project team to implement independently. The ML team member can be present and receive feedback throughout the implementation about the training materials but will avoid giving direct technical assistance (and thereby	

	undermining the independence that the training materials should facilitate). Feedback is given during and after.	
10.	Training materials are adjusted based on the feedback. Guidance for when to contact the ML team is always included. If minor feedback, move to step 10 if major repeat step 8.	
11.	The final presentation is given to members of the ML team and sent through quality assurance.	ML lead reviews the conceptual part of training material. EPPI superuser reviews practical part. Implementation lead approves entire training material.
12.	Training materials are shared on the ML's SharePoint site for project teams to implement the function independently. Project teams can request help from the ML team if they are stuck.	

Appendix 8: Support suggestions for sustainment of the implemented function

Below we describe the different forms of user support provided by the ML team to employees, either personally or based on training material and -resources developed by the team. With all types of implementation efforts, continuous endorsement from leadership on the importance to participate in training sessions and engage with other learning resources is central. This support creates a culture that encourages continuous learning and professional development among the employees.

Types of support provided to project teams can include one-to-one support to project teams, easy access, e.g., via an internal website, to in-house developed training materials and other relevant resources.

One-to-one support
To support project leaders with the implementation of new ML functions, we provided one-on-one training and technical assistance. Each project lodges a support request form specifying the details of the project. The request is then added to the agenda for the upcoming team meeting, where it undergoes thorough discussion. Subsequently, each project received a dedicated ML team member who trained the project leader first, and then the rest of the team, and was available for immediate assistance when needed. A follow-up plan with timelines, action points, and communication methods is agreed upon. The process extends to ongoing support, quality-control of language in protocol and final report, training sessions, and continuous updates until project completion. However, we found that sole one on-one trainings were not sufficient for immediate method independence of project leaders and members to apply newly learned ML functions. To address this, additional training materials were developed.

Training materials
Based on our experiences, we have found that training materials developed in-house have proven to be highly popular among employees. This success can be attributed to the targeted approach that aligns with the specific needs of our workforce. Additionally, our employees have expressed a clear preference for customized training methods that are tailored to their daily work. They appreciate learning through simplified examples that relate to their responsibilities. Furthermore, workshops incorporating a step-by-step approach, actively involving participants, have received more positive feedback compared to traditional classroom-based training methods. Employees favour a hands-on learning experience over passive observation of

presentations, especially when dealing with the setup or utilization of ML functions in the data management tool. Also, short e-learning courses developed by the team have gained significant approval by the employees. The flexibility of these courses, allowing employees to complete them at their own pace and during breaks between tasks, has been very well-received.

Reporting templates

The ML team developed reporting templates that are available to all employees and whose implementation is strongly encouraged. The development of the templates was based on two main goals: 1) To ensure a standardized way of reporting the use of ML in our reports and 2) to encourage employees that use ML to reflect on the reasons for why they use ML. The reporting template should be incorporated as early in the project phase as possible, ideally in the protocol development phase. We have provided an English version of our reporting templates in Appendix 10.

Website with relevant resources

An ML SharePoint room available to all employees, as well as interested external parties, was created with the aim to be a “one stop shop” for HTVs ML resources (see Figure 9 for front page of our SharePoint room). This contains all written material created on each ML function as well as a syllabus with key papers relating to the use of ML in evidence synthesis, recordings from previous training sessions and link to e-learning courses. It also provides links to relevant external resources like review tool resources (user guides, tutorials, YouTube videos, SR Toolbox).

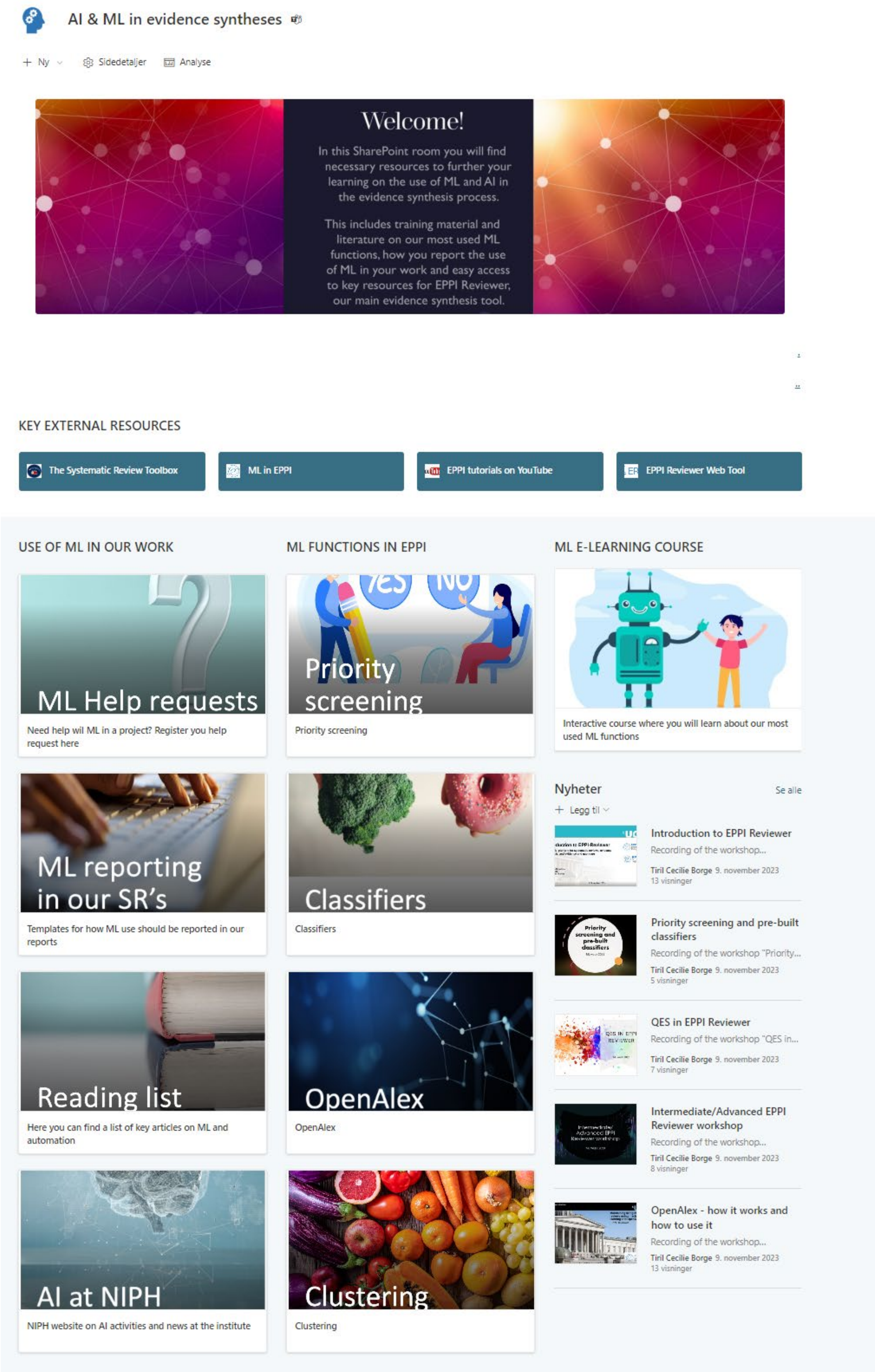

Figure 9: Illustration of the welcome page of the HTV ML teams SharePoint room

Appendix 9: Evaluation approach and template

Here is a general approach to what we have found to be important to consider when doing in-house evaluations:

1. Identify what you want to evaluate, and provide background information on this, including why it should be evaluated. One ML function that we have evaluated are custom classifiers. Custom classifiers can be used during the title and abstract screening phase and is a supervised form of machine learning. Here you build a model based on your decisions on for example inclusion and exclusion of a set of studies at title and abstract level, and the model will then predict the probability of fitting your inclusion criteria for any new studies fed into the model.
2. Determine the metrics to be measured—whether performance-related (time and resource use), accuracy-related (recall and precision), or qualitative outcomes (acceptance and feasibility). Our focus has primarily been on measuring time use and acceptability, aligning with the mandates from leadership.
3. Then you need to decide on what type of product you want to evaluate the ML function on. Do you want to assess performance of one ML function across distinct types of reviews to find which review-types the function works best on? During the first year of the ML team, nine projects contributed to the evaluation of custom classifiers for screening, where three were prospective and six were retrospective (an update of a covid-19 rapid review, one EUnetHTA rolling collaborative review and two updates, three scoping reviews, three reviews of RCTs/cohort studies, and one overview of reviews). Some results were that we found that between 18-90% fewer studies can be screened at title and abstract level and that Auto-screening studies the custom classifier predicted to be less than 10% relevant saved 48 hours, which corresponded to 36% of total screening time, with complete accuracy (4).
4. Consider the composition of the ideal review team based on the complexity of the ML function and the team's familiarity with ML. In our case, embedding an ML team member into the review team was crucial, and team-wide acceptance of ML integration was paramount. If this is challenging, retrospective evaluations can be an alternative option.
5. Integral to these evaluations is identifying potential workflow changes to optimize the benefits of ML integration. For example, with custom classifiers, you can indicate any changes to screening practice, for example prioritize/deprioritize screening according to the custom classifiers predicted relevance of fitting your inclusion criteria, and you can start screening on full text level in parallel.

Machine learning evaluation template

Written by: / Peer-reviewed by:

Status:

Title:

Machine learning (ML) function:

Systematic review phase:

1. Background
    1.1. ML function in plain language
    1.2. Previous evaluations
    1.3. Need for this evaluation

2. Methods
    2.1. Research aims:
    2.2. Evaluation requirements
        2.2.1. Review type:
        2.2.2. Other review characteristics:
        2.2.3. Review team characteristics:
        2.2.4. Support from ML team:
            a. One option is minimal support. The ML team can .... independent from the review team.
            b. If this evaluation is instead used to train a review team, then one ML team member will spend approximately ... sessions to train the review team leader and any other members interested.
        2.2.5. Overlap with other ML evaluations/learning:
    2.3. Data collection

Procedures described here are to train the review team leader, but all steps can also be conducted by the ML team if the minimal-involvement option 2.2.4 (a) is chosen.

2.4. Outcomes and analytic plan

Outcome 1:

Outcome 2:

3. Use of evaluation results in ML team
4. Suggested evaluation administration

ML evaluation lead:

Reviewers also on ML team:

Management contact:

Potential partners or sources of technical guidance:

Appendix 10: Reporting templates

The following text is included in the reporting template developed by the ML team at NIPH and can be used in its entirety or as inspiration by other groups or institutes that want to implement ML functions into the evidence synthesis process:

This Appendix describes how we will use/have used machine learning. At the end of the Appendix, there is an explanation of the terms we have used for the various ML functions that will be used/have been used in this template. In this Appendix, we present the most common ML functions used in the development of evidence synthesis products at [insert institute/group name]. This is not an exhaustive list of possible functions: it is solely based on what we use.

- Remember that each project is unique and that judgements about which ML features are suitable will vary.
- For help on how to use the different ML functions, see the help material in our SharePoint site first. If you can't find an answer, talk to your ML contact, or contact the ML team.
- Text boxes in blue are for guidance only and should not be included in the actual project plan/final report.
- Remember to merge the individual step tables into a single table at the end.
- It is possible to subtract or add steps and functions as needed.
- At the end of the appendix there is a list with explanations of the terms we have used for the different ML functions. Only include the terms you use in the reporting.
- This appendix is written for a project plan, remember to change the verb tense in the final report.
- Remember that this is a dynamic process. We therefore recommend that you start reading full texts in parallel with assessing the title and abstracts and remember to resolve any conflicts frequently along the way.
- This symbol indicates areas where the team needs to think through options and make decisions for using ML in the project.

Procedure

Suggested main text for project plan and report: **Project plan:** In the process of selecting references, we plan to use the following machine learning (ML) functions in the EPPI Reviewer software (47) [possibly other software]: [insert name of features]. See the procedure for using ML in Appendix [x]. **Final report:** In the process of selecting references, we used the following machine learning (ML) features in the EPPI Reviewer software (47) [possibly other software]: [insert name of functions]. For a description of the procedure, see Appendix [x].

Step 1:
In this step, you should think about the functions you will use before you start to screen the title and abstracts with priority screening. Common functions to use in this step are the Cochrane RCT classifier*, systematic review classifier, automatic text clustering and OpenAlex.

For example:
- To update a review and identify new relevant references, it may be appropriate to use OpenAlex to search for references based on the studies you have included in the previous review (i.e., your seed studies). In the final report, the number of references used, parameters, search date and number of hits from OpenAlex should be reported, in addition to the references used as seed studies. You can also use OpenAlex if you are writing a Single Technology Assessment to check if there are studies that have been omitted from the documentation packages.

- If you are looking for specific study designs, you can run one of the study design classifiers, either to exclude references that do not have the desired study design(s) early in the process or to prioritise the selection of references with your desired study design(s). When using such a classifier, clearly report which threshold values you use to exclude references without manual review, or to prioritise a specific group of references in the priority screening and whether you have used single screening on certain groups of references.

- To try to identify groups of relevant references that fulfil the inclusion criteria, you can test automatic text clustering. If you identify groups that are relevant, these can be quickly assessed for inclusion. The following parameters should be reported when using the function: Maximum hierarchy depth, maximum cluster size, maximum label length, minimum cluster size, single word label weight. Remember to write the cluster name and the number of references that were assessed by one person or excluded without manual review.

*Only the Cochrane RCT classifier has been tested and validated.

Step	Description of approach
Step 1	*Describe the ML functions you will use <u>before</u> starting to manually screen references with priority screening.* *Example text:* - We will use automatic text clustering to identify relevant groups of references that possibly fulfil the inclusion criteria to screen these first. - We will use the Cochrane RCT classifier, as recommended by Cochrane, to quickly identify references that fulfil/do not fulfil our inclusion criteria for study design. - We will use the systematic review classifier to quickly identify potential systematic reviews. - Any other ML features

Step 2:

This step will usually describe the use of priority screening and when in the process you will consider changing from two to one person screening references, stop screening references or switch to using other ML functions. What determines this could for example be that a) a certain number of references have been screened without including a single reference (e.g., 100) or b) you have screened for a certain amount of time without including any references (e.g., one hour). When using priority screening, clearly report when you changed the screening procedure from double to single screening and/or what you used as basis to exclude references without manual review or prioritise a specific group of references in the priority screening.

Step	Description of approach
Step 2	*Describe the ML function(s) you will use to select (screen) references and what you will use as a decision basis to switch from two to one person, stop selecting or use other ML functions. Example text:* "To more quickly identify references that fulfil the inclusion criteria in the process of assessing titles and abstracts, we will use priority screening." Examples of decision basis for changing screening practices: - At an inclusion rate of ≤ x% of the last x references read - After assessing x references without including one - After x number of minutes of assessing references without finding one relevant study -Describe any other solution that was used

Step 3:
In this step, you describe which ML functions you will use if you have many remaining references after having screened references at title and abstract level for a while. The purpose of this step is to see if there are more relevant references among the remaining references. Common functions to use in this step are automatic text clustering (see reporting step 1) and custom classifiers*.

When using custom classifiers, we recommend that you have included some references based on full text review. Custom classifiers can also be based on assessments of the title and abstract if it is a project where full texts have not been read.

In the final report, state the threshold value for which references were only screened by one person or excluded without manual review. Example text:
"References that were predicted to have (above/below) x% probability of meeting the inclusion criteria:
-- were reviewed by only one project member
-- were automatically excluded without manual assessment**"
-- describe other method

* It is not recommended to use custom classifier on your own if you are not familiar with the function and ML.
** This should only be done if the custom classifier has been tested on sufficient references already included at title and abstract level

Step	Description of approach
Step 3	*Describe ML functions that are used, for example, if you have been screening for a while and do not see signs that you have found most of the references that fulfil the inclusion criteria and there are still many references left. Examples:* - We will use automatic text clustering to identify obviously irrelevant/relevant groups of references - We will build and test a custom classifier in collaboration with the ML team once we have included [e.g., 15-25] references in full text. If the testing of the model is deemed satisfactory, we will use the model and report how it is used and what threshold values were used for changing screening practices (e.g., from double to single screening or excluding references without manual review).

Step 4:
The purpose of this step is to identify relevant references that are not captured in the database searches using already included references*. This step can be used to supplement or replace grey literature searches and/or searches in individual databases. In the final report, the number of references used to run OpenAlex, parameters, search date, any filters, and the number of hits from OpenAlex must be reported. Also refer to the references used as seed studies for the search.

For the most comprehensive search mode, choose "bi-directional citation AND recommendations ('bi-citation AND recommendations')".

Describe how you proceed when you screen the references identified by OpenAlex. Should all references be assessed or only selected ones? Do two or one person assess? Also describe any use of priority screening or other relevant ML functions you will use to assess these references.

*For updates of previous reviews or quality control of documentation packages for single technology assessments, this can also be step 1 in the approach

Step	Description of approach
Step 4	*Describe any use of OpenAlex to identify additional relevant references that are similar to references you have already included.* *Example text:* - We will use OpenAlex to identify references using references we have included at full text level. We will review references from OpenAlex [describe procedure].

Reflexivity

[If the final report has a separate section on reflexivity, reflexivity regarding ML should be included there. If not, a small paragraph should be written here]

ML reflexivity is about how the decisions you make affect ML, as the algorithms learn from your decisions. For example, if you only manually review 25 per cent of the references and exclude the remaining 75 per cent based on the machine's assessments, any bias from the manual review process will affect how the machine has assessed the remaining 75 per cent of the references. It is therefore not necessarily about the number of studies that are manually assessed, but how precise and coordinated the team members' assessments are in relation to understanding the inclusion and exclusion criteria. It is important to recognise and reflect on how team members' preconceptions of the topic may affect the process of selecting references. For example, a doctor will probably wear different "glasses" than a sociologist, and previous experience in relation to the topic, both on a professional and a personal level, may also affect the choices made.

Awareness and open discussion about different understandings of the problem/theme in the project group will be important in order to avoid this potential bias, which can influence how well the ML functions work.

Definition of ML terms

Algorithm can be explained as a complete, precise and step-by-step description of a procedure of operations intended to solve a problem (55).

Automatic text clustering is a process that analyses the distribution and context of words, parts of words or terms in titles and abstracts, to find patterns in the data. The function generates groups based on common features in the titles and abstracts, where each group is given a name based on some semantic similarity of the references included in the group. Each reference can be assigned to one or more groups. This function can be used to identify groups of relevant or irrelevant references (10).

Classifiers, or classification algorithms are trained on prelabelled data to be able to make predictions on new data according to whether or not it has these characteristics. All classifiers we use are binary classifiers. Three classifiers[1] in the EPPI-Reviewer software (47) are:

- ***Cochrane RCT classifier*** - This classifier has been trained and validated on 280,000 health-related references, which means that it can distinguish between randomised controlled trials (RCTs) and other study designs with a high degree of certainty (56;57). This classifier classifies the references into two groups "likely to be an RCT" and "unlikely to be an RCT", with 99 per cent recall. Cochrane recommends that all systematic reviews of RCTs use this classifier and only consider references classified as "likely to be an RCT", i.e., they recommend that all studies categorised as "unlikely to be an RCT" can be excluded without manual review.
- ***Custom classifier*** - A bespoke classifier model that you build yourself that is specific to your project. Here you use your already included and excluded references at title and abstract level to train and test the model. When presented with new references (i.e., the unscreened references), it makes predictions on relevance based on the previously included and excluded studies used for training. Custom classifiers are often used to categorise remaining references by probability of meeting/not meeting the inclusion criteria. Classifiers can also be used to categorise objects dichotomously (category x/not category x). The custom classifier model in EPPI Reviewer presents the references according to percent probability of meeting the inclusion criteria (58).
- ***Systematic review classifier*** is a model that has been trained and validated on a large number of health-related references from the University of York's "Database of Abstracts of Systematic Reviews of Effect", which enables it to distinguish between systematic

[1] There are also other pre-build classifiers in EPPI, such as "covid 19 classifier" and "health economic classifier". These have not been tested or validated by the ML team.

reviews and other study designs (47). Unlike the Cochrane RCT classifier, we have no Figures on precision and recall, and judgements about how to use the classifier are made on a project-by-project basis.

Machine learning (ML) is a sub branch within artificial intelligence where statistical methods are used to make predictions about new data (59). In simple terms, machine learning means that we use algorithms that enable the computer to learn from and continuously develop its predictions based on empirical data that we feed it.

OpenAlex is an open access dataset with more than 250,000,000 scientific objects (references including institutional reports, grey literature, conference abstracts etc.) It is a scientific knowledge graph that uses graph neural networks to identify references in the dataset based on either a search with keywords/search strings, or seed studies. Instead of searching for subject headings or keywords in the titles and abstracts of studies, OpenAlex links references based on the content and meaning of the text (60).

Priority screening is a ranking algorithm in the EPPI Reviewer software (47) that is trained by the researchers' decisions on inclusion and exclusion of references at title and abstract level. Ranking algorithms work in a similar way to binary classifiers, and references that the algorithm considers more relevant based on the researchers' inclusion decisions are pushed forward in the "reference queue". In this way, we get a quicker overview of how many references possibly fulfil the inclusion criteria than if we were to read the references in random order.

Norwegian Institute of Public Health

Published by the Norwegian Institute of Public Health
April 2024
PO Box 222 Skøyen
N-0213 Oslo
Tel.: (+47) 21 07 70 00
The report can be downloaded as a pdf at www.fhi.no