Implementation of Machine Learning in Division for Health Services: Strategy Proposal from 2024

NIPH
Norwegian Institute of Public Health

Contents

Key Messages

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. The ML team has since become an international leader in integrating and implementing ML into evidence synthesis.

The overall goal of the ML 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 review process.

This report presents ML 3.0s strategy suggestion from 2024 and onwards based on the team's experiences from its inception in 2020.

In response to the evolving needs of our institute and in alignment with the strategic objectives of both the institute and the division, this proposal advocates for the elevation of the ML team to division level from 2024 and onwards. This will ensure long-term sustainability and decrease financial burdens for HTV. In addition, we propose a restructured organizational framework with three teams: Innovation and Horizon Scanning, Evaluation and Evidence Building, and Implementation and Support, as well as a steering committee to coordinate activities and engage in external networking.

Title:
Implementation of Machine Learning in Division for Health Services: Strategy Proposal from 2024

Publisher:
The Cluster of Reviews and Health Technology Assessments, Division for Health Services, ML team 3.0

Type of publication:
Report

Executive summary

Background
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, aligning with NIPH's strategies for automation and workflow innovation. This was also in line with the goals of the Division for Health Services which sought to automate work processes and efficiently summarize evidence using ML. Since its inception, the ML team has positioned NIPH as a leader in implementing ML into evidence synthesis, achieving significant milestones, and securing official financing in November 2022.

The report presents suggestions for the strategy from 2024 and onwards based on the experiences of ML implementation from 2020 up to the latest iteration of the team, ML 3.0.

Suggested short-term strategy and focus areas
The proposed short-term strategy aims to address the imminent challenges faced by the ML team within HTV, particularly the departure of key members. The focus is twofold: first, to keep the ML team operational throughout 2024, and second, to prioritize activities aligning with HTV's overarching ML goals. The activity plan includes recruiting a new leader with ML experience and 1-2 team members, including at least one information specialist. The short-term strategy centres on maintaining and enhancing current employees' ML knowledge. Capacity-building activities will be guided by employees' preferences, including peer-to-peer support, technical workshops by function experts, and expanding the successful e-learning initiative from 2023 to cover additional ML aspects. This approach aims to ensure continuity, upskilling, and sustained focus on ML goals in the short term.

Suggested long-term ML strategy
In response to the evolving needs of our institute and in alignment with strategic objectives of both the institute and the division, this proposal advocates for the elevation of the ML team from cluster level to division level as a long-term strategy. The proposal aligns with both the institute and division strategies, emphasizing the importance of evolving infrastructure, knowledge support, and expertise in innovative methods like ML. Elevating the ML team is seen as imperative for innovation and collaboration in evidence synthesis and public health, preventing limitations in capacity and demotivation of team members.

Elevation of the team to division level will benefit HTVs work with evidence syntheses as it will ensure long-term sustainability of our ML efforts, more freedom in exploring ML tools aimed at specific evidence synthesis processes, as well as decreasing financial burdens for HTV.

Without this transition, there is an elevated risk of limited capacity, tasks, and time for the team, hindering its ability to keep up with advancements in the field. There is also the potential of becoming merely a "maintenance team," leading to demotivation and high turnover.

<u>ML Team 4.0 organization</u>

To address the challenges of rapid growth in ML and artificial intelligence (AI) tools, a restructured organizational framework for the ML efforts is proposed, dividing the team into three distinct teams: Innovation and Horizon Scanning, Evaluation and Evidence Building, and Implementation and Support. Within each team we suggest a further two-pronged division of activities: One related to the use of ML and AI in evidence synthesis and one related to the use of ML and AI in other research areas, like primary research, registries etc. This will ensure a continued investment and development of ML in the evidence synthesis process, as well as acknowledging the increased focus on primary research and registry work within the division and NIPH as a whole. Each team would have specific responsibilities, and a steering committee comprising the team leads would be established to coordinate activities and engage in external networking. The restructuring aims to enhance resource efficiency, introduce new competencies, increase the talent pool of potential new members, and facilitate collaboration across departments. Clear criteria for team members are outlined based on their roles, emphasizing skills in ML and AI, communication, teaching, and experience in evaluation and implementation.

Suggested focus areas for long-term strategy

The current iteration of the ML team has put forward suggestions for key focus areas for the next iteration of the team, which align with the overarching goals of the division and institute. These include seeking external funding for financial sustainability, continue exploring the use of OpenAlex in evidence synthesis in collaboration with the librarian team, and fostering continued collaboration with external institutions, particularly EPPI Centre. The team also aims to strengthen interdisciplinary collaboration within NIPH, build expertise in Generative AI, develop a comprehensive ML implementation package for training other institutions, explore and evaluate data extraction tools, and investigate the application of ML/AI in the institute's registry work.

Preface

This report presents a strategy suggestion for the next iteration of the ML team, "ML 4.0". The current iteration of the team, ML 3.0, has crafted these suggestions based on our reflections of our successes, challenges, and learning past and present.

Financing
Much of the work, particularly relating to innovation activities, was externally funded. The remaining work was self-initiated and financed by the Cluster for Reviews and Health Technology Assessments, Division for Health Services at the NIPH.

With appreciation
The current team's learning and strategizing 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 team of librarians who have started evaluation work on OpenAlex. 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.

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

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

Background

Using machine learning (ML) in the evidence synthesis process optimally utilizes scarce human resources by handling complex, repetitive tasks, allowing reviewers to focus on more thought intensive parts of the evidence synthesis process. 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. Since its inception, the ML team has positioned NIPH as a leader in integrating and implementing ML into evidence synthesis processes.

To address the challenges of the recent major reorganization, which has seen a drastic reduction in employee numbers in HTV, the rapid growth in ML and AI tools in combination with the evolving needs of our institute and in alignment with strategic objectives of both the institute and the division, this proposal advocates for the elevation of the ML team to the division level. Such a move is strategically positioned to enhance collaboration and ensure sustainability of ML efforts in the evidence synthesis processes in the division and the opportunity to expand to other divisions and research methods.

Suggested ML strategy from 2024 and onwards

Below we present our ML strategy proposition, which involves elevating the ML work from HTV to division level. It is a two-phased strategy divided into short- and long-term goals. The short-term strategy highlights our proposal of the ML work that should be conducted in HTV during the remainder of 2024, while the long-term strategy is our more ambitious proposal towards making the ML work in the division sustainable in the long-term.

This proposal resonates with NIPHs strategy (1) of evolving infrastructure to meet new challenges and supports the increasing need for knowledge curation and development in health decision-making. It also aligns with the division strategy (2), which emphasizes autonomous employees, expertise in project management, and the integration of new methods such as machine learning and automation.

Short-term strategy: ML maintenance and developments within HTV

Recognizing that the proposed significant shift in direction requires time for implementation, we emphasize the need for a short-term strategy to guide immediate ML activities within HTV. A critical challenge addressed by this strategy is the departure of four out of six ML team members, including the team and implementation leads, from January to May 2024. The short-term strategy concentrates on two primary objectives:

1. Ensuring the viability of the ML team throughout 2024. This will be done by recruiting a new leader with ML expertise and 1-2 team members, with at least one being an information specialist. There will be a period of overlap between leaders to ensure a smooth transition for both the new leader and team members. This period will be used to set specific goals and develop a working plan to achieve them during 2024.
2. Determining the activities that align with HTV's overarching ML goals. The overall goal of the ML team is to use ML in a way that best combines human intelligence and machine learning, to enhance human activities, by figuring out how best to integrate ML and workflow changes, throughout the review process. The team's activities since initiation have covered the following areas:
   - **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 suggested focus areas for the short-term strategy will primarily involve implementation and capacity building activities, to maintain and enhance HTV employees' ML knowledge. To guide capacity-building decisions, we will prioritize activities that have been well-received by employees or have proven effective in facilitating learning. This involves offering peer-to-peer support to project teams, conducting workshops - similar to the previous ML week structure or more dispersed one-day workshops where function experts provide technical interactive sessions on using various ML functions. This initiative may extend to cover additional ML or AI applications if necessary. Additionally, building on the positive feedback from the e-learning courses developed in 2023, the focus will be on creating more e-learning courses, preferably in microlearning format, covering various aspects of ML, including other ML or AI functions, suggestions for when to stop screening and how to use the ML reporting template.

In parallel, preparations for implementation of the long-term strategy should be conducted by the leader(s) of the ML team in close collaboration with designated leaders in the division. This comprehensive approach aims to address immediate challenges and ensure a seamless transition while staying aligned with HTV's overarching ML goals. Although we will focus on implementation and capacity building, we do aim to further explore the potential use of large language models in the evidence review process.

Long-term strategy: ML Team 4.0 as part of a division portfolio

Our rationale for the elevation of the ML team from HTV to the division level encompasses a multitude of factors that underscore the necessity of this organizational transition, which are presented below:

1. **Facilitating cross-collaborations**: Elevating the ML team to division level presents better opportunities for collaborations across clusters and divisions, fostering a more

integrated and interdisciplinary approach to our work. Cross-collaboration has been talked about for a long time at the institute, without much progress, which this proposition aims to change.

2. **Ensuring team sustainability:** Operating at division level secures the long-term sustainability of the ML team, by a) increasing the pool of potential team members, ensuring a diverse skill set and a broader range of expertise; and b) reducing vulnerability to employee turnover as there will be a larger pool of potential team members. Together this will safeguarde the continuity of core team members and ML/AI knowledge within the ML team, and making the ML team more robust against future reorganizations.

3. **Resource benefits for HTV:** Elevation decreases the financial burden for HTV, as the cost for ML team members will be divided across departments involved with the ML team. Overall resources within HTV will be less strained, as the pool of potential team members increases to all employees working within the division.

4. **Testing ML tools on primary research**: Being at division level better enables the evaluation and testing of ML tools on primary research considering the majority of the primary research conducted in the division is not based in HTV. Also, our existing ML expertise in evidence synthesis can help inform the work on primary research.

5. **Utilizing EPPI Reviewer across products and teams:** . The data management tool we use, EPPI Reviewer, is used by many groups across NIPH. The suggested transition facilitates further increased and optimal use of EPPI Reviewer across products and teams, improving the efficiency of evidence synthesis processes across the division.

6. **Global leadership position:** To maintain the ML teams position as a global leader in ML implementation, it is imperative to expand the pool of talent, ensuring the team's motivation is driven by innovation rather than mere fulfillment of requirements. By elevating the team to division level, puts our ML work in better position to remain an implementation leader of ML in evidence synthesis processes, as well as opening up for ML innovation within other research fields, like primary research.

ML Team 4.0 division level organization

By elevating the ML team efforts to division level, we not only ensure the sustainability of the team but also contribute significantly to the overarching goals of our division and institute. To further facilitate the divisions ML and AI work, we propose a restructured framework for Team 4.0 with a clearer division of tasks and responsibilities, driven by two primary factors. Firstly, the rapid growth and expansion of the field of ML and AI make it impractical for a single team to adequately perform horizon scanning, evaluation, and implementation. This has currently led to a strain on the team's resources, hindering the execution of activities at a satisfactory level. The second is the opportunity to bring in team members with specific competencies and/or particular interest within evaluation, implementation, or teaching. These new team members will bring the skills and motivation we need to continue our work and receive support and onboarding from the ML team.

The proposed structure involves dividing the existing team into three distinct teams, see Figure 1; Innovation and Horizon Scanning, Evaluation and Evidence Building, and Implementation and Support, each with a team lead, and establishing a steering committee to

coordinate activities. The steering committee would also be responsible for planning the implementation of the long-term strategy in parallel with executing the short-term strategy.

Within each team we suggest a two-pronged division of activities: One related to the use of ML and AI in evidence synthesis processes and one related to the use of ML and AI in other research areas, like primary research. This will ensure a continued investment and development of ML in the evidence synthesis processes, as well as acknowledging the increased focus on primary research and registry work at both division and institute level. The teams and their respective roles, responsibilities and necessary skill sets are outlined below. Alternatively, the two different directions (evidence synthesis vs other research areas) can be organized in separate teams, if this is seen as a better option.
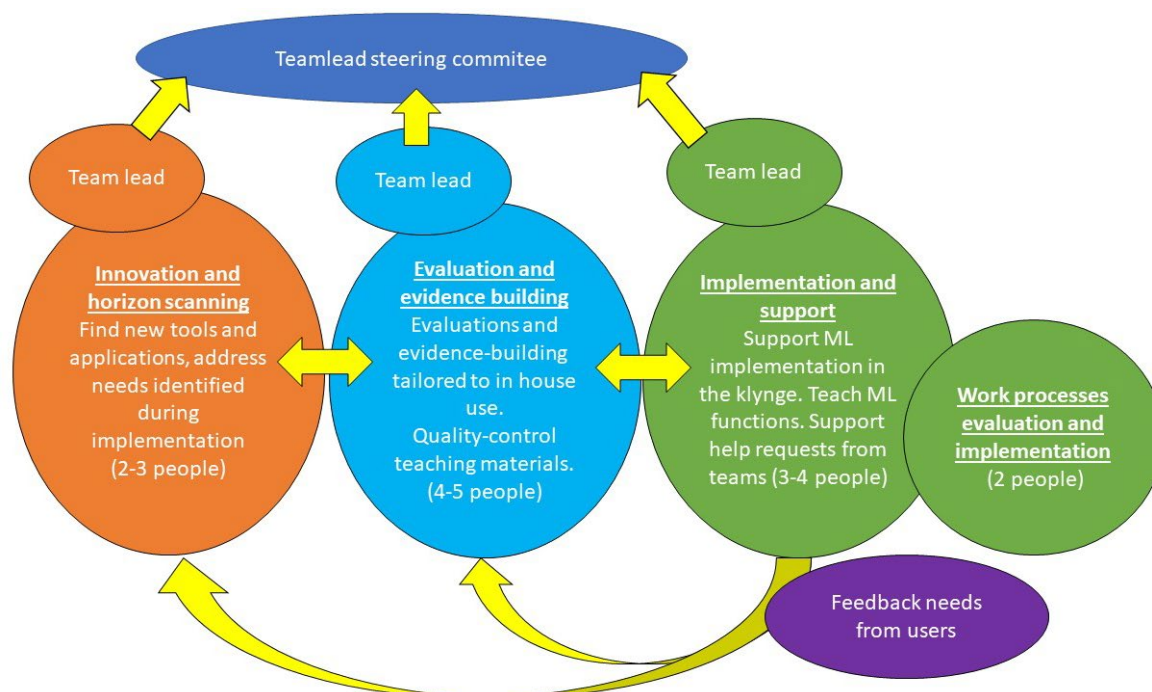

Figure 1: Proposed structure for ML 4.0

1. Innovation and horizon scanning
This team will consist of a team lead and 1-2 team members. They will have responsibility for horizon scanning for new tools within ML and AI, mapping these tools, doing a preliminary test, and then deciding which tools should be evaluated. They should also be responsible for mapping already published evaluations for new relevant tools. They will also have responsibility for the OpenAlex automated search for new research on ML and AI that is relevant to our evidence synthesis work.

Effectively conducting horizon scanning and innovation requires individuals to possess a combination of essential qualities. These include being organized, having good searching skills and a forward-thinking vision and curiosity to anticipate and explore emerging trends. Furthermore, they should have an understanding of ML and AI and potential ways it can be used in our work, and clear communication skills.

11

2. Evaluation and evidence building

This team will consist of a team lead and 3-4 team members (dependant on the number of ongoing evaluations). When the innovation and horizon scanning team identifies a new tool or function that they feel would be a good fit for our work, the evaluation and evidence building team will plan and conduct an evaluation of the new tool, if we have not already identified an evaluation already conducted by another evidence synthesis group. This team will ensure that all new tools are beneficial to our work by saving time, increasing productivity, or helping with complex problems. They will also ensure that the new tool maintains or improves the quality of our products. This team will feed back to the innovation and horizon scanning team about any specific tools or needs they feel the team should be trying to identify.

If an evaluation shows that a tool should be implemented into our work processes, the evaluation lead will present the tool to the implementation and support team and have a discussion around implications for implementation and the way forward.

Team members need to have a good understanding of ML and AI functions and how they work as well as our internal work processes to plan and conduct evaluations. It is a clear advantage if some team members have a background in analysis/statistics and experience in planning and conducting evaluations.

3. Implementation and support

This team will consist of a team lead and 4-5 team members broken into two working groups. This team will feedback user needs to the two other teams. The first working group of 2-3 team members will have responsibility for implementation and peer-to-peer support. The second working group of two people will have responsibility for addressing how using ML/AI changes our workflow processes and implications for project management.

The implementation and user support working groups will develop new teaching materials and implementation strategies in communication with the evaluation lead when a new tool is identified for widespread implementation. The team will create an implementation plan and identify what type of teaching materials and user support are needed. The team will also give peer-to-peer support to project teams on request.

The work processes evaluation and implementation work group will focus on evaluating changes related to workflow and assessing implications for implementation related to these changes. They will oversee implementing any work process changes that are identified as being beneficial through evaluations.

Team members need to have a good understanding of ML and AI functions and how they work as well as good teaching and communication skills. A further benefit would be experience with implementation and evaluation of implementation projects and/or qualitative research. Team members should have excellent communication skills, be patient and have an ability to speak with hesitant or sceptical colleagues without being judgemental.

Team lead steering committee

The three team leads will sit in a steering committee where they will coordinate activities and discuss tools that are ready of evaluation and implementation. This group will also have responsibility for networking with external actors and sitting in working groups linked to project partnerships in, for example, ICASR. A fourth member of the team lead steering committee will provide administrative coordination. This role will entail having an overview over what is happening, setting up meetings, coordinating funding applications, sending out the newsletter and keeping the website and SharePoint room up to date.

Suggested focus areas for long term strategy

In the following section are some suggestions for possible long term focus areas for the ML team. These focus areas align well with both the division and institute strategies, as well as supporting the team's overarching goals and contributing to its continued positioning as a leader in ML implementation within the evidence synthesis field.

1. **Strengthen interdisciplinary collaboration within NIPH.** During 2023 we have seen a lot of interrest of HTVs ML work in evidence synthesis from other clusters and divisions in the institute, which reflects the need for the ML work to become more available to the whole institute. We have provided peer-to peer support for groups in the of Global Health cluster, the Norwegian Scientific Committee for Food and Environment and Division of Infection Control. Increased networking across NIPH divisions Elevation of the ML work enables for increased interdisciplinary collaboration and aids in the dissemination of our work.

2. **Towards a more standardized evidence synthesis practice across the institute.** Groups across the whole institute perform in various degrees evidence synthesis work, and these efforts will most likely increase in the coming years. Unfortunately, there is no standardized practice for how evidence syntheses should be conducted at institute level. Elevating the ML work to division level better positions us to disseminate our work and providing capacity building efforts across the institute, which could also turn out to be the initiative which drives the evidence synthesis work at the institute towards more standardized practices

3. **Develop a complete ML implementation package**: We have during 2023 made a scalable e-learning course on the conceptual parts of our most used machine learning functions. We want to continue developing the e-learning course as well as adding technical how-to components, to be able to provide a complete training package that can be delivered to other groups in-house as well as outside of NIPH and to collaborating public health institutes globally. This aligns with the division strategy's emphasis on collaboration to strengthen public health institutions by sharing expertise. This has also the potential of creating revenue. This should also be integrated into the evidence synthesis teaching materials that are already being provided by HTV, to create a complete evidence synthesis "how-to" package.

4. **Seek funding from external and international sources:** Seeking external funding contributes to the financial sustainability of the team and is in line with the overarching goals of the division to collaborate and strengthen public health institutions globally. Potential relevant Norwegian funding sources is the Norwegian Research Councils upcoming [research efforts](#) in artificial intelligence and digital technologies which will be

increased by at least NOK 1 billion over the next five years, aimed to contribute to greater insight into the consequences of technological development for society. Possible international funding opportunities include the EU Networking grant (ICASR collaboration) and the UK better methods grant, which could be a joint application with the EPPI Centre.

5. **Increase visibility of the divisions published work that is not indexed.** This enhances the accessibility and usability of the reports, which aligns with the divisions strategy for open science. Two examples of this is:
   a. To apply for DOI numbers for our reports (e.g., when uploading publications in ResearchGate you have the option to generate a DOI number for that publication). This aligns with the divisions strategy for open science. Implementing the use of DOI numbers will allow our publications to be "web scraped" to find the evidence we produce as it is not indexed in databases.
   b. To facilitate machine readability for open data sharing: Engage with other systematic review groups such as The FHIR Resources for Evidence-Based Medicine Knowledge Assets (EBMonFHIR) (3) to begin working towards standardized formatting and language for evidence syntheses to make open data sharing possible.

6. **Further explore and evaluate AI and ML tools that are highly relevant for our work in the division**: This aligns with the division and institute strategies for adopting and increasing expertise in new methods, contributing to efficiency and advancements in evidence synthesis, as well as staying at the forefront of innovation. Tools of particular relevance might be:
   a. **OpenAlex:** Using OpenAlex to create libraries of relevant literature for different projects or topic areas. These libraries would be housed in EPPI Reviewer where research teams could code and categorise them for easy sifting.
   b. **RobotReviewer**: This tool has been dormant for a while, but work will start up again. As this is a tool we have evaluated some parts of in previous work in 2022 (4), it will be very useful to continue evaluations of this tool.
   c. **Generative AI tools**: Capacity building and evaluations of generative AI tools is particularly relevant, as there are many tools available, and due to the nature of the tools and their AI functionality, many aspects of e.g., data extraction can be explored, for example PICO extraction, risk of bias evaluations for many different study designs as well as extraction of results. A focus on generative AI also ensures that the team remains at the forefront of innovative methodologies and stays on top of the generative AI field, which is the AI area that has seen the most explosive development since the launch of ChatGPT in November 2022.

7. **Continued external collaboration with EPPI Centre and other key institutions:** This aligns with the division strategy's emphasis on collaboration to strengthen public health institutions and fosters interdisciplinary collaboration. Also, this will increase visibility of both the ML work being conducted in the division and well as safeguard against unnecessary duplicate work. Also, assisting EPPI Centre with evaluating and testing various tasks puts us in a prime position of being a key user organization that can provide input to EPPI Centre which will benefit our institution, with regards to improved interface and functionality of the EPPI Reviewer tool. Evaluations that EPPI Centre are interrested in involving us in are related to:
   a. Semi-automated data extraction and mapping of data

    b.   New clustering functions with large language models

    c.   Stopping criteria for priority screening- testing not only the different criteria but testing the usability of the function.

    d.   Vector database work

    e.   Custom classifiers

8. **Explore possibilities with using ML/AI for the institutes registry work:** Registries will be a key focus area for the institute the coming years due to the centralization of the health registries to NIPH. Exploring how AI/ML can be used in this work should be considered, for example in relation to using generative AI to extract data directly from registry data in tabular format.

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. Område for Helsetjenester. Område for helsetjenester 1-5 år framover. Oslo: Folkehelseinstituttet; 2022.
3. Brian A, Martin M, Khalid S, Joshua R, Lisa S, Mario T, et al. 20 Achieving evidence interoperability in the computer age: setting evidence on FHIR. BMJ Evidence-Based Medicine 2019;24(Suppl 1):A15. DOI: 10.1136/bmjebm-2019-EBMLive.28
4. 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 Med Res Methodol 2022;22(1):167. DOI: 10.1186/s12874-022-01649-y